CONTENTS

Symposium

FAITH IN THE PUBLIC SPHERE
Shami Chakrabarti ................................................................. 483

FACE VEIL BANS IN THE EUROPEAN COURT OF HUMAN RIGHTS: THE IMPORTANCE OF EMPIRICAL FINDINGS
Eva Brems .................................................................................... 517

GENDER AND (RELIGIOUS) ATTIRE: A MATTER OF (FREE) SPEECH
Alejandro Madrazo, LL.B., LL.M., J.S.D. ....................................... 553

RELIGIOUS FREEDOM: EXPRESSING RELIGION, ATTIRE, AND PUBLIC SPACES
Lucy Vickers .................................................................................. 591

(WHEN) CAN RELIGIOUS FREEDOM JUSTIFY DISCRIMINATION ON THE BASIS OF SEXUAL ORIENTATION?—A CANADIAN PERSPECTIVE
Noa Mendelsohn Aviv ...................................................................... 613

RELIGIOUS FREEDOM AND EQUAL TREATMENT: A UNITED KINGDOM PERSPECTIVE
Karon Monaghan QC ....................................................................... 673

INCONVENIENCE OR INDIGNITY? RELIGIOUS EXEMPTIONS TO PUBLIC ACCOMMODATIONS LAWS
Marvin Lim, J.D. and Louise Melling, J.D. ........................................ 705

WHO DECIDES CONSCIENCE? RFRA’S CATCH-22
Priscilla J. Smith, J.D. ..................................................................... 727
ARTICLE

COMMON INTEREST COMMUNITY COVENANTS AND THE FREEDOM OF CONTRACT MYTH
Andrea J. Boyack, J.D. ................................................................. 767

NOTES AND COMMENTS

ACCOMMODATING EMPLOYERS’ INTERESTS INTO THE DISCUSSION OF EMPLOYMENT PROTECTIONS FOR VICTIMS OF DOMESTIC VIOLENCE
Timothy John Durbin ................................................................. 845

MOVING PAST A “POCKET CHANGE” SETTLEMENT: THE THREAT OF PREEMPTION AND HOW THE LOSS OF CHANCE DOCTRINE CAN HELP NFL CONCUSSION PLAINTIFFS PROVE CAUSATION
John Guccione ........................................................................... 907

INTERSEX CHILDREN IN FOSTER CARE: CAN GOVERNMENT ELECT SEX ASSIGNMENT SURGERY?
Ashley Huddleston ................................................................. 957

DRUGS: YOU USE, YOU GAIN? WHY COURTS SHOULD UPHOLD LONG-TERM DISABILITY BENEFITS FOR RECOVERING ADDICTS
Gregory M. Juell ................................................................. 1005
INTRODUCTION

When people discuss different equality rights conflicting with each other, they often have in mind conflicts between religious beliefs and gender or sexual orientation rights. Conflicts between other equality rights seem (so far at least) to arise less frequently. Economic difficulties and increased competition for resources may change this, but my right not to be discriminated against (at least directly) on the grounds of my sex or race rarely causes difficulties for others. Equally, it is difficult to think of examples where preventing discrimination on the grounds of disability, sexual orientation, or age will involve significant compromise to someone else’s protected rights.

The challenge (and the point) of religious and philosophical beliefs is that they inevitably invoke moral structures, which are not universally shared, and which may not be reflected in modern legal norms. For example, many religions are based on ancient doctrines reflecting patriarchal ideals. This inevitably leads to conflict between those beliefs and the rights of people of different faiths or consciences, women, or gay people.

But these specific and predictable conflicts are not the only problems. Issues of particular faith identity (or the lack thereof) seem to have become particular sore points in our cultural discourse at both the national and local level. You hear of people taking personal exception to accommodations being granted to a fellow employee on the grounds of his or her religion. Their reasoning is that it’s “not fair” for someone to be allowed time off for religious observance when others are not allowed to leave

* Shami Chakrabarti has been the Director of the U.K. human rights group Liberty since September 2003.
early, for example, to play sport. I suspect you would not hear the same complaints—at least not so publicly—about accommodations being made for disabled employees or perhaps even for female colleagues with childcare difficulties.

Ten years ago, the Employment Equality (Religion or Belief) Regulations of 2003\(^1\) came into force in the U.K., and have subsequently been superseded by the Equality Act of 2010.\(^2\) It has also been thirteen years since the Human Rights Act of 1998 came into force incorporating Article 9 (freedom of thought, conscience, and religion) of the European Convention on Human Rights into the U.K.’s domestic law.\(^3\) The cases about faith in the public sphere have generated some of the most outspoken commentary, in part because of the huge divergence in views towards religion in the U.K.

Some people have adopted a new breed of aggressive secularism (perhaps an inevitable instinctive response to the rise of international fundamentalist Christian and Islamic movements) that seeks to eradicate religion from public life altogether.\(^4\) Meanwhile, there are substantial minorities of individuals with strongly held religious beliefs involving strict doctrine and practice. Some traditionalists mourn a perceived decline in the Church of England as dominant faith in the land.\(^5\)

Others have encouraged a political and legislative culture that concedes irritation, offence, alarm and distress, as evidenced by our public order and anti-social behavior statute books and promotes a general fear of difference and dissent.\(^6\) This in turn

---

6. See, e.g., George Monbiot, *At Last, A Law to Stop Almost Anyone*
produces both the nonsense of nervous “Winterval” celebrations\(^7\) and the disgrace of a young British man being arrested for calling Scientology a cult.\(^8\) There has also been an increase in hostility towards religious minorities which has manifested itself recently in calls to ban the wearing of the burkha in public places, most vocally and stringently in continental Europe, but also in the U.K.\(^9\)

Society has three choices in dealing with the question of the extent to which people have the right to express their religion in the public sphere. The first choice is to select and elevate an approved faith to the point of giving it dominant status over all other belief systems. That faith is completely and formally interwove into the entire legal, political, and social system—every sphere of public life and as much of private life as can be achieved. An extreme example might be Afghanistan under the Taliban, and a more moderate example would be Britain at earlier stages in its history.

The second option is in many ways both equal and opposite. It is based on the view that faith conviction should be viewed as dangerous and divisive. If faith conviction cannot be eradicated altogether, it must be chased from the public to the private sphere—confined to a place of worship, the home, or upstairs under the bed with the pornography. An extreme example would be Stalin’s Russia, and a more moderate one would be the French Republic.

There is also a third option: a more human rights-based approach and one that resonates well with a society like Britain,

---

\(^7\) Winterval was the name given to Birmingham City Council public events in 1997 and 1998. “Winterval” has since become shorthand in the U.K. for attempts to “rebrand” Christmas so as not to exclude non-Christians.


a country where the struggle for religious freedom has been so connected with the struggle for democracy itself.

Human beings are creatures of faith and logic, emotion and reason, and this is reflected in the law. It may be true that religion has inspired considerable war and prejudice, but it has also been responsible for art, music, and compassion. While scientists and engineers have produced some of the greatest advancements in human history, their work has also been the stuff of nightmares. If we really believe in freedom of thought, conscience, and religion, then such freedom must include the right to the faith or belief of your choice, the right to no faith, and crucially, to be a heretic to any religion.

Inseparable, enumerated rights like freedom of conscience, expression, and association, and the right to private and family life, all flow from foundational human rights ideals of dignity, equal treatment, and fairness. Lord Nicholls in Williamson (a case concerning corporal punishment in schools) said:

Religious and other beliefs and convictions are part of the humanity of every individual. They are an integral part of his personality and individuality. In a civilised society individuals respect each other’s beliefs. This enables them to live in harmony. This is one of the hallmarks of a civilised society. Unhappily, all too often this hallmark has been noticeable by its absence. Mutual tolerance has had a chequered history even in recent times. The history of most countries, if not all, has been marred by the evil consequences of religious and other intolerance. ¹⁰

Some of those historical examples highlight one of the largest dangers that can arise from religious discrimination, namely religion being used as a proxy for race. Sadly, this is not a phenomenon confined to history, as is amply demonstrated by the evolving—and increasingly toxic—debate on the wearing of the burkha in public spaces, which is considered below.

As with other forms of individual expression and autonomy,

we should be slow to interfere with the expression or manifestation of any religious or other belief—doing so only when such intervention is necessary and proportionate to protecting the rights and freedoms of others. This can of course be a difficult exercise in practice, and there are a collection of core issues which have proved consistently controversial.

In this Article, I focus on two issues. First, I consider what religious and philosophical beliefs the law deems to be worthy of protection. I will look at cases that discuss the scope of “philosophical belief” in the U.K. Employment Appeal Tribunal, the seminal decision of the European Court of Human Rights in Eweida v. United Kingdom,11 which considers the issue of beliefs held by relatively few people, and a subsequent decision in a U.K. Tax Tribunal on the same issue. Second, I consider how far the law requires us to go to protect the manifestation of those beliefs. Specifically, I will analyze the legal position of public officials and business owners providing services to the public and the rights of individuals to wear religious clothing in public, both in the U.K. and in France. Finally, I will conclude by analyzing recent developments in the European Court of Human Rights’ approach to religious freedom.

I. WHICH RELIGIONS, BELIEFS, OR MANIFESTATIONS ARE WORTHY OF PROTECTION?

How serious does a belief have to be in order to deserve protection? While U.K. courts and the European Court of Human Rights have said such a belief must “attain a certain level of cogency, seriousness, cohesion and importance,”12 what does that mean? How do you draw the line between “beliefs” and convictions that are idealistic, scientific, or political? What about beliefs that are offensive or discriminatory? What if you are the only person who holds your particular belief or you interpret your religion in an idiosyncratic way?

---


These are questions that some find contentious. But in general, a human rights-based approach lends itself to a generous interpretation of the concept of what constitutes religion and belief for the purposes of defining the scope of legal protection. The seriousness of the belief, the extent to which it affects others, and the number of people sharing it might all be relevant factors in deciding whether any interference is justified, but it is surely better not to shut out certain beliefs from being protected at all. The last thing we want are judges—or employers, for that matter—making value judgments about the types of beliefs that are worthy of respect.

This is broadly what the House of Lords, formerly Britain’s highest domestic court, was contemplating in *Williamson.* One of the judges in the court below, the Court of Appeal, had thought that a belief in the principle of “spare the rod and spoil the child” did not qualify for protection as a religious belief at all. However, that view was firmly rejected by the House of Lords. Lord Nicholls said:

> When the genuineness of a claimant’s professed belief is an issue in the proceedings the court will inquire into and decide this issue as a question of fact. This is a limited inquiry. The court is concerned to ensure an assertion of religious belief is made in good faith: “neither fictitious, nor capricious, and that it is not an artifice”, . . . . But, emphatically, it is not for the court to embark on an inquiry into the asserted belief and judge its “validity” by some objective standard such as the source material upon which the claimant founds his belief or the orthodox teaching of the religion in question or the extent to which the claimant’s belief conforms to or differs from the views of others professing the same religion. Freedom of religion protects the

---


14 Regina (Williamson & Others) v. Sec’y of State for Educ. & Empl’t, [2002] EWCA (Civ) 1926, [23], [2003] QB 1300 at 1310 (Eng.).

15 *Williamson*, [2005] UKHL 15 at [87].
subjective belief of an individual. [R]eligious belief is intensely personal and can easily vary from one individual to another. Each individual is at liberty to hold his own religious beliefs, however irrational or inconsistent they may seem to some, however surprising.\(^\text{16}\)

**A. What Beliefs Are Worthy of Protection: The Approach of the U.K. Employment Appeal Tribunal**

The U.K. Employment Appeal Tribunal (“EAT”) considered the scope of the concept of “philosophical belief” in *McClintock v. Department of Constitutional Affairs*.\(^\text{17}\) The case concerned a magistrate who refused to officiate because he might have had to decide whether children should be placed for adoption with same-sex partners and then resigned from his role in family law cases.\(^\text{18}\) The magistrate claimed that, in breach of the 2003 Regulations, he had been discriminated against on the basis of his philosophical beliefs. He did not say that he believed adoption by same sex couples was wrong as a matter of principle; just that he thought that there was no convincing evidence that it could be in a child’s best interests.\(^\text{19}\) It also appears that the magistrate would have been willing to change his mind in light of further research.\(^\text{20}\)

The EAT adopted the test for “philosophical belief” set out by the European Court of Human Rights in *Campbell & Cosans v. United Kingdom*:\(^\text{21}\) that the belief must have sufficient cogency, seriousness, cohesion, and importance, and be worthy of respect in a democratic society.\(^\text{22}\) The EAT found that because Mr. McClintock had never framed his objections on the basis of any

\(^\text{16}\) Id. at [22].
\(^\text{18}\) Id. at [4].
\(^\text{19}\) Id. at [7].
\(^\text{20}\) Id.
\(^\text{22}\) Id. at para. 36.
religious or philosophical belief, he fell outside the scope of the 2003 Regulations. The tribunal had correctly observed that it is not enough “to have an opinion based on some real or perceived logic or based on information or lack of information available.”

McClintock demonstrates that while the courts will not judge the “validity” of a claimant’s belief—a possibility ruled out in Williamson—it will consider whether the purported belief is in fact a belief based on principle, rather than a mere opinion based on the available evidence. As it happens, if Mr. McClintock had maintained a protected belief that it was simply wrong for same-sex couples to adopt, then the outcome would surely have been the same. But if religion is to enjoy neither a punished nor privileged status in society, and accepting that all human beings are to some extent creatures of logic and emotion, faith and reason, there is no real justification for attempting to distinguish a deeply held belief based on evidence from one taken on faith.

The scope of protection for religious and philosophical beliefs in the U.K. has undoubtedly been extended by Grainger Plc v. Nicholson, the “green martyr” case. Mr. Nicholson had been dismissed by the defendant and the defendant claimed that the dismissal was due to redundancy. Mr. Nicholson claimed that he was discriminated against based on his asserted philosophical belief in relation to climate change and the environment. The question for some might be: why would a climate change campaigner want a tribunal to treat his convictions as a “philosophical belief” rather than as a scientific fact? The cynical answer would be that it was the only way he could challenge his dismissal. But actually, it seems the belief in issue was much more than just a belief in climate change itself. It was also a belief that we are all morally obliged to take urgent steps to address the causes of climate change though our lifestyles and any other means available. The EAT said that a belief of this kind—provided it was of a similar cogency or status to a religious

---

25 Id. at [2].
26 Id.
belief—could fall within the legal framework designed to protect faith and conscience in the workplace. If Mr. Nicholson was made redundant simply for holding this belief, then why shouldn’t he be entitled to a remedy from the Tribunal?

In his judgment, Justice Burton summarized the limitations on the concept of “philosophical belief”:

(i) The belief must be genuinely held.
(ii) It must be a belief and not . . . an opinion or viewpoint based on the present state of information available.
(iii) It must be a belief as to a weighty and substantial aspect of human life and behaviour.
(iv) It must attain a certain level of cogency, seriousness, cohesion and importance.
(v) It must be worthy of respect in a democratic society, be not incompatible with human dignity and not conflict with the fundamental rights of others.

The scope of protection was considered again in Power v. Greater Manchester Police Authority. Alan Power, a former employee of the Police Authority, claimed that he was dismissed because of his spiritualist faith and that his belief that psychics should be used in criminal investigations. The judge in the Employment Tribunal found that a belief in life after death and the capacity to communicate with spirits “on the other side” had the necessary cogency, seriousness, cohesion, and importance to qualify as a belief worthy of respect in a democratic society. The EAT upheld this decision and found that the test adopted in Grainger was satisfied.

---

27 Id. at [26].
28 Id. at [24].
30 Id. at [3]–[7].
31 Id. at [17].
B. Beliefs Held by Few People: The Approach of the European Court of Human Rights

The cases decided by the U.K. EAT, discussed above, demonstrate the breadth of the different types of beliefs that are capable of protection in the U.K. What, though, about those beliefs held by very few people? As Lord Nicholls recognized in Williamson, religious belief is “intensely personal,” and it would seem odd for protection to depend on whether the belief in question is shared by others who are also put at a disadvantage.

However, that appeared to be the effect of the EAT’s judgment in the case of Eweida v. British Airways Plc. The issue in that case was whether British Airways’ (“BA”) uniform policy—which prohibited Ms. Eweida from wearing a small cross around her neck—was indirectly discriminatory on religious grounds and therefore needed justification. Although it was not in dispute that Ms. Eweida was a committed Christian, and that it was a genuine and important part of her faith to wear her cross visibly, the EAT found that there was no indirect discrimination because Ms. Eweida had not shown that BA’s uniform policy disadvantaged Christians as a group.

The Court of Appeal upheld the decision of the EAT because there was no evidence that any other BA employee had ever requested to wear a visible cross, or been deterred from doing so. Liberty argued that indirect discrimination should not require a manifestation of belief to be shared between a group of people with the same protected characteristic. Specifically, that religion and the manifestation of belief is a deeply personal matter and a human rights-based approach should be sensitive to

---

34 Id. at [62]–[63].
36 Id. at [8], [38].
37 Id. at [7].
a genuine personal assessment of the requirements a faith places on its adherents.\textsuperscript{38} Whether an act is a “manifestation” of what is found to be a sincerely held religious belief should be judged by the believer, him or herself. Care must be taken to avoid engaging in any assessment of the validity of the belief that drives certain actions.

Ms. Eweida successfully pursued her claim in the European Court of Human Rights, under claims found through Article 9, the right to freedom of thought, conscience, and religion, and Article 14, the right to be free from discrimination.\textsuperscript{39} During its judgment, the court considered the scope of the right to freedom of thought, conscience, and religion. It reiterated that this right protects views that attain a certain level of cogency, seriousness, cohesion, and importance. The court indicated that the view reaches this level, and the state’s duty of neutrality and impartiality is incompatible with any action by the state to assess the legitimacy of someone’s religious beliefs or the way in which those beliefs are expressed.\textsuperscript{40}

However, the court acknowledged that not every act which is in some way inspired, motivated or influenced by it constitutes a “manifestation” of the belief. Thus, for example, acts or omissions which do not directly express the belief concerned or which are only remotely connected to a precept of faith fall outside the protection of Article 9, § 1. In order to count as a “manifestation” within the meaning of Article 9, the act in question must be intimately linked to the religion or belief. An example would be an act of worship or devotion which forms part of the practice of a religion or belief in a generally recognized form. However, the manifestation of religion or belief is not limited to such acts; the existence of a sufficiently close and direct nexus between the act and the


\textsuperscript{39} Id. at para. 95.

\textsuperscript{40} Id. at para. 81.
underlying belief must be determined on the facts of each case. In particular, there is no requirement on the applicant to establish that he or she acted in fulfilment of a duty mandated by the religion in question.\textsuperscript{41}

The court considered that Ms. Eweida’s insistence on wearing a cross visibly at work was a manifestation of her religious belief and that “the domestic authorities failed sufficiently to protect the first applicant’s right to manifest her religion.”\textsuperscript{42} This was the case notwithstanding that there was no evidence that any other BA employee wished to manifest his or her religion in this way.\textsuperscript{43}

The court found that BA’s uniform policy pursued a legitimate aim “to communicate a certain image of the company and to promote recognition of its brand and staff.”\textsuperscript{44} However, it noted that Ms. Eweida’s desire to manifest her religious belief was a fundamental right “because a healthy democratic society needs to tolerate and sustain pluralism and diversity; but also because of the value to an individual who has made religion a central tenet of his or her life to be able to communicate that belief to others.”\textsuperscript{45} The court found that the domestic courts had given too much weight to BA’s desire to maintain a certain corporate image, especially since Ms. Eweida’s cross was discrete and did not detract from her appearance, and there was no evidence that it would have impacted BA’s brand or image.\textsuperscript{46} Moreover, BA was able to amend its uniform policy to allow for the wearing of religious symbolic jewelry, which demonstrated that the prohibition was not fundamentally important.\textsuperscript{47}

The European Court’s finding that Ms. Eweida’s rights were breached even though no other BA employee had been shown to have been affected by the rule is the correct approach. A particular method of manifesting a belief does not need to be

\textsuperscript{41} Id. at para. 82 (citations omitted).
\textsuperscript{42} Id. at para. 95.
\textsuperscript{43} Id. at para. 94.
\textsuperscript{44} Id. at para. 93.
\textsuperscript{45} Id. at para. 94.
\textsuperscript{46} Id.
\textsuperscript{47} Id.
widely shared to be worthy of protection.

C. Beliefs Held by Few People: The Approach of the U.K. Tax Tribunal

The European Court of Human Rights’ decision in *Eweida* was applied by the first-tier Tribunal (Tax Chamber) in *Blackburn v. Revenue and Customs Commissioners*.\(^{48}\) The claimants were members of the Seventh-day Adventist Church who were refused an exemption from a requirement to file VAT returns online on religious grounds.\(^{49}\) The judge found if he had to make a decision purely using the normal rules of construction, without reference to the Human Rights Act 1998, he would have found that the claimants were *not* entitled to an exemption.\(^{50}\) While the claimants were members of the Seventh-day Adventist Church, the Church did not consider its beliefs to be incompatible with the use of electronic communications.\(^{51}\) Indeed, the claimants did not object to the use of all electronic communications, but just to the use of computers, the internet, television, and mobile phones.\(^{52}\) However, the judge reached a different conclusion in light of the claimants’ rights under Article 9.\(^{53}\)

Continuing the reasoning applied in *Eweida*, the judge took a broad approach to the assessment of whether the claimant’s Article 9 rights were violated and found that the claimants were manifesting their religious beliefs through their refusal to use computers.\(^{54}\) While the Revenue and Customs Commissioners did accept that the claimants’ beliefs attained the necessary “level of cogency, seriousness, cohesion, and importance to obtain


\(^{49}\) Id. at [12], [16].

\(^{50}\) Id. at [33].

\(^{51}\) Id. at [12].

\(^{52}\) Id.

\(^{53}\) Id. at [44]–[62].

\(^{54}\) Id. at [52].
protection” under Article 9\textsuperscript{55} they still argued that there was not a “sufficiently close and direct nexus between the act and the underlying belief.”\textsuperscript{56} The judge rejected this argument saying:

Indeed, as [the claimant] explained it, in shunning computers he and his wife are acting in what they see as fulfilment of a duty mandated by their religion, in that he and his wife believe that they must act in accordance with their conscience in order to be judged righteous at the second coming. And their conscience dictated that they shun computers. In this, therefore, it is apparent to me the manifestation of their religious beliefs in shunning computers is acting in fulfilment of a duty mandated by their religion as they perceive it to be. This is clearly within the meaning of “manifestation” in Article 9 as explained by the ECHR in [\textit{Eweida v. United Kingdom}].\textsuperscript{57}

The judge found that the requirement to file VAT returns online was in fact a restriction on the claimants’ rights under Article 9 and that there was no justification for the restriction.\textsuperscript{58} These cases demonstrate the effectiveness of a practical human rights based approach to the protection of religious and philosophical beliefs. The courts correctly acknowledge that it would be inappropriate to adopt a narrow definition of “belief” in order to exclude protection of certain groups. Furthermore, the courts recognize that a measure does not need to affect a wide group of people who share the same beliefs in order to infringe on a person’s right to express his or her belief; even beliefs held by a small minority are worthy of protection.

II. HOW FAR MUST WE GO TO ACCOMMODATE THE MANIFESTATION OF BELIEFS?

Real respect for freedom of thought, conscience, and belief

\textsuperscript{55} \textit{Id.} at [50].
\textsuperscript{56} \textit{Id.}
\textsuperscript{57} \textit{Id.} at [51]–[52].
\textsuperscript{58} \textit{Id.} at [59]–[62].
requires that we be slow to interfere, doing so only when such intervention is necessary and proportionate to protecting the rights and freedoms of others. Generally, it is easier to justify intervention in the context of young children than with adults. It is also easier to justify intervention in the context of employment when a public official, in particular, cannot practically perform his or her reasonable duties or refuses to apply the law of the land and the principle of non-discrimination to those that he or she serves. Intervention is also easier to justify with regard to the provision of goods and services when those engaged in commercial activity seek to discriminate when deciding who they will and will not serve.

A. The Religious Beliefs of Public Officials

The question of how to deal with religious beliefs that are discriminatory in nature has recently come to the forefront in a number of important cases involving public officials. One such case involved Ms. Ladele, a Christian registrar in Islington Council who said that she could not conduct Civil Partnerships because it would involve her participation in creating a union that was “contrary to God’s laws.” Although an Employment Tribunal originally upheld Ms. Ladele’s claim that she had been directly and indirectly discriminated against on grounds of her religion, that decision was reversed by the EAT. Ms. Ladele’s appeal was dismissed by the Court of Appeal, since the local authority was pursuing the legitimate aim of providing effective service by requiring Ms. Ladele to be designated as a registrar for civil partnerships. The court also found that the local authority was complying with its overarching policy of being committed to the promotion of equal opportunities, which required its employees to act in a way that does not discriminate against others.

60 Id. at [3]. It is believed that Liberty was the first NGO to ever intervene in the Tribunal in the public interest.
61 Id.
62 Id. at [40].
Ms. Ladele complained to the European Court of Human Rights, which found that there had been no breach of Ms. Ladele’s rights under Article 14 in conjunction with Article 9. Unfortunately, while the European Court upheld the U.K.’s position, it found that the issue fell within the Contracting States’ margin of appreciation. The court noted that the consequences for Ms. Ladele were particularly serious as her refusal to be designated as a civil partnership registrar resulted in her facing disciplinary action and losing her job. However, the national authorities were pursuing a legitimate aim and they had not exceeded the wide margin of appreciation that the court generally allows national authorities when balancing competing rights. The local authority had offered Ms. Ladele a compromise whereby she would be required to carry out straightforward signings of the civil partnership register and administrative work in connection with civil partnerships, but she would not be required to conduct ceremonies.

This case is perhaps a paradigm of a justified interference with someone’s expression of his or her religion. Ms. Ladele was a public official who would not carry out functions which she thought conflicted with her beliefs, notwithstanding that those functions had been introduced by a democratically elected Parliament, and the refusal to do the work amounted to unlawful

---

64 Article 14 of the European Convention on Human Rights states: The enjoyment of the rights and freedoms set forth in this Convention shall be secured without discrimination on any ground such as sex, race, colour, language, religion, political or other opinion, national or social origin, association with a national minority, property, birth or other status.
65 Convention for the Protection of Human Rights and Fundamental Freedoms, art. 14, Nov. 4, 1950, 213 U.N.T.S. 230 [hereinafter European Convention on Human Rights]. Article 14 does not provide a freestanding right not to be discriminated against, and in order to rely on Article 14, a claimant must be able to show that another of their rights under the Convention is engaged.
67 Id. at para. 102.
68 Id. at para. 106.
69 Id. at para. 26.
discrimination. Islington’s stance was not based on practicality—it could have provided the civil partnerships service without her—but was a matter of principle. The local authority could not be seen as condoning unlawful discrimination.

It would be nonsense if Islington were obliged to accommodate Ms. Ladele’s belief on the one hand, and have a duty not to discriminate on grounds of sexual orientation on the other. Discrimination on grounds of sexual orientation is now unlawful in the U.K. and is treated equally with discrimination based on race, sex, and indeed religious discrimination. It would undermine the whole system of equality protection if public officials were allowed to engage in what would otherwise be unlawful discrimination because of their personal beliefs.

B. The Religious Beliefs of Those Providing Goods and Services to the Public

In a recent case the U.K. Supreme Court had to decide the issue of how far we should go to accommodate religious belief. The case of Bull v. Hall69 involved Christian bed and breakfast owners who turned away a gay couple—who had booked a double room—because of a sincerely held belief that sexual intercourse outside of traditional marriage is sinful.70 The court used the broad approach to the assessment of Article 971 rights taken in Eweida and Blackburn, but while the court accepted that the right to manifest religious belief was clearly engaged, the reasoning in

70 Id. at [9]–[10].
71 Article 9 of the European Convention on Human Rights provides:
   1. Everyone has the right to freedom of thought, conscience and religion; this right includes freedom to change his religion or belief and freedom, either alone or in community with others and in public or private, to manifest his religion or belief, in worship, teaching, practice and observance.
   2. Freedom to manifest one’s religion or beliefs shall be subject only to such limitations as are prescribed by law and are necessary in a democratic society in the interests of public safety, for the protection of public order, health or morals, or for the protection of the rights and freedoms of others.

European Convention on Human Rights, supra note 64, art. 9.
Ladele prevailed.\textsuperscript{72} Supreme Court Justice Lady Hale, strongly made the case that the moniker of religious freedom did not sanction discrimination in the provision of goods and services: Homosexuals . . . were long denied the possibility of fulfilling themselves through relationships with others. This was an affront to their dignity as human beings which our law has now (some would say belatedly) recognised. Homosexuals can enjoy the same freedom and the same relationships as any others. But we should not underestimate the continuing legacy of those centuries of discrimination, persecution even, which is still going on in many parts of the world. It is no doubt for that reason that Strasbourg requires “very weighty reasons” to justify discrimination on the grounds of sexual orientation. It is for that reason that we should be slow to accept that prohibiting hotel keepers from discriminating against homosexuals is a disproportionate limitation on their right to manifest their religion.\textsuperscript{73}

This conclusion aligned with Liberty’s intervention before the U.K. Supreme Court in this case. Liberty argued that the better approach to balancing competing rights is to broadly read Article 9, treat the limitation as interference, and when it comes to the issue of justification, give significant weight to the importance of affording lesbians and gay men equality in accessing services and in the enjoyment of other social privileges. The court also made it clear that its decision was not a matter of preferring one protected characteristic or one set of rights to another. The result would have been the same if a gay hotel owner sought to turn away a Christian couple on the grounds of their beliefs.\textsuperscript{74}

When sexual orientation regulations first came into force in the U.K., there was considerable debate about whether exemptions should be allowed on religious grounds leading to some concessions. Religious bodies continue to be allowed to

\textsuperscript{72} Bull, [2013] U.K.SC 73.

\textsuperscript{73} Id. at [53].

\textsuperscript{74} Id. at [54].
discriminate on the grounds of sexual orientation in certain limited circumstances.\textsuperscript{75} Those lines are drawn by Parliament and those individuals who disagree with those lines should lobby for a change in the law. However, no individual can ask their employer or the courts to extend the scope of the exemptions. I accept that it is difficult for people who are in public or business roles to adapt to changes in the law with which they fundamentally disagree, but which also have a significant impact on how they conduct their role or business. But that’s what it means to live in a democracy and you either accept it or, if you feel that strongly about it, you should find another job. There comes a time when the pacifist has to leave the army rather than insist on his pacifism therein.

There will be many other cases that are not as clear cut as those described above, and it is the task of employers and courts to try to come to sensible conclusions. One such example is the case of the Christian bus driver who refused to drive buses carrying the slogan “There’s probably no god.”\textsuperscript{76} His employer recognized that this might be upsetting for him and agreed to try to put him on other routes, as long as this did not inconvenience other drivers.\textsuperscript{77} The driver accepted this and agreed that if it became impracticable to accommodate him, he would have to find another job.\textsuperscript{78} Sadly, those sorts of stories of tolerance and common sense are either rare, or more likely, rarely reported.

The cases of \textit{Ladele} and \textit{Bull v. Hall} demonstrate that interference with respect for religious freedom in the provision of public and business services can be justified. For example such interferences may be justified in particular security or safety scenarios where an item of clothing must be temporarily removed to allow for a respectful identity check at an airport or sterile

---

\textsuperscript{75} For example, the exemptions from employment equality legislation allow religious employers to discriminate against potential applicants for jobs on grounds of religion or belief and of sexual orientation, and to discriminate against current employees on those same grounds. The Equality Act, 2010, c. 15, § 196, sch. 23 (U.K.).


\textsuperscript{77} \textit{Id}.

\textsuperscript{78} \textit{Id}.
conditions in parts of a hospital. But the rights and freedoms of others, in my view, do not include protection from difference, irritation, and offense, as opposed to real harm, whether the individual concerned is in a religious, political, or other minority.

C. The Right to Wear Religious Clothing in Public: The French Approach

An example of grossly disproportionate interference with religious freedom is the recent introduction in France of a law banning the wearing of clothing designed to conceal the face in public spaces. The law imposes penalties on individuals who coerce others into wearing clothing that covers their face and on those wearing such clothing. There are only limited exceptions to the ban, for example: clothing permitted by law or on medical or other grounds, worn for sport, festivities, or artistic or traditional events. Although the law is framed in neutral terms, one effect is to prohibit the wearing of the burkha in public places, and it is a troubling example of the rising anti-Islamic sentiment in Europe.

The French law has been challenged in the European Court of Human Rights and was recently transferred to the Grand Chamber where judgment is pending. The challenge was brought by a devout Muslim French national who wears a burkha.

---

80 Id. arts. 3, 4.
81 Id. art. 1.
and niqab\textsuperscript{84} because of her faith, culture, and personal convictions, and who is not pressured to do so by her husband or her family. The applicant is happy not to wear the niqab in certain circumstances, but would like to have the option of wearing it in public. The applicant is relying on Articles 3 (prohibition of torture and inhuman and degrading treatment), 8 (right to respect for private and family life), 9 (right to freedom of thought, conscience and religion), 10 (right to freedom of expression), 11 (right to freedom of assembly and association), and 14 (prohibition on discrimination) of the European Convention on Human Rights.\textsuperscript{85} Liberty has intervened in the application and argued that the French ban is an unjustified interference with various human rights.\textsuperscript{86} Our submission was based on three propositions.

First, the law clearly interferes with freedom of religion. Following the decision of the European Court in \textit{Eweida},\textsuperscript{87} it is clear that it is not for a court to decide whether the applicant’s choice to wear a veil is a valid manifestation of her religion. Nor is it a court’s role to question the extent to which other members of the applicant’s religious group share her belief in the importance of wearing the burkha and niqab. In previous cases, the European court had dismissed claims for religious discrimination in the employment context because the applicants could choose to resign from their position if it conflicted with their religious beliefs.\textsuperscript{88} The court moved away from this approach in \textit{Eweida}, but even if it had not done so, these cases

\textsuperscript{84} \textit{Id.} There are many different recognized spellings of “niqab.” I will adopt this spelling throughout this article, except in quotations that use another recognized spelling.


\textsuperscript{86} Brief for Intervenor, S.A.S. v. France, at 5–7.


can be distinguished from the French ban—women who wish to wear a burkha in public do not have the option of resigning to avoid the impact of the law; it affects every aspect of their lives.

Second, the ban interferes with an applicant’s right with respect to her private and family life under Article 8 because it affects her ability to establish a social life and develop relationships with others. The ban also affects the applicant’s right, under Article 10, to express her faith by wearing a burkha.

Third, the effect of the law is discriminatory as it significantly disadvantages Muslim women who choose to wear the burkha. The French law appears to have three potential justifications: (1) it is contrary to Republican values for a person to be cut off from others; (2) there may be a danger to public safety; and (3) the wearing of the burkha is a public manifestation of a lack of equality between men and women. Liberty, however, argued that these interferences with individuals’ human rights cannot be justified.

On closer inspection, these justifications are flawed. Whilst secularism is an important value in France, the law specifically affects the wearing of the burkha but does not prevent people from wearing other religious dress or symbols. There is no sound reason for this difference in treatment. Security concerns also do not provide an answer. A requirement to remove a face covering in certain circumstances may be justified, but a complete ban on wearing the burkha is clearly disproportionate. Many women who wear the burkha would be willing to show their faces for identification purposes and it is not clear that the existing French laws regarding identity checks provide insufficient protection for public security. Finally, while promoting equality is an important aim, punishing women for expressing their faith does not support equality. A law based on clumsy assumptions about what drives a woman to wear the veil disregards her individual autonomy. By forcing women to comply with a particular notion of equality, the law undermines their dignity as women and as Muslims and has the effect of barring them from some public spaces altogether. 89

89 It is important to note that the applicants in these cases do not consider themselves to be pressured into wearing the burkha and the niqab, but rather, it is an expression of their religious faith.
D. The Right to Wear Religious Clothing in Public: the U.K. Approach

The issues surrounding the wearing of religious clothing in public are not limited to France and have been considered in the U.K. One example is the case of Azmi, which concerned a classroom assistant who was not allowed to wear a niqab.\(^\text{90}\) Although the school’s decision to refuse to allow Azmi to wear the niqab was ultimately found to be justified, the EAT rigorously scrutinized the school’s reasons.\(^\text{91}\) This was an unusual case because a religious dress requirement arguably did have a negative impact on others. Ms. Azmi’s job primarily involved language support for pupils for whom English was not their first language. She was permitted to wear the niqab outside the classroom but not while teaching. General research and observation of her teaching showed that language support could be carried out more effectively if her face was visible. On that basis, the EAT decided that the school’s approach was not unlawful.\(^\text{92}\)

A similar issue was considered by the House of Lords, formerly the U.K.’s highest court, in the cases of \(R\) (\(SB\)) v. Head Teacher and Governors of Denbigh High School\(^\text{93}\) and \(R\) (\(X\)) v. Head Teacher and Governors of \(Y\) School.\(^\text{94}\) These cases concerned claims by Muslim girls who asserted a right to wear a jilbab and a niqab, respectively, at school. The court dismissed these claims on the basis that there was no interference with the girls’ right to freedom of religion under Article 9 because they could have gone to schools that would have allowed them to wear


\(^{91}\) Id. at [62]–[74].

\(^{92}\) Id. at [66], [80]. One interesting aspect of the case was that Ms. Azmi suggested that the situation could be resolved by isolating her from male teachers. The school refused to do this, which was surely right because it could have led to claims of direct sex discrimination by male teachers.


the religious garments. 95

A more recent example involves the issue of whether a defendant charged with witness intimidation should be allowed to wear the niqab during a trial in the Crown Court. 96 In his judgment on September 16, 2013, H.H. Judge Peter Murphy set out general principles on when defendants in the Crown Court should be allowed to wear clothing that covers their face. 97 The judge gave detailed consideration to the human rights issues involved. He noted the importance of the right to freedom of religion, but stated that the corollary of this right is a duty to respect legal institutions and a court’s rules and practices. The judge also considered the fundamental requirements of an adversarial trial and the need for the court to be able to judge the defendant’s reaction and to prevent the defendant from being immunized from effective cross-examination. Furthermore, in order to protect the administration of justice, the court—and not the defendant—must be in control of its procedure. 98

The judge recognized the importance of wearing a niqab to many Muslim women. He said:

I also recognise the intrinsic merit which the niqab has in the eyes of women who wear it. I reject the view, which has its adherents among the public and the press, that the niqab is somehow incompatible with participation in public life in England and Wales; or is nothing more than a form of abuse, imposed under the guise of religion, on women by men. There may be individual cases where that is true. But the niqab is worn by choice by many spiritually-minded, thoughtful and intelligent women, who do not

97 Id.
98 Id.
deserve to be demeaned by superficial and uninformed criticisms of their choice. The Court must consider the potential positive benefits of the niqaab.\textsuperscript{99}

In the end, the judge conducted a balancing exercise between the defendant’s right to freedom of religion and the rights of others involved in the trial such as the victims, the jurors, and the rights of the public generally. He concluded that it would be appropriate to have some restrictions on when a niqab could be worn during the trial and set out principles on how the issues should be dealt with. For example, a female officer could be asked to confirm the defendant’s identity to the court, and while the defendant would have to remove the niqab to give evidence, she could give evidence behind a screen or by video link so that she could not be seen by the general public.

The H.H. Judge Peter Murphy found that restrictions on the niqab in court furthered the legitimate aim of protecting the fair and effective running of the criminal courts. He also held that some restrictions on the defendant’s right to freedom of religion were necessary and proportionate to uphold the rule of law in a democratic society.\textsuperscript{100} This judgment is an example of a court adopting a nuanced, principled, and practical approach that respects a defendant’s religious convictions while protecting the administration of justice. This can be contrasted with the blanket ban in France on wearing the burkha or niqab in public, which is dismissive of an individual’s right to express his or her religious convictions.

Unfortunately, not everybody in the U.K. adopts such a sensible approach to this issue. There have been calls for the U.K. to introduce a ban—similar to the one introduced in France\textsuperscript{101}—on wearing a burkha in public. A Conservative Member of Parliament, Philip Hollobone, has introduced a private member’s bill, the Face Coverings (Prohibition) Bill, which would make it an offense for a person to wear a garment with the primary purpose of obscuring one’s face in a public

\textsuperscript{99} Id. at para. 67.

\textsuperscript{100} Id. at paras. 81–85.

\textsuperscript{101} See Law 2010-1192 of October 11, 2010 (Fr.), supra note 79.
place.\textsuperscript{102} While the U.K. government does not support the Bill, and it currently has little prospect of success, its mere introduction demonstrates the rising tide of intolerance that is sweeping across Europe.\textsuperscript{103} This trend is further evidenced by a recent YouGov poll conducted in the U.K. in September 2013 that showed that 61\% of British adults agreed with the statement: “the burka should be banned in Britain.”\textsuperscript{104} It is to be welcomed, therefore, that the European Court of Human Rights has been developing a more thoughtful approach to religious freedom in its recent cases.

\textit{E. The Approach of the European Court of Human Rights to Religious Freedom}

The European Court’s approach, which is perhaps inevitable for an international court grappling with such diverse national traditions, initially seems to favor secularism. There are a number of cases in which the court said that a person is entitled to his or her beliefs, but there are limitations on his or her right to express those beliefs in the public sphere. Following this reasoning the court found no interference with Article 9 in the case of a woman who was refused permission to graduate from university unless she was prepared to be photographed without a headscarf,\textsuperscript{105} or in the case of a teacher who was not allowed time off to attend religious worship on a Friday.\textsuperscript{106} Even when the court did find that there was interference, it was often accepted that the restrictions were justified. One example is the court’s refusal to hear a complaint about a requirement mandating turban

\textsuperscript{102} Face Coverings (Prohibition) Bill, 2013-14, H.C. Bill [31] (Eng.).
removal during airport security screening.\textsuperscript{107}

The European Court’s approach of favoring secularism is best demonstrated by the case of \textit{Dahlab v. Switzerland}.\textsuperscript{108} In this case the European Court found that a refusal to allow a primary school teacher to wear the hijab (not the niqab, just the headscarf) was justified in view of the “powerful external symbol” that wearing a headscarf represented, specifically that the hijab could be seen as having a kind of proselytizing effect since it appeared to be imposed on women by a religious precept that was hard to reconcile with the principle of gender equality.\textsuperscript{109} The court found that wearing the hijab undermined the message of tolerance, respect for others, and equality and non-discrimination that all teachers in a democratic society must convey to their pupils.\textsuperscript{110}

However, in more recent cases the European Court has begun to show a greater tolerance for religion in the public sphere. A recent example is the case of \textit{Lautsi v. Italy},\textsuperscript{111} which concerned a state school in Italy that had a crucifix fixed to the wall in each of its classrooms.\textsuperscript{112} The applicant wanted to give her two children, who attended the school, a secular upbringing, and thought that the crucifix displays interfered with that goal.\textsuperscript{113} She claimed that the crucifix presentation breached her right under Article 2 of Protocol No. 1\textsuperscript{114} to educate her children in accordance with her religious and philosophical beliefs. She also claimed that it breached her right to freedom of religion under Article 9 and was discriminatory and contrary to Article 14.\textsuperscript{115}

In 2009, a chamber of the court adopted the secularist


\textsuperscript{109} \textit{Id.} at 450.

\textsuperscript{110} \textit{Id.}


\textsuperscript{112} \textit{Id.} at para. 11.

\textsuperscript{113} \textit{Id.} at para. 12.

\textsuperscript{114} European Convention on Human Rights, \textit{supra} note 64, art. 2 (“No person shall be denied the right to education. In the exercise of any functions which it assumes in relation to education and to teaching, the State shall respect the right of parents to ensure such education and teaching in conformity with their own religions and philosophical convictions.”).

\textsuperscript{115} Lautsi, 54 Eur. Ct. H.R. at para. 29.
approach and found that there had been a breach of Article 2 of Protocol No. 1 and Article 9. The court found that the state had “an obligation to refrain from imposing beliefs, even indirectly, in places where persons are dependent on it or in places where they were particularly vulnerable.” It noted that in countries where the majority of the population is members of one religion, the use of the symbols of that religion without restriction as to place and manner could constitute pressure on students who do not practice that religion. The court found that while the crucifix had a number of meanings, the predominant meaning was a religious one. The crucifixes could be considered “powerful external symbols” and could be emotionally disturbing for children of other religions or those who were not religious at all. The court could not see “how the display of a symbol that it is reasonable to associate with Catholicism . . . could serve the educational pluralism which is essential for the preservation of ‘democratic society’ . . .”

This decision was reviewed by the Grand Chamber, which came to the conclusion that there had been no violation of Article 2 of Protocol No. 1 or of Article 9. The Grand Chamber acknowledged that states are responsible for ensuring neutral and impartial exercise of various religions, faiths, and beliefs. It also noted though that states are not prohibited from imparting religious or philosophical knowledge either directly or indirectly. In addition, the aim of Article 2 of Protocol No. 1 is to safeguard pluralism in education and to ensure that knowledge is conveyed in “an objective, critical, and pluralistic manner,

116 Id. at para. 30.
117 Id. at para. 31.
118 Id.
119 Id.
120 Id. at para. 73.
121 Id. at para. 31.
122 The Grand Chamber is made up of seventeen judges: the court’s President and Vice-Presidents, the Section Presidents and the national judge, together with other judges selected by drawing of lots. The judgment of the Grand Chamber is final.
enabling pupils to develop a critical mind.”124

The Grand Chamber thus held that the decision to display crucifixes in state schools fell within the state’s “margin of appreciation and therefore was allowed.”125 The court said that they had a duty to respect states’ decisions relating to the organization of the school environment, and the setting and planning of the curriculum, provided that they did not lead to a form of indoctrination.126 Since the crucifix is essentially a passive symbol, its display alone is insufficient to denote a process of indoctrination and did not have the same effect as “didactic speech or participation in religious activities.”127

The court’s softer approach is also demonstrated by the cases of Eweida and Ladele, discussed above, as well as Chaplin, and McFarlane, which were all heard simultaneously by the European Court of Human Rights.128 In its decision, the European Court noted that previously it had held that the possibility of resigning from a job meant that there was no interference with the employee’s religious freedom.129 However, it suggested that those decisions were not consistent with the court’s approach to other rights, such as the right to respect for private life under Article 8, or the right to freedom of expression under Article 10, and said:

Given the importance in a democratic society of freedom of religion, the Court considers that, where an individual complains of a restriction on freedom of religion in the workplace, rather than holding that the possibility of changing jobs would negate any interference with the right, the better approach would be to weigh the possibility in the overall balance when considering whether or not the restriction was proportionate.130

It is therefore clear that, from the perspective of the

124 Id. at para. 62.
125 Id. at para. 70.
126 Id.
127 Id. at para. 72.
129 Id. at para. 83.
130 Id.
Strasbourg jurisprudence, employers with policies that restrict their employees’ ability to manifest their religious beliefs will potentially be interfering with their employees’ rights under Article 9. The employer must then demonstrate that its policies are justified. The court applied this approach to the individual circumstances of the four claimants with differing results. The facts of *Eweida* and *Ladele* have already been discussed above. The facts of the other two cases, *Chaplin* and *McFarlane*, also illustrate the court’s approach.

Ms. Chaplin was a nurse on a geriatric ward who wished to wear a cross on a chain around her neck; however, this was contrary to the ward’s uniform policy. Her managers believed that there was a risk of injury if one of the patients pulled the chain or if it swung forward and came into contact with an open wound. The court found that here the hospital’s goal of protecting health and safety was more important than British Airways’ goal of protecting its corporate image. In this instance, the court gave the domestic authorities a wide margin of appreciation since the hospital managers were best placed to make decisions about clinical safety.

Mr. McFarlane was a counselor who, because of his orthodox Christian beliefs, refused to provide psycho-sexual counseling to same-sex couples. This breached his employer’s policy that required employees to provide services equally to heterosexual and homosexual couples and McFarlane was let go. The court did note that the loss of Mr. McFarlane’s job was a serious sanction. However, when Mr. McFarlane had begun his training course, he was aware of his employer’s equal opportunities policy and that he would not be able to filter clients on the ground of sexual orientation. The most important factor for the court was that the employer’s action was intended to

---

131 See *supra* Part I.A.
133 *Id.*
136 *Id.*
secure its policy of providing services without discrimination. The state authorities therefore had a wide margin of appreciation.

These cases provide grounds for optimism that, when the Grand Chamber gives its judgment in *S.A.S. v. France*, it will take the opportunity to build on its recent jurisprudence in *Eweida* and *Lautsi*.

**CONCLUSION**

I have sympathy with a human rights-based approach to grappling with discrimination arguments, particularly in the context of belief. Domestic law governing faith and belief in the U.K. and all sensible workplace policies should be applied with the fundamental right to freedom of thought, conscience, and religion in mind.

I advocate for a broad and generous approach to what is considered a “protected religion or belief,” such as the approach adopted by the U.K. EAT. This approach minimizes unattractive, divisive, and counterproductive arguments about which personal beliefs are worthy and unworthy of protection per se.

Having adopted a broad approach to what constitutes a protected religion or belief, it is then necessary to accommodate the manifestation of those beliefs and to only interfere with them when it is necessary and proportionate to do so in order to protect the rights of others. Deciding what is and is not a proportionate interference or unreasonable accommodation can of course be a tricky task in the workplace and the public sphere. There will of course be situations in which it is appropriate to interfere with those rights, the cases of *Ladele* and *Bull v. Hall* being obvious examples. However, the best discipline comes from testing alternative scenarios with the principle of non-discrimination and equal treatment itself. Would British Airways have banned the wearing of a headscarf or turban amongst its workforce? On the evidence, patently not. Should a council accommodate a registrar who refuses to officiate over mixed-race weddings? Is an atheist who believes that Christian doctrine is counter to the “laws of

---

137 *Id.* at paras. 108–10.
physics” best qualified to be a minister of that religion? I suspect most of us would answer my last two questions in the negative.

It is encouraging to see the European Court of Human Rights moving away from a secularist approach and adopting a more balanced approach to issues of religious freedom that gives appropriate weight to individuals’ religious convictions. However, it is unfortunate that some national governments have moved the other way, as is demonstrated by the French ban on the wearing of the burkha or niqab in public, which undermines many Muslim women’s rights to express their religious beliefs.

Nobody ever said that life in a rich, diverse democracy was easy, or that the public sphere and workplace wouldn’t be a place of occasional tension and strife. Our human rights framework offers a robust tool for negotiating the limits of otherwise vague terms like “tolerance” and “cohesion.” Inevitably, the laws that afford some protection to those whose beliefs are irritating, or even offensive to us, protect us as well. To quote St. Matthew’s Gospel: “Judge not, that ye be not judged. For with what judgment ye judge, ye shall be judged: and with what measure ye mete, it shall be measured to you again.”139 Or, if you prefer a secular Matthew, try the words of fictional Congressman Matt Santos from the Gospel according to Aaron Sorkin’s The West Wing:

The framers of our Constitution believed that if the people were to be sovereign and belong to different religions at the same time then our official religion would have to be no religion at all. It was a bold experiment then as it is now. It wasn’t meant to make us comfortable, it was meant to make us free.140

139 Matthew 7:1–2.
140 The West Wing: Mr. Frost (NBC television broadcast Oct. 16, 2005).
FACE VEIL BANS IN THE EUROPEAN COURT OF HUMAN RIGHTS: THE IMPORTANCE OF EMPIRICAL FINDINGS

Eva Brems*

INTRODUCTION

European societies’ recent struggle with the integration of their Muslim minorities has resulted in many challenging legal debates, particularly with regard to the accommodation of religion in the workplace and in educational settings. Recently, such debates have extended to the proper role of religious expression in the public space. The most widespread example of this new phenomenon is the criminalization of the wearing of the niqab, or Islamic face veil, in public.

One of the most remarkable aspects pertaining to the European bans on face coverings and the surrounding debates is that they proceed on the basis of assumptions about women wearing face veils without any factual support. At the time the bans in Belgium and France were adopted, there was no empirical research available that documented the experiences and motives of the women who wore face veils. Nor was there any effort undertaken to consult those women in the process leading up to the ban.

One such example is the report by the Parliamentary Commission of Inquiry in France before the ban on face veils was adopted.¹ The Commission of Inquiry consisted of 32 members

* Eva Brems is a Professor of Human Rights Law at Ghent University, Belgium, and was a member of the Belgian Federal Parliament (June 2010–May 2014). The research for this paper was possible thanks to the European Research Council (Starting Grant).

¹ A. GÉRIN, ASSEMBLÉE NATIONALE N° 2262, RAPPORT D’INFORMATION FAIT EN APPLICATION DE L’ARTICLE 145 DU RÈGLEMENT AU NOM DE LA MISSION D’INFORMATION SUR LA PRATIQUE DU PORT DU VOILE INTÉGRAL SUR
representing all parliamentary groups. It heard about 200 witnesses and experts and sent out questionnaires to several French Embassies. After six months, it produced a 658-page report. However, the commission had not planned to hear a single woman who actually wore a face veil. The only person they did interview who wore a face veil, Kenza Drider, was only heard upon her own request. In the Belgian parliament, a large majority of legislators rejected a request for expert hearings as well as a referral of the bill for advice to the Council of State, the state body that controls amongst others the conformity of proposed new legislation with higher law, such as constitutional and European human rights provisions. Today, however, such qualitative research on the experiences of women who wear the face veil exists. Specifically it has been conducted in France and Belgium, as well as in the Netherlands, the United Kingdom, and Denmark. It is worth noting that the findings of these studies are very similar. While the data on which this paper relies are mostly from my own study in Belgium, it needs to be

LE TERRITOLE NATIONAL (2010).

2 Id.
3 Id.
6 See Naima Bouteldja, “France vs. England,” in THE EXPERIENCES OF FACE VEIL WEARERS, supra note 5.
7 UNIV. OF COPENHAGEN, RAPPORT OM BRUGEN AF NIQAB OG BURKA (2009), available at http://www.e-pages.dk/ku/322/. See also Kate Østergaard et al., Niqabis in Denmark: When Politicians ask for a Qualitative and Quantitative Profiling of a Very Small and Elusive Sub-Culture, in THE EXPERIENCES OF FACE VEIL WEARERS, supra note 5.
8 Unless mentioned otherwise, quotes from niqabis (i.e. face veil wearers) are from EVA BREMS ET AL., HUMAN RIGHTS CENTRE OF GHENT UNIVERSITY, WEARING THE FACE VEIL IN BELGIUM (2012) [hereinafter BREMS ET AL., WEARING THE FACE VEIL], available at http://www.ugent.be/re/publiekrecht/en/research/human-
emphasized that these findings are entirely consistent with those of the French study.

A challenge to the French face veil ban is currently pending before the European Court of Human Rights, the case of \textit{S.A.S. v. France}. This paper assesses what chance the applicant has of succeeding. It argues that a crucial factor will be the extent to which the European judges will be willing to base their reasoning on empirical findings regarding face veils in Europe, rather than on prevalent myths embraced by European audiences and policymakers. This Article will also argue that these empirical findings are crucial for an adequate legal analysis of the human rights dimension involved in face veil bans.

First, this Article will set out the facts and context of face veil bans in Europe and the legal challenges surrounding them. Then, the Article will explain the legal criteria that will be used by the European Court of Human Rights when deciding on this issue. Next, in its central argument, this Article will discuss the possible outcome of \textit{S.A.S. v. France}, by assessing whether the arguments advanced by European governments to ban face veils can pass the human rights test instituted by the court. This assessment will rely on the court’s case law, as well as on the case file of the case currently pending before it. It will also include the results of empirical research conducted among women wearing face veils in Europe and analyze whether the government’s claims are consistent with those empirical studies.

\section{Face Veil Bans in Europe and the Human Rights Challenge}

Throughout (western) Europe, there is a trend to ban “face coverings” in public spaces, which targets women who wear the Islamic face veil.\footnote{Formally, these bans apply to “face covering” in general. Yet, both the parliamentary debates and the political discourse surrounding their adoption, as well as the practice of their implementation, indicate that in fact these bans target only Islamic face veils.} This phenomenon started with municipal bans; such bans are in place today in Belgium, the Netherlands, Italy,
and Spain.\textsuperscript{10} Nationwide bans were adopted in France in 2010\textsuperscript{11} and in Belgium in 2011,\textsuperscript{12} and a regional ban was voted into

\begin{quote}
10 In Belgium, the “geographical coverage” of these local prohibitions appears to be the widest, with virtually all major cities and towns disposing of a prohibition, which is regularly enforced (and continues to be enforced despite the nationwide ban, presumably on account of the lighter procedure of administrative sanctions). \textit{See, e.g.}, \textit{Belgium’s Lower House of Parliament Bans Burqa-type Islamic Dress in Public}, \textit{DAILY NEWS} (Apr. 30, 2010), http://www.nydailynews.com/news/world/belgium-house-parliament-bans-burqa-type-islamic-dress-public-article-1.169905. In the Netherlands, such local bans are quite rare. There, as the legality and constitutionality is widely considered controversial, they hardly seem to be enforced in practice. \textit{See, e.g.}, Ofrit Liviatan, \textit{From Abortion to Islam: The Changing Function of Law in Europe’s Cultural Debates}, 36 \textit{FORDHAM INT’L J.} 93, 103 (2013). In Italy, local bans can be found particularly in the north and northeast of the country. \textit{See, e.g.}, Evan Darwin Winet, \textit{Face Veil Bans and Anti-Mask Laws: State Interests and the Right to Cover the Face}, 35 \textit{HASTINGS INT’L & COMP. L. REV.} 217, 247 (2012). In Spain in 2010, a relatively small number of towns and cities in Catalonia (including, most notably, Barcelona), started to pass regulation banning face covering in municipal buildings. \textit{See, e.g.}, Natalie Orenstein, \textit{France Hardly Alone on Burqa Ban}, NEWSDesk (July 21, 2010), http://newsdesk.org/2010/07/21/france-hardly-alone-on-burqa-ban/.

11 Loi 2010-1192 du 11 octobre 2010 interdisant la dissimulation du visage dans l’espace public [Law 2010-1192 of October 11, 2010 on the Prohibition of Concealing the Face in Public Space], \textit{JOURNAL OFFICIEL DE LA REPUBLIQUE FRANÇAISE} [J.O.] [OFFICIAL GAZETTE OF FRANCE], Oct. 12, 2010, p. 18344. Article 1 states: “No one may, in spaces open to the public, wear a garment that has the effect of hiding the face.” Exceptions apply when “clothing is prescribed or authorised by legal or regulatory provisions,” when the clothing “is justified by reasons of health or professional motives,” or when the clothing is “part of sports activities, festivities or artistic or traditional manifestations.” \textit{See id.} art. 2, § II. Sanctions consist of fines for the wearer of up to 150 euros and/or participation in a citizenship course. Additionally, the Act penalizes anyone who forces another “through threats, violence, constraint, abuse of authority or power for reason of their gender” to wear face coverings, with a fine of 30,000 euros and one year imprisonment. \textit{Id.} art. 4. The latter penalties can be doubled if the victim is a minor. \textit{Id.} On October 7, 2010, the Constitutional Council upheld the constitutionality of the ban, with only minor reservations. \textit{Conseil constitutionnel} [CC] [Constitutional Court] decision No. 210-613DC, Oct. 7, 2010 (Fr.). Most notably the Council determined that the ban could not be enforced in places of worship. \textit{Id.} ¶ 5.

12 Loi visant à interdire le port de tout vêtement cachant totalement ou de manière principale le visage [Prohibition on Wearing Clothing Fully or Mostly Covering One’s Face] of June 1, 2011, \textit{MONTEUR BELGE} [M.B.] [OFFICIAL
effect by referendum in the Swiss canton of Ticino. Proposals for similar nationwide bans have been dismissed—at least temporarily—in Denmark, the Netherlands, Spain, the United

FACE VEIL BANS IN THE ECHR

Gazette of Belgium], July 13, 2011, http://www.staatsblad.be (Belg.). It inserts an Article 563bis into the Belgian Criminal Code. In practical terms and “subject to legal provisions to the contrary,” this provision punishes persons “who appear in places accessible to the public with their faces covered or concealed, in whole or in part, in such a manner that they are not recognisable” with a monetary fine of fifteen to twenty-five euros (increased with the legal surcharge factor, i.e., multiplied by 5.5) and/or a prison sentence of one to seven days. Id. An exception applies when face covering is permitted or imposed by “labour regulations or municipal ordinances due to festivities.” Id. Moreover, the law continues the application of local bans imposing administrative sanctions. In Belgium, too, the law was unsuccessfully challenged before the Constitutional Court, which like the French Constitutional Council, made only a minor reservation for places of worship. Cour Constitutionelle [CC] [Constitutional Court] decision no 145/2012, July 13, 2011, Moniteur Belge [M.B.] [Official Gazette of Belgium], Dec. 6, 2012, http://www.grondwettelijkhof.be (Belg.).

13 The referendum in September 2013 obtained a 65.4% majority. Gerhard Lob, Burka Ban Approved in Italian-Speaking Switzerland, Swissinfo.CH (Sept. 22, 2013), http://www.swissinfo.ch/ita/politica/Il_Ticino_met_tale_al_bando_il_burqa_nella_costituzione.html?cid=36936130. The federal parliament will have to rule on the constitutionality of the rule. Id.

14 In 2009, the Danish Minister of the Interior set up an ad hoc committee to study the desirability of banning face veils in public. They commissioned an empirical study, which showed that the number of face veil wearers in Denmark was very small, and that many were Danish converts. Subsequently, no ban was adopted. See Østergaard et al., supra note 7.

15 The Dutch government agreements of 2007 and 2010 announced the introduction of a face-covering ban. Such a bill was introduced in Parliament in early 2012. Yet, after the fall of the cabinet, the new coalition announced in its agreement only a set of functional face-covering bans (in the context of education, health care, and public transportation, as well as for access to government buildings), rather than a general ban. See Bruggen Slaan, Regeerakkoord (2012), available at http://www.rijksoverheid.nl/bestanden/documenten-en-publicaties/rapporten/2012/10/29/regeerakkoord/regeerakkoord.pdf.

In July 2010, Spain’s lower chamber of parliament rejected a bill to ban the wearing of face-covering garments in public. At the regional level, the Catalan Parliament rejected two motions aiming to introduce a face veil ban in public spaces presented by the Popular Party on July 1, 2010 in the Plenary, and on April 5, 2011, in the Commission on Welfare and Immigration. See Amnesty Int’l, Choice and Prejudice: Discrimination Against Muslims
judges have occasionally ruled that the application of a local ban on face veils violated fundamental rights. The French

---


19 The parliamentary Commission on Constitutional Affairs approved a bill on August 2, 2011 (“Divieto di Indossare gli Indumenti Denominati *Burqa* e *Niqab*”) that would prohibit persons from going in public wearing any garment covering the face, rendering it punishable with fines of 100 to 300 euros. *Legge 24 Ottobre 2011, n. 216/3, A.C. 627- A (It.)*, available at http://www.camera.it/701?leg=16&file=AC0378C.

20 In February 2013, the Supreme Court of Spain overturned a city authority ban in Catalonia on the basis that it limited religious freedom, and that the city lacked the authority to order such a prohibition. S.T.S., Feb. 6, 2013 (R.A.J., No. 4118/2011) (Spain). In Belgium, contradictory case law on the application of local bans to face veils was one of the reasons for the enactment of the general ban. See Politierechtbank [Pol.] [Police court], Jan. 26, 2011, (Belg.) (on file with author); Politierechtbank [Pol.] [Police court] Tongeren, June 12, 2006, no. 05A79 (Belg.) (on file with author) (finding no such violation). In Italy, two courts found that the 1975 Public Order Protection Act could not be regarded as grounds for municipalities prohibiting face veils in public space in general. See TAR Trieste, 16 Ottobre 2006, nr. 645; Diritto & Giustizia, n° 44, 2006, 111–13, Giurisprudenza di merito, n° 9, 2007; Giur. it. 2007, 2423 (It.); Cons. Stato, 19 Giugno 2008, no. 3076, available at http://religare-database.eu/PDF/PDFwp5Italy/ConsiglioBurqa2008.pdf; see also Mathias Möschel, *La Burqa en Italie: d’une Politique Locale à une Législation Nationale, in QUAND LA BURQA PASSE À...*
Conseil d’État even advised in a report against the adoption of the nationwide ban since “no incontestable legal basis” could be provided for such a general ban. 21 Similarly, the Dutch Council of State advised against the adoption of a face-covering ban.22 Such advice has been ignored. Both the French Conseil Constitutionnel23 and the Belgian Constitutional Court24 have validated the nationwide bans on face veils. Those courts held that such bans did not violate any fundamental rights, as protected in their respective constitutions and the European Convention on Human Rights (ECHR).

However, it is the European Court of Human Rights (“the court”) that will deliver the final word in this matter. As mentioned above, a challenge to the French face-covering ban in the case of S.A.S. v. France is currently pending before a Grand Chamber of the Court.25 A public hearing was held on November 27, 2013,26 and the court’s judgment is expected sometime this year.

The applicant in S.A.S. v. France is a French citizen born in Pakistan who lives in the Paris region.27 She is a law graduate

L’OUEST; ENJEUX ETHIQUES, POLITIQUES ET JURIDIQUES (David Koussens & Olivier Roy eds., 2013).


22 Parliamentary document, Feb. 3, 2012, TK 33165, at nr. 2, available at https://zoek.officielebekendmakingen.nl/kst-33165-2.pdf; the advice of the Council of State, Nov. 28, 2011, TK 33165, at nr. 4, available at https://zoek.officielebekendmakingen.nl/kst-33165-4.pdf. In the Netherlands, as in Belgium and France, one of the powers of the Council of State is to give advice on proposed legislation, concerning its legality and in particular its conformity with higher law. Such advice, however, is not binding.

23 See infra note 11 and accompanying text.

24 See infra note 12 and accompanying text.


who completed an internship with a law firm in Birmingham, with whom she submitted the case before the European Court of Human Rights.\(^{28}\) She stated that, before the ban, she had been wearing the face veil on a regular basis since she was 18 years old:

Gradually, I wore my full face veil whenever I passed through public areas, traveled on public transport or visited public buildings (generally three times a week) . . . . Of course, for instance, I would take off my veil if I needed to visit the doctor or keep an official appointment.\(^{29}\)

Since she submitted the application on the day the ban went into effect, at that point she had not yet been stopped by the police or fined for wearing her veil. However, in a witness statement submitted two years later, she discussed how the ban had negatively affected her life, stating that as a result of the implementation of Loi no. 2010-1192 I now live under the threat of both state prosecution and public persecution. As a result of the implementation of Loi no. 2010-1192 I am now vilified and attacked on the streets of the Republic I live, effectively reduced to house arrest, virtually ostracized from public life and marginalized.\(^{30}\)

In that same statement she continues to provide additional information about the negative impact the ban has had on her daily life:

criminalisation, or rather the political scaremongering that preceded it, has incited members of the public to now openly abuse and attack me whenever I drive wearing my veil. Pedestrians and other drivers routinely now spit on

\(^{28}\) Id. \(\S\) 15.

\(^{29}\) Id. \(\S\) 22–23. This is a correction to the statement in the application that “[t]he Applicant does not wear the niqab in public places at all times . . . . As to when the Applicant chooses to wear the niqab in the public place depends very much on her introspective mood, spiritual feelings and whether she wishes to focus on religious matters.” Application \(\S\) 3, S.A.S. v. France, Eur. Ct. H.R. (App. No. 43835/11).

\(^{30}\) Witness Statement of the Applicant, supra note 27, \(\S\) 6.
my car and shout sexual obscenities and religious bigotry. Consequently, I now feel like a prisoner in my own Republic, as I no longer feel able to leave my house unless it is essential. I leave the house less frequently as a result. I wear my veil with even less frequency when out in public as a result. Indeed, I also feel immense guilt that I am forced to no longer remain faithful to my core religious values.\footnote{Witness Statement of the Applicant, supra note 27, ¶ 26. A car on a public road is not considered part of the “public space” under the French ban, as per an interpretative circular: Circulaire du 2 mars 2011 relative à la mise en œuvre de la loi n° 2010-1192 du 11 octobre 2010 interdisant la dissimulation du visage dans l’espace public, JORF n°0052 (Mar. 3, 2011), at 4128.}

In addition, the applicant’s fear of harassment motivated her request for anonymity,\footnote{Application, supra note 29, ¶ 1.} which was granted by the court. Her testimony is consistent with the testimony of other individuals who wear face veils about the impact that the ban has had on their daily lives.\footnote{In particular, see OPEN SOC’Y JUSTICE INITIATIVE, AFTER THE BAN: THE EXPERIENCES OF 35 WOMEN OF THE FULL-FACE VEIL IN FRANCE (2013) [hereinafter AFTER THE BAN], which was submitted as additional evidence with the Third Party Intervention of Open Society Justice Initiative, S.A.S. v. France, App. No. 43835/11, (Eur. Ct. H.R. filed Oct. 2, 2013), available at http://www.opensocietyfoundations.org/sites/default/files/after-the-ban-experience-full-face-veil-france-20140210.pdf.}

However, in one respect, the applicant is not so representative. At the end of her witness statement, she put forward several “compromise proposals”:

Firstly, I would be willing to accept restrictions regarding the visibility of the veil’s material, i.e. to wear only veils that were diaphanous “see-through,” thus ensuring that my facial features remained essentially visible . . . . Secondly, I would be willing to accept exemptions enabling full face veils to be worn during the fasting periods and festivities of Ramadan and Eid.\footnote{Witness Statement of the Applicant, supra note 27, ¶ 33.}
It seems unlikely that other individuals who wear face veils in France or Belgium would agree with these proposals. Even though the willingness to compromise may be present (e.g., one Belgian respondent had approached her mayor with a proposal to avoid the color black for her face veil), these options seem to affect the essence of the women’s claims, and hence to go beyond a reasonable, compromised solution.

As the European Court of Human Rights is situated on French territory, the face-covering ban prevented the applicant from attending the hearing in her case. The applicant’s attorney informed the court that the applicant would like to attend, yet wished for “confirmation from the court that there will be provision for her to wear her full face veil during the proceedings including but not limited to transit to and from the court.” The reply stated that the court could not guarantee that the applicant would be able to wear the full face veil while traveling to and from the court. With regard to whether the applicant could wear her face veil during the hearing, the deputy Grand Chamber registrar who signed the reply wrote that:

the President has asked me to draw your attention to the fact that the applicant’s request confronts him with a question that is complex and sensitive, since it places the Court in a situation where the answer could be seen by the parties and external observers as an indication of an opinion on the merits of the issues to be examined at the hearing. He invites the applicant to contemplate her request in the light of the foregoing and to inform him . . . if she wishes to maintain it.36

35 Letter from Sanjeev Sharma, counsel at J.M. Wilson Solicitors LLP, to the president of the Grand Chamber of the European Court of Human Rights (Nov. 11, 2013) (on file with author) (“If she is not permitted to cover her face then her anonymity status becomes redundant. She is in a quandary. She does not wish to find herself in the position of having broken the law by attending Court yet she wishes to exercise her fundamental right to be present at her own hearing.”).

As a result of this reply, the applicant did not attend.

In this case, the court accepted third party interventions from the Belgian government, as well as from the Human Rights Centre of Ghent University, the NGO Liberty, Amnesty International, Open Society Justice Initiative (“Open Society”), and the group Article 19. Among the third party interveners, both Open Society and the Human Rights Centre of Ghent University submitted empirical data in addition to legal arguments. In April 2011, when the French ban went into effect, Open Society published data from interviews with 32 women wearing the face veil in France, during which it inquired into their motivations and experiences. This report was added to Open Society’s written submission and was referred to extensively in Liberty’s submission. In addition, and specifically in light of *S.A.S v. France*, Open Society conducted a follow-up

---


study, documenting the experiences of individuals who wore face veils after the ban. It received the court’s permission to add the report to its submission. In Belgium, the Human Rights Centre at Ghent University conducted similar research by interviewing 27 women, partly before and partly after Belgian’s ban on face veils went into effect. The argumentation in the Centre’s third party intervention relies strongly on that data. Such data are vital to understanding the legal rationale behind the applicant’s case.

II. RELEVANT PROVISIONS AND LEGAL TEST

The applicant in S.A.S. v. France alleges a violation of Articles 3, 8, 9, 10, 11, and 14 of the ECHR. Relying on

43 See BREMS ET AL., WEARING THE FACE VEIL, supra note 8; see also Eva Brems et al., The Belgian “Burqa Ban” Confronted With Insider Realities, in THE EXPERIENCES OF FACE VEIL WEARERS, supra note 5.

The relevant articles are as follows:

Article 3 ECHR:
No one shall be subjected to torture or to inhuman or degrading treatment or punishment.

Article 8 ECHR:
1. Everyone has the right to respect for his private and family life, his home and his correspondence.
2. There shall be no interference by a public authority with the exercise of this right except such as is in accordance with the law and is necessary in a democratic society in the interests of national security, public safety or the economic well-being of the country, for the prevention of disorder or crime, for the protection of health or morals, or for the protection of the rights and freedoms of others.

Article 9 ECHR:
1. Everyone has the right to freedom of thought, conscience and religion; this right includes freedom to change his religion or belief and freedom, either alone or in community with others and in public or private, to manifest his religion
or belief, in worship, teaching, practice and observance.
2. Freedom to manifest one’s religion or beliefs shall be subject only to such limitations as are prescribed by law and are necessary in a democratic society in the interests of public safety, for the protection of public order, health or morals, or for the protection of the rights and freedoms of others.

Article 10 ECHR:
1. Everyone has the right to freedom of expression. This right shall include freedom to hold opinions and to receive and impart information and ideas without interference by public authority and regardless of frontiers. This Article shall not prevent States from requiring the licensing of broadcasting, television or cinema enterprises.
2. The exercise of these freedoms, since it carries with it duties and responsibilities, may be subject to such formalities, conditions, restrictions or penalties as are prescribed by law and are necessary in a democratic society, in the interests of national security, territorial integrity or public safety, for the prevention of disorder or crime, for the protection of health or morals, for the protection of the reputation or rights of others, for preventing the disclosure of information received in confidence, or for maintaining the authority and impartiality of the judiciary.

Article 11 ECHR:
1. Everyone has the right to freedom of peaceful assembly and to freedom of association with others, including the right to form and to join trade unions for the protection of his interests.
2. No restrictions shall be placed on the exercise of these rights other than such as are prescribed by law and are necessary in a democratic society in the interests of national security or public safety, for the prevention of disorder or crime, for the protection of health or morals or for the protection of the rights and freedoms of others. This Article shall not prevent the imposition of lawful restrictions on the exercise of these rights by members of the armed forces, of the police or of the administration of the State.

Article 14 ECHR:
The enjoyment of the rights and freedoms set forth in this Convention shall be secured without discrimination on any ground such as sex, race, colour, language, religion, political or other opinion, national or social origin, association with a national minority, property, birth or other status.
Article 3, the prohibition of torture and inhuman and degrading treatment, the applicant contends that since it is illegal to wear a garment designed to conceal the face in public places, if she wears her face veil in public she risks incurring a criminal penalty, as well as suffering harassment and discrimination.\footnote{ECHR, Fifth Section, Questions for the Parties, S.A.S. v. France, App. No. 43835/11, (Eur. Ct. H.R.) (undated) (on file with author).} In addition, under Article 8, which protects the right to privacy in the home and family, the applicant complains that the ban, which prohibits her from dressing as she chooses in public, is a violation of her right to respect for private life.\footnote{ECHR, \textit{supra} note 44. Under the right to protection of private life, the European Court of Human Rights protects a wide range of autonomy-related interests, arguably including dress styles. The (former) European Commission of Human Rights (EComHR) has ruled, for instance, that constraints imposed on a person’s choice of mode of dress constitute an interference with private life as ensured by Article 8 of the Convention. \textit{See} Kara v. United Kingdom, Y.B. Eur. Conv. on H.R. (Oct. 22, 1998).} The applicant also claims that Article 9 of the ECHR, which protects manifestation of religion or belief, is violated by the ban on face veils. Effectively, she argues that her inability to wear the full veil in public places denies her the freedom to manifest her religion or belief.\footnote{ECHR, \textit{supra} note 44.} Additionally, she contends that the ban violates Article 10, the right to freedom of expression. Specifically, the ban prevents her from wearing in public a garment that expresses her faith, as well as her religious, cultural, and personal identity.\footnote{\textit{Id.}} Furthermore, she alleges that the ban violates Article 11’s freedom of assembly and association, since if she cannot wear her veil, she cannot go into public, and thus cannot associate with others.\footnote{\textit{Id.}} Finally, she contends that the ban ignores Article 14, which prohibits discrimination based on gender, race, language, religion, or any other status. In particular, the applicant complains that the face covering ban, by its very nature, engenders discrimination based on sex, religion, and ethnic origin against women who, like her, wear the full
It is important to begin with the understanding that Article 3 is an absolute right: once a certain treatment falls under its scope, it is automatically a violation, regardless of its justification. The main legal question is therefore one of a threshold of severity: can the treatment that the applicant complains of be labeled “inhuman” or “degrading?” The idea that the risk of a criminal penalty—in this case a fine—could be considered degrading or inhuman is farfetched. Yet, the claim that the ban exposes women who wear a face veil to harassment raises a relevant issue. The applicant in S.A.S. “believes that if she wears the niqab[,] members of the public . . . will request her without proper justification to remove it and will in the process harass and discriminate against her thereby exposing her to degrading treatment.”\(^{51}\) The research in France and Belgium demonstrates that women who wear a face veil in those countries suffer serious harassment. This research also strongly suggests that such harassment has increased as a result of the bans, with many citizens acting as vigilante policemen.\(^{52}\) It may legally be possible to hold the French government accountable under Article 3 for not adequately protecting women who wear a face veil. Yet the applicant is making a different point, namely that by introducing the face covering ban, the French state has implicitly, if not explicitly, encouraged aggression against women who wear face veils. Accepting this line of reasoning would be innovative. However, since there is almost no hard evidence linking the aggressions directly to the adoption of the law, it is unlikely that

\(^{50}\) Id.


\(^{52}\) BREMS ET AL., WEARING THE FACE VEIL, supra note 8, at 17–21. AFTER THE BAN, supra note 33, at 13, reports that the majority of interviewees experience verbal abuse on a regular basis, and that twelve respondents reported physical assaults, such as having their veil pulled off and being violently pushed or spat on. It notes that “the ban and public discourse seems to have implicitly legitimized the abusive treatment of veiled women. With a widespread condemnation of the full-face veil, the women’s testimonies reveal that some members of the public seem to think that the law allows for or legitimizes private enforcement.” Id.
the court will be persuaded by this argument.

Under all the other ECHR provisions that are invoked in this case, the court instead engages in a balancing exercise between the right that is at stake and the interest that is invoked by the government as a “legitimate aim” that may justify a proportionate restriction of that right. It is likely that the court will discuss the impact of the ban under Article 9, stating that its reasoning applies mutatis mutandis to the claims under Articles 853 and 10.54 The court may dismiss the claim under Article 11.55 But the court should address the discrimination claim under Article 14 in combination with one or more of the other Articles of ECHR separately.56 This is because the face covering bans are manifest examples of seemingly neutral legislation that is in fact targeted at a specific group, namely Muslim women who wear a face veil. The European Court of Human Rights recognizes that “a general policy or measure that has disproportionately prejudicial effects on a particular group may be considered discriminatory notwithstanding that it is not specifically aimed at that group.”57

53 The applicant argues applicability of Article 8 of the ECHR on multiple grounds: (1) the full face veil provides the person wearing it a sense of acute or extended privacy; (2) the matter relates to the applicant’s individual autonomy concerning her identity and dress code; (3) wearing the face veil is related to her ability to interact with others; (4) the exposure to public hostility infringes upon her physical and psychological integrity; and (5) the face veil is a cultural practice of a minority group. See Final Observations, S.A.S. v. France, App. No. 43835/11, ¶¶ 39–47 (Eur. Ct. H.R. July 4, 2013).

54 The relevance of freedom of expression for this case was particularly emphasized in the submission by Article 19 before the Grand Chamber, i.e., after the Fifth Sections “questions to the parties,” where it was not included. See Third Party Intervention Submissions by Article 19, supra note 40.

55 In this respect, it is noteworthy that before relinquishing jurisdiction to the Grand Chamber, the Fifth Section of the Court sent three questions to the parties, pertaining to their opinion on a positive violation of Articles 8, 9, and 14 only, the latter moreover being restricted to discrimination based on religion or sex. Exposé des faits et Questions aux parties, supra note 51.

56 Article 14 of ECHR prohibits discrimination in the exercise of any of the Convention rights, and hence has to be invoked together with another ECHR provision. See ECHR, supra note 44.

FACE VEIL BANS IN THE ECHR

i.e. the concept of indirect discrimination. This does not require proof of discriminatory intent even though in these cases this proof would not be hard to find.

It is clear that “the legislative history of the law demonstrates that the intent was to regulate the burqa and niqab, which were specifically identified as the target of the ban."58 The effective move towards a general ban on face covering in France began not long after MP André Gérin, along with others, filed a resolution on June 9, 2009 aimed at establishing a commission of inquiry concerning the face veil on French territory.59 Not long afterwards, President Nicolas Sarkozy, in a speech on June 22, 2009, stated that such veils were not welcome in France and that legislation was necessary “to protect women from being forced to cover their faces and to uphold France’s secular values.”60 That the ban is aimed at the face veil, despite its neutral language, is furthermore obvious on account of the constant references to the face veil throughout the parliamentary debates. Moreover, the opinion requested by the Prime Minister from the Conseil d’Etat in the early drafting stages concerned the “legal grounds for a ban on the full veil.”61 Clearly the ban was not about the visibility of faces in general, but specifically about the Islamic face veil. In that sense, the applicant is right to state that “this is a case where the discriminatory treatment comes very close to direct discrimination,”62 on grounds (e.g., sex,63 religion64) for which

58 Written Comments of the Open Society Justice Initiative, supra note 39, ¶ 6.
61 CONSEIL D’ETAT, SECTION DU RAPPORT ET DES ETUDES, supra note 21, at 7 (emphasis added).
62 Final Observations, supra note 53, ¶ 141.
the court exercises strict scrutiny.

Regardless, once the applicant has demonstrated that there is a difference in treatment or a disproportionate prejudicial effect, the burden shifts to the government to prove that the difference in treatment pursued a legitimate aim in a proportionate manner. At that point the review of any claim for discrimination would be examined under a similar analysis as that used for Articles 8, 9, 10, or 11 of the ECHR. Under this analysis, an interference with a right can only be justified if it has a legal basis, pursues a legitimate aim from among those listed in the restriction clause, and—most importantly—if there is a reasonable relationship of proportionality between the restrictive measure and that aim. Hence, whether the court conducts its analysis of the ban under the prohibition of discrimination or religious freedom, or both, the focus of the court’s reasoning will be on the examination of whether the ban can be considered proportionate to one or several legitimate aims.

According to the French Government in S.A.S., the “legitimate aim” underlying the face covering ban, involves three policy goals: (1) the protection of public safety; (2) respect for “compliance with the minimal requirements of life in society”; and (3) “equality between men and women and respect for the dignity of the person.”

III. UNCOVERING IN THE NAME OF “THE MINIMAL REQUIREMENTS OF LIFE IN SOCIETY?”

“Public safety” is explicitly mentioned in Article 9(2) as a

---


65 Several provisions of the ECHR are drafted in such a way (similar to other conventions such as the ICCPR), that the first paragraph sets out the scope of the right, whereas the second paragraph contains the conditions for its legitimate restriction. These restriction clauses set out three conditions: the restrictive measure should have a legal basis; it needs a legitimate aim from among those listed in the restriction clause; and it has to be “necessary in a democratic society,” i.e., proportionate with respect to the realization of that aim.

legitimate aim that may justify proportionate restrictions of religious freedom. Additionally, the protection of equality between men and women and of human dignity could also be a legitimate aim under Article 9 due to language regarding “the protection of the rights and freedoms of others.” Yet, it is not clear that respect for “the minimal requirements of life in society” fits under any of the “legitimate aims” under Article 9. The French Council of State was the first to suggest that this idea of “minimal requirements of life in society” could be legally translated into a novel interpretation of the concept of “public order,” building on the idea of “non-material public order.”

The Council of State noted that such a concept had not been developed in French legal doctrine or case law, and was not found in any neighbouring legal system either. Hence, the Council of State considered the concept of non-material public order vulnerable to constitutional challenges and advised against its use. However, French MPs borrowed the new concept, also naming it “social public order,” and built their case for a ban on face coverings largely on that ground. It is, however, far from certain whether the court will agree with this line of reasoning.

At the public hearing, one of the judges asked the representative of the French government how she viewed the policy goal of “minimal requirements of life in society” coexisting with the list of legitimate aims in the second paragraph of Article 9 of the ECHR. The representative’s answer referred to “the protection of the rights and freedoms of others.” This answer, however, supposes the existence of a right to see the face of others in a

---

67 See ECHR, supra note 44.
68 Id.
69 Id.
70 CONSEIL D’ETAT, SECTION DU RAPPORT ET DES ETUDES, supra note 21, at 26–27. On the concept of non-material public order in this context, see Rim-Sarah Alouane, Bas les Masques! Unveiling Muslim Women on Behalf of the Protection of Public Order: Reflections on the Legal Controversies Around a Novel Definition of “Public Order” Used to Ban Full-Face Covering in France, in THE EXPERIENCES OF FACE VEIL WEARERS, supra note 5.
71 This quote was taken from notes the author took during oral hearing of the Grand Chamber of the European Court of Human Rights in S.A.S. v. France in Strasbourg, France, on November 27, 2013.
public space. This is certainly not a legal right.

In the French debates surrounding the face-covering ban, seeing the face of others has been put forward as a moral right. This view is based on the work of the French sociologist Elisabeth Badinter, who was interviewed by the parliamentary commission of inquiry, and of the French philosopher Emmanuel Levinas. The latter’s discourse about the “face of the other,” as the basis for meeting another person and being morally involved, was interpreted literally (as referring to the actual face of a human being as the crucial building stone for moral relations in a society, rather than the presence of the other in our midst) by the Commission of Inquiry, and by the French government in its arguments in S.A.S. v. France. However, the applicant rightly alleges that

the government is treading on dangerous ground when it attempts to justify a legal measure by postmodern philosophy, which by its very nature is highly complex and not capable of clear-cut interpretations, let alone one “correct” interpretation. Arguably, the law prohibiting covering the face in public is not at all in line with the spirit of Levinas’ philosophy as his idea of “face-to-face encounter” is centred on inherent respect for the other—the opposite of what the law in question achieves.

72 GÉRIN, supra note 1, at 118.
74 GÉRIN, supra note 1, at 117–18.
76 Final Observations, supra note 53, ¶ 90. It is added, moreover, that [e]ven if one accepts a literalist interpretation of Levinas that it is crucial actually to see the other’s face for an ethical obligation towards her to arise, it is still absurd to conclude that that leads to the compulsory uncovering of the face, so
The level of subjectivity and cultural bias inherent in the “minimal requirements of life in society” argument becomes even more apparent in the submissions by the Belgian government, which advance the fact that it is considered respectful to take off sunglasses during conversation, and states without any reference to authority that appearing on the street implies a readiness to be looked at that is an inherent limit to the right to isolation.\textsuperscript{77} This that such an obligation towards the legally sanctioned person can be created. In other words, even if one accepts the idea that the face has special importance in the western tradition, it is a far cry to enforce uncovering the face by penal sanctions. If the Respondent followed its own logic through, it ought also to introduce a law making touching the other in the public space compulsory, as skin contact is also a feature of Levinasian discourse. The absurdity of the Respondent’s argument is evident and fails to take into account the cultural practices of minorities which does not necessarily subscribe to this philosophical ideology.

\textit{Id.} \textsuperscript{77} Third Party Intervention of the Belgian Government, \textit{supra} note 75, at 6.

A parallel can be drawn: it is today still generally considered more respectful toward others to take off one’s sunglasses in conversation so as to allow real and complete interaction. So many emotions pass through our face and specifically through our eyes. The eyes even promote listening. Trying to interact with respect for others without the classical rituals of greeting and looking (“les rituels classiques du bonjour et du regard”), is not well preparing the field of the relation. The right to isolation has its limits. I can go out on the street and not feel like engaging in a long conversation, but I have be ready to be watched and, ideally, greeted, even by someone who is a stranger to me. The notion of dignity in dressing and social contacts is relatively subjective yet the more a society is multicultural and the more types of religious and philosophical convictions and types of cultural customs coexist, the more persons have to be careful to not show them in a too demonstrative manner on the public street. The vestimentary codes in our societies are the product of a societal consensus, they are the fruit of a balanced compromise between our individual liberty and our codes of interaction in society.

\textit{Id.} (translated by author from French).
view results from translating the philosophical rhetoric into concrete behavioral requirements. The idea being that people who choose to be in public are not allowed to make themselves unavailable for interaction with others or to give such an impression. It also assumes that not showing one’s face should automatically be considered as a signal of withdrawal from, or unavailability for, social interaction. Yet, even if one accepts furthering social interaction as a “legitimate aim” that may justify restrictions of fundamental rights as a matter of principle, it appears difficult to justify the necessity of a face-covering ban in all public spaces to further this aim.

First, it is important to consider if wearing a face veil really prevents communication. In an age of mobile phones and online communities, the philosophical claim that someone cannot meet someone else without looking him or her in the face seems detached from reality. Several women interviewed expressed a self-image that included them as open or sociable persons. Many of the women stated that, from their perspective, communication is perfectly possible, even if they recognize that the veil could be experienced as a communicative barrier by those they speak to. Within their familiar environment, especially before the ban, women who wear a face veil participated in a range of social activities involving contact with others at schools, in shops, and administrative offices among others.

For example, one interviewee described her experience as follows, “Me, I talk to everybody, everybody sees me laugh; they answer me in the same tone if they want to. When they don’t want to, that’s another matter.”

Another interviewee also shared the following recollection.

At the time I lived in a neighborhood of old
people . . . . And these people recognized me without any problem and they acted toward me as if they saw whoever else in the street . . . . We were good neighbors, and I remember that when we moved, the old people were even sad because they told us: “Oh, we knew you so well and we knew that we could count on you, that we could ask you something.” There was even an old lady, who lived upstairs where I lived, and whose children did not visit her. And she told me, “it is so good of you, that you come and visit the elderly,” because I visited her from time to time with my children. So it does not stand in the way of anything at all. It is enough to want to accept the difference and to understand that behind that face veil, there is a person who is completely normal.  

Similarly, the applicant in S.A.S. v. France states:

The most important thing about how I communicate is my words and how I follow them up with deeds—not the visibility of my face. My veil compels others to respond to my brains, not my body; to respond to what I say, not how I look when I say it; on my character, not my clothes.  

Moreover, many interviewees reported positive contacts when they interacted with people who they had previously been unacquainted with. Several women told stories of how a conversation with an initially suspicious stranger turned into a positive exchange. For example,

In a supermarket, people told me “madam, why do you wear that?” I came closer to a gentleman and told him what the religion says. It is not mandatory but if you do it on your own initiative . . . . He told me “maybe your husband forced you.” I said “you see, I do my shopping all alone, and I drive alone, there is nobody with me.” And he was

---

80 Id. at nr 25.
81 Witness Statement of the Applicant, supra note 27, ¶ 20.
satisfied.\textsuperscript{82}

The empirical findings thus reveal the erroneous character of one of the main assumptions legislators had for banning the face veil, namely that the face veil indicates a withdrawal from social interaction. At least before the ban, women wearing a face veil were in fact interacting in numerous ordinary ways with society at large. Moreover, it appears that the ban has not increased these women’s social interactions. In fact, the ban may have decreased these women’s social interactions. Many women who choose to wear a face veil are strongly attached to it and continue to wear it despite the ban, meaning that they now avoid going into public except by car. These women are afraid of an encounter with the police as well as of the harassment and aggression by strangers. Hence, instead of increased social interaction, the effect of the ban on these women is a deterioration of their social life, their interactions with society at large, and their mobility.

Women report that the ban has led to women who wear face veils being harassed even more than before, severely limiting their ability to interact with society. A French face veil wearer reports a typical incident of harassment:

Last time I went to Auchan (supermarket E), a mob formed around me and people were saying “what are you doing here? It’s forbidden! You have no right to go out entirely veiled. It’s banned, it’s illegal. Go back to your country.” [I] feel like a monster. Even pedophiles and criminals, are not treated like that . . . . We are seen as less than nothing, not as human beings.\textsuperscript{83}

Another woman reports the negative effects of the ban on her life as a mother, “I don’t go out at all with my son. At two and a half he’s at the age to go to the park, to play outside. It’s not something I can do with him because I’m scared.”\textsuperscript{84}

Even if some women did take off the face veil as a result of the ban,\textsuperscript{85} and it did result in their being more easily approached.

\textsuperscript{82} Interviews with Belgian niqabis, supra note 79, at nr 2.
\textsuperscript{83} AFTER THE BAN, supra note 33, at 15.
\textsuperscript{84} Id. at 12.
\textsuperscript{85} Among the thirty-five women interviewed in France after the ban, eight
by others in the public sphere, the ban is a disproportionate measure to achieve that effect. The research suggests that the ban is a disproportionate measure to achieve that effect. In practice, women can and do communicate with their faces veiled. As the above interviews have shown, women who wear a face veil do experience positive exchanges. Social integration is a matter of goodwill on both sides and is likely to be better realized by inclusive means rather than ones that are repressive.

IV. BANNING IN THE NAME OF SAFETY?

In the public discussion surrounding face-covering bans, the safety argument often concerns subjective feelings of danger that are generated by the sight of a woman who covers her face. In the Belgian Parliament, for instance, the introduction of a face veil ban was compared to placing street lamps in a dark alley in order to reduce the fear of crime and provide a feeling of public security. In our interviews, veiled women recounted experiences of engendering feelings of unease or fear in others. One woman stated, “I understand completely that people are scared. It’s normal, because it’s covered, it’s hidden, you don’t know what’s underneath. At first sight, it’s shocking.” At the same time, these women’s stories also show that it is possible to overcome these feelings and to establish meaningful contacts and relationships with others. These contacts and relationships appear to exist in particular with persons with whom there is regular interaction and who may therefore be assumed to be “used to” the veil. Examples include neighbors, teachers, and shopkeepers. Thus, any feelings of danger do not necessarily accompany contact with veiled women, and as such those feelings can be overcome, it is

respondents have removed their full-face veil, while twenty-seven continue to wear it despite the ban. See id. at 2.

It should be noted, however, that the aggression of the public at large appears to extend to women who dress in a conservative, Islamic manner, even without a face veil. Hence, it is far from certain that those same women, when they take off the face veil yet otherwise keep dressing as they did before, would be more easily approached by members of the public.

an issue that might be better addressed by other means than those that are repressive.

Moreover, in the European Court of Human Rights, a feeling of danger can only serve as a ground for the restriction of human rights if there is an objective foundation for such a feeling. Therefore, a religious practice cannot be prohibited merely on account of the fact that a part of the population finds it offensive or even alarming. The court has made this particularly clear in its case law, holding that “a legal system which applies restrictions on human rights in order to satisfy the dictates of public feeling—real or imaginary—cannot be regarded as meeting the pressing social needs recognised in a democratic society, since that society must remain reasonable in its judgement.”

Following this line of reasoning, the argument of subjective safety, in this case with regard to face veils, cannot serve as a justification for restricting human rights.

By contrast, the promotion of objective public safety can be regarded as a legitimate government objective. Yet, any measure must still address an actual safety concern. Restriction of civil liberties—here, the right to freedom of religion and expression—cannot be based on speculation alone. One example of this is in *Arslan v. Turkey*, where members of a religious order were criminally convicted for wearing distinctive religious clothing in public. The court held that the convictions violated their freedom of religion since there was no evidence that the applicants represented a threat to public order, or that they were involved in proselytism.

Even if the ban on face veils did promote public safety, it would still be difficult to reconcile the ban’s broad scope—all public spaces—with the proportionality principle. That principle requires that a measure restricting a fundamental freedom must not burden that freedom any more than necessary to achieve its purpose. In most cases, safety risks can be overcome by

---

90 See supra note 52–55 and accompanying text for an explanation of the test applied by the European Court of Human Rights.
measures less restrictive than a ban, such as the obligation to lift the face veil upon a legitimate request. Public safety only requires the ability to identify someone when needed rather than permanent recognizability. Among our interviewees, we found that women were in general willing to identify themselves to police or other authorities by lowering their veils. Many interviewees explicitly stated that they would be willing to identify themselves to male as well as female officials. The applicant in S.A.S. v. France is no exception. In her application to the court, she confirmed her willingness “to show her face when a security check is required” as well as “to lift her veil when requested to do so for necessary identity checks.”

Hence, the risk that the face veil poses for objective safety in the general public sphere appears exaggerated, if not unfounded. This provides support that the ban is too broad and disproportionate to be justified as a safety measure.

V. BANNING FACE VEILS IN THE NAME OF WOMEN’S RIGHTS?

During the parliamentary debates concerning the French and Belgian legal bans the discourse emphasizing women’s rights and women’s dignity was abundant. Strong language was used, branding the face veil as a “mobile jail,” a “textile prison,” or the “shroud of freedom.” The underlying assumption of this argument is that women wearing a face veil are (mostly or always) forced to do so. In its submission in S.A.S. v. France, the French government stated that

to consider that women, for the sole reason that they are women, must hide their faces in public

91 Application, supra note 29, ¶ 4. See also Witness Statement of Applicant, supra note 27, ¶ 30 (“Of course, I accept that—at specific times, locations, contexts—legitimate public safety issues do require those wearing full face veils to satisfy security identity checks. At an airport or bank, for instance, or during heightened localized security threats, I accept that security identity checks (i.e., lifting my veil to reveal my face) are entirely justified and reasonable.”).

space, is to deny them the right to exist as individuals in this space and to reserve the expression of their individuality to the private family space, since only the men of the family have the right to see their face, or to an exclusively female space. Only men, according to such a view, are entitled to such public individual existence. Hence there is an absolute, publicly asserted, negation of equality between men and women.93

Yet, all available empirical research demonstrates that this central assumption is erroneous. While the research does not allow a conclusion as toward whether or not (and if so, how many) women are being forced to wear a face veil,94 it does clearly show that for a significant number of women who wear a face veil, the face veil is the result of an autonomous choice.95 All interviewees describe the decision to start wearing the face veil as a well-considered and free decision, a personal trajectory of deepening and perfecting one’s faith. One woman describes her trajectory this way: “I wore my veil first on the inside before I wore it on the outside. For me, the veil on the inside is the first thing. My veil is my chastity, it is my behavior, it is my politeness, it is my respect.”96

Another woman expresses how she sees herself and others who decide to wear a face veil: “A woman who is completely

---


94 In its submissions in S.A.S. v. France, the French government refers to the testimony of Sihem Habchi, chair of the organization Ni Putes Ni Soumises, who describes the testimony of a woman named Karima about the domestic violence she suffered from her husband, including her being forced to wear a face veil. See French Government Submissions in Response to the Third Party Interventions at 7, supra note 37. This is the only concrete example in the debates that we found. In the Belgian study, explicit efforts were undertaken to identify similar cases, yet none were found. See BREMS ET AL., WEARING THE FACE VEIL, supra note 8.

95 See THE EXPERIENCES OF FACE VEIL WEARERS, supra note 5, for reports on empirical research from Belgium, the Netherlands, France, the United Kingdom, and Denmark.

96 Focus group discussion with Belgian niqabis (on file with author).
veiled, for me she is a woman with strength, with enormous self-confidence . . . . You need it very much.”

In France and Belgium, nearly all women who choose to wear a veil were confronted with strong negative reactions from their relatives and friends, sometimes even their husbands. There is no evidence, in either France or Belgium, of pressure from husbands or relatives to wear a face veil; while there is recorded pressure from husbands and relatives to not wear a face veil.

By contrast, forcing a woman to wear a burqa or niqab amounts to an impermissible oppression of women, a type of domestic violence. It is doubtful whether criminalizing and fining the women in question can be considered a relevant measure to combat this oppression. This approach treats the oppressed woman as a perpetrator rather than as a victim. Hence, the idea of protecting women against the imposition of a face veil cannot justify a face-covering ban under Article 9 of the ECHR.

Overall, a woman’s agency appears as a strong and determining factor in her journey toward the face veil. These women generally see themselves, and each other, as “strong” women. They experience the ban as a denial of their autonomy and hence as anti-emancipatory. The applicant in S.A.S. v. France stated this view eloquently:

To be clear, neither my faith nor any man is dictating to me what I can wear in public: only the State is dictating that to me. Neither my faith nor any man is restricting my liberty to choose my own clothing: only the State is restricting that

---

97 Interviews with Belgian niqabis, supra note 79, at nr. 19.

98 While the Belgian ban only criminalizes the person who covers her face, the French law creates a separate offense for forcing another to cover her face. Yet, that provision does not appear to be applied in practice. One year after the ban was implemented, the French Ministry of Interior reported that 299 women had received a fine or warning wearing the full-face veil, yet there was no mention of any application to men. See Written Comments of the Open Society Justice Initiative, supra note 39, ¶ 12. Moreover, there is no evidence (nor even any claim in that sense by the French government in S.A.S.) that such warnings or fines are used to help women who might be victims of abuse.

99 See also Witness Statement of the Applicant, supra note 27, ¶ 15 (“I am a strong, independent Muslim woman.”).
liberty. Neither my faith nor any man is compelling me to obey clothing restrictions under specific threat of punishment: only the State would punish me if I failed to comply with its clothing restrictions.100

It should be noted that the court has ruled that personal autonomy “can also include the possibility of devoting oneself to activities perceived as being of a nature physically or morally damageable or dangerous to oneself,” and that “particularly serious reasons” are required for state interference.101 Hence, to the extent that the face veil is chosen freely, a ban based on the protection of women’s autonomy does not make sense. Moreover, research completed after the French ban went into effect shows that the ban has made women more dependent on their husbands because they go out less, either to avoid a fine because they cover their faces or, if they stopped using a face veil, because they feel uncomfortable.102 The French government rejects the relevance of free choice in this debate, stating that the face veil “effaces persons from public space,” and that “regardless [of] whether this effacement is desired or suffered, it is necessarily dehumanizing and can therefore hardly be regarded as consistent with human dignity.”103 The government also suggests those women who wear a face veil suffer false consciousness,104 having internalized an oppressive rule.105 The French government’s former claim, as the S.A.S. v. France applicant discusses, is “an abstract assumption based on stereotyping and chauvinistic logic that does not survive scrutiny.”106 Their latter claim is “deeply paternalistic and

100 Id. ¶ 14 (emphasis added).
102 AFTER THE BAN, supra note 33, at 8.
103 French Government Observations in Reply to Application, supra note 93, ¶ 92 (translated by the author from French).
104 The term “false consciousness” denotes the inability of members of subordinated groups to recognize their subordination, on account of their interiorization of the views and values that support this subordination.
105 Id. ¶ 105.
106 Final Observations, supra note 53, ¶ 95.
selective as it assumes that non-Muslim French women are not subjected to cultural influences in their personal choices.”

However, the focus on women’s autonomy is not all there is to say about the face veil from a gender or feminist perspective. From a more radical feminist angle, a practice that makes women—yet not men cover up—whether voluntarily or not, is an expression of male dominance. Yet this is just one of many “cultural” expressions of patriarchy. In the same sphere, mainstream French and Belgian culture makes women—yet not men—go to great lengths to be pretty. Applying make-up, shaving armpits and legs, and wearing uncomfortably high heels are just some of the expressions of French and Belgian women’s submission to patriarchy. Eradicating all such expressions is a valuable feminist project, but this is manifestly not the French or Belgian government’s project. If it were, it would be difficult to justify an exclusive focus on face veils while other equally patriarchal practices, that are immensely more widespread, remain unchallenged. Hence, this line of reasoning cannot justify the ban under the auspice of the protection of women’s rights.

In addition, a close reading of the parliamentary debates surrounding face veil bans reveals that the women’s rights argument is not necessarily focused on the rights of women who

---

107 S.A.S Response to the French Government’s Submissions ¶ 30, in S.A.S. v. France (Eur. Ct. H.R. filed on May 29, 2012) (“Western ideas of feminine sexuality and fashion are assumed to be unproblematic for gender equality, despite the feminist tradition that has long challenged such ideas as oppressive. The law banning face covering purports to promote gender equality, but does so in an ill-informed and discriminatory way.”).

108 See Dolores Morondo Taramundi, Women’s Oppression and Face Veil Bans: A Feminist Assessment, in The Experiences of Face Veil Wearers, supra note 5.

109 Taramundi defines a radical feminist perspective as “a form of critical theory originating in the second-wave, whose interest lies mainly with social power structures, and in particular, with patriarchy or sex-gender systems.” Id. at 1 n.1.

110 Yet, it might be doubted if criminal law is an appropriate way to realize such a project.

111 It is estimated that around 1,900 women wear the face veil in France; in Belgium, between 200 and 270 women. See GÉRIN, supra note 1, at 29, 74. This is less than 0.5% of the Muslim minority in these countries.
wear face veils. Instead, the argument is that the rights of all women would be offended by the use of the face veil. The French government in *S.A.S. v. France* states, “[t]he entire concealing of the face also affects the dignity of the persons who share the public space with the fully veiled person and who are treated as person from whom one has to protect oneself by refusing all exchange, including visual exchange.”¹¹² In legal terms, the question becomes, whether the face veil can be banned as a *symbol* of women’s oppression or lack of respect for human dignity. Prohibiting symbols is virtually always at odds with freedom of speech. In *Vajnai v. Hungary*,¹¹³ the European Court of Human Rights ruled that the application of the Hungarian ban on communist symbols to someone wearing a red star during a demonstration violated Article 10 of the ECHR. The court was mindful of the fact that for many Hungarians, communist symbols are associated with painful memories¹¹⁴ but stated, however, that the symbol did not *exclusively* represent totalitarian rule,¹¹⁵ nor had the Hungarian government shown that the use of the star had generated any danger of violence or disorder, or that there was a “pressing social need” for the interference with free speech.¹¹⁶ To the extent that the face veil can be analyzed as a symbol, the same reasoning should apply and thus any ban should be struck down.

Moreover, the empirical findings reveal another major problem with this line of reasoning—namely that such an analysis involves a strictly outsider interpretation of the face veil as a symbol carrying a message. However, as the interviews with women who wear a face veil show, those women do not intend to reject the outside world or to send any message. While most interpretations consider the face veil as a message to the world saying that “women should cover themselves,” or even that “women should be submitted to men,” the women themselves do

¹¹² French Government Observations in Reply to Application, *supra* note 93, ¶ 93.
¹¹⁴ *Id.* at para. 55.
¹¹⁵ *Id.*
¹¹⁶ *Id.*
not see it that way. Yet, those women who wear a face veil are well aware of these interpretations, and find them a source of major frustration. One woman stated: “It’s really, really humiliating and degrading for the personality of a Muslim girl, to hear someone say morning and evening ‘it’s the men who submit you, it’s the men who oblige you . . .’”117 Another interviewee reported, “That is really something I want for myself. For me.”118

For these women, the veil is not a message to the outside world. It is a very personal thing—a choice they make for themselves concerning their relationship to God. Proselytizing is far from these women’s minds. If there is a message, it is certainly not a normative one, in that it is not about telling others how to behave. Nor is it a message about gender relations, but instead about religion. If the face veil is a symbol at all, for the women wearing the face veil it is a symbol of religious devotion. At most, the veil could be considered a symbol of chastity. But chastity to gender inequality is a stretch.

We further analyzed our interviews from a gender perspective. There was a wide variety of views on gender relations ranging from very conservative to quite progressive. For example, one woman expressed her view in this manner, “I do not want to be equal to a man, I want to remain a woman, I don’t want to do the work of a woman and a mother and do the work of a man on top of that.”119

Yet, other women saw things differently, as illustrated by the following quotes:

I went to Egypt and met women with a burqa who were lawyers, doctors. I also want to be like that, to achieve something more. I can study, I can work, I can do the same things as any other woman or man.120

. . .

At my house, we are two to vacuum, two to

---

117 Interviews with Belgian niqabis, supra note 79, at nr. 21.
118 Id. at nr. 11.
119 Id. at nr. 5.
120 Focus group discussion with Belgian niqabis, supra note 96.
prepare meals, two to bathe the children, two to change the diapers, two to do the shopping, two to babysit.\textsuperscript{121}

A large majority of our interviewees were housewives. For some women, the life of a housewife is the expression of a commitment to traditional gender roles. Yet, there are others for whom the life of a housewife is not their first choice; they dream of a society in which they would not have to choose between a career and a face veil. Some women who wear a face veil express assertive emancipated views against traditional role patterns and against unequal gender practices in the Muslim community. Clearly, the face veil is not an indicator of its wearer’s approval of male dominance, let alone of its promotion. With such a gap between insider and outsider interpretations, it appears that the face veil is not truly a symbol at all. Instead, for some, it is an excuse to engage in textbook prejudice and stigmatization.

CONCLUSION

From the perspective of the women concerned, bans on face veils are counterproductive with respect to all three of the stated purposes for the ban: (1) they restrict women’s rights instead of furthering them; (2) they reduce social interaction; and (3) they expose women to serious safety risks. The French and Belgian legislators were not concerned with the impact on face veil wearing women, but rather with the effect on people who are confronted with women wearing the face veil. People for whom the sight of a face veil is an affront to women’s dignity, who do not want to interact with a woman wearing a face veil in shops or on the street, and who feel unsafe when they come across a face veil because they associate it with terrorism and fundamentalist Islam—it is those people whom the ban seeks to protect.

Will the \textit{S.A.S. v. France} judgment force European states to withdraw their face veil bans? If the European Court of Human Rights takes empirical reality seriously, it cannot uphold the bans. Yet the risk remains that the European Court of Human Rights cannot bring itself to look through the eyes of such a very

\textsuperscript{121} \textit{Id.}
different Other, and that it will simply accept the French state’s justifications based on widely shared assumptions and majority concerns—much like the French Constitutional Council and the Belgian Constitutional Court already have done.
GENDER AND (RELIGIOUS) ATTIRE: A MATTER OF (FREE) SPEECH

Alejandro Madrazo*

I. INTRODUCTION

Both gender and freedom of speech are topics of growing importance in Mexico. This is an undeniable observation when viewed in the light of constitutional development and debate. In recent years, the Mexican Supreme Court has decided a number of important cases affecting both gender and freedom of speech.¹


¹ For freedom of speech cases, not directly related to sexuality or gender, see Comisión federal de telecomunicaciones. El artículo 9o.-A, fracción XVI, de la ley federal de telecomunicaciones, al otorgarle facultades exclusivas en materia de radio y televisión, no viola los artículos 49 y 89, fracción I, de la constitución federal, Pleno de la Suprema Corte de Justicia [SCJN] [Supreme Court], Semanario Judicial de la Federación y su Gaceta, Novena Época, tomo XXVI, Diciembre de 2007, Tesis P. XXVII/2007, 26/2006, Página 963 (Mex.); Libertades de expresión e imprenta y prohibición de la censura previa, Pleno de la Suprema Corte de Justicia [SCJN] [Supreme Court], Semanario Judicial de la Federación y su Gaceta, Novena Época, tomo XXV, Febrero de 2007, Tesis 1a. LVII/2007, 1595/2006, Página 655 (Mex.); Libertad de expresión y el derecho a la información. Su importancia en una democracia constitucional, Pleno de la Suprema Corte de Justicia [SCJN] [Supreme Court], Semanario Judicial de la Federación y su Gaceta, Novena Época, tomo XXX, Diciembre de 2009, Tesis 1a. CCXV/2009, 2044/2008, Página 287 (Mex.); Primera Sala SCJN, amparo directo 6/2009, sentencia de 7 de octubre de 2009; Medios de comunicación. Su consideración como figuras públicas a efectos del análisis a los límites de la libertad de expresión, Pleno de la Suprema Corte de Justicia [SCJN] [Supreme Court], Semanario Judicial de la Federación y su Gaceta, Décima Época, tomo II, Noviembre de 2011, Tesis 1a. XXVIII/2011 (10a), 28/2010, Página 2914 (Mex.).

For cases regarding gender, reproduction or sexuality, not directly related

The Mexican Supreme Court has decided several cases that are related to the expression of gender and sexuality. The first, and most prominent, is the Amparo Directo Civil 6/2008, in which the Court considered sexual and gender identity, and affirmed that the right to freely develop one’s personality allows an individual to “project” his or her life “in all ambit of life,” including one’s identity. See Primera Sala SCJN, Amparo Directo Civil 6/2008, sentencia de 14 de mayo 2008, at 90. This line of argument was used by the Mexico City Assembly in its defense of same-sex marriage in the Acción de Inconstitucionalidad 2/2010. See Acción de inconstitucionalidad. La inclusión del artículo 391 del código civil para el distrito federal en el decreto de reforma a dicho ordenamiento, publicado en la gaceta oficial de la entidad el 29 de diciembre de 2009, así como su vinculación con un precepto que fue modificado en su texto, constituye un Nuevo acto legislativo susceptible de impugnarse en aquella vía, Pleno de la Suprema Corte de Justicia [SCJN] [Supreme Court], Semanario Judicial de la Federación y su Gaceta, Novena Época, tomo XXXIV, Agosto de 2011, Tesis P. XIX/2011, 2/2010, Página 869 (Mex.). The Mexico City Assembly argued that marriage is a form of freedom of expression both because of its connection to one’s right to freely develop one’s personality and contribution to public debate. The Supreme Court decided the case in August 2010. However, the Court did not explicitly affirm the expressive dimensions of marriage until 2012 in the Amparo en Revisión 581/2012, in which it spoke of the “expressive benefits” of marriage. Matrimonio entre personas del mismo sexo. Perspectivas para analizar su constitucionalidad, Pleno de la Suprema Corte de Justicia [SCJN] [Supreme Court], Semanario Judicial de la Federación y su Gaceta, Décima Época, Libro XIX, tomo I, Abril de 2013, Tesis 1a. XCVIII/2013 (10a.), 581/2012,
I have published elsewhere on the intersection of gender and free speech, exploring the theoretical and normative implications of understanding gender as a form of expression. Here, I hope to use that earlier work as a platform to address, specifically, the question of religious attire in public spaces and its intersection with gender equality.

I propose, first, that we understand gender as a form of expression, and second, that we understand religious attire (e.g., head gear worn by women belonging to a particular religious group) as not only (or mainly) religious attire, but also as attire that expresses gender roles. Furthermore, that the main function of freedom of speech is the protection and promotion of diversity in speech. Starting from these premises, I propose we take the debate over religious, female-worn head gear (i.e., head scarves) and recast it in terms of freedom of speech. That is, instead of framing the issue as one where there is tension between (religious) freedom and (gender) equality, the debate can be framed under the free speech analytic framework and recast as a tension within free speech. On one hand, we have the importance of women’s gender expressions; and on the other hand, a state’s interest in promoting gender equality. Discussing these issues under the free speech framework allows us to accommodate both a woman’s desire to wear religious head gear and the state’s

---

Página 965 (Mex.).

Lastly, in 2012, the Court decided the Amparo Directo en Revisión 2806/2012, a case about homophobic expressions. See Primera Sala SCJN, Amparo Directo en Revisión 2806/2012, sentencia de 6 de marzo 2013. This case is not so much about what can be expressed through one’s gender or sexuality, but about what others cannot discern about people’s gender and sexuality.

I want to thank my colleague, Estefanía Vela, an acute and systematic observer of the Court, for keeping all these cases on the radar and, specifically, for helping me prepare this footnote.

2 See Alejandro Madrazo, Género y libertad de expresión, in LIBERTAD DE EXPRESIÓN: ENTRE LA TRADICIÓN Y LA RENOVACIÓN. ENSAYOS EN HOMENAJE A OWEN FISS 257–87 (Esteban Restrepo Saldarriaga ed., 2013). The text was written for a Mexican legal audience, who was unfamiliar with both gender studies and the free speech doctrine in the United States. I use ample portions of that text here, and I would like to thank Pamela Ruiz Flores for her help with the translation of the sections used here.
attempt to ban it, while simultaneously empowering those same women.

We tend to discuss the question of whether specific head gear used by specific groups (women) within a larger religious community (Muslim) should be banned as a tension between religious freedom and gender equality. This framework relies on some individuals' beliefs that wearing a headscarf is valuable and should be protected by law because it is a religious and collectively-held practice. Furthermore, it also assumes that the practice of having women wear headscarves disempowers them and subjects them to traditional gender roles, undermining gender equality. The question between these two sides then becomes whether religious freedom should prevail over gender equality.

The problem with this framework is twofold. First, it sets the stage for arbitrary solutions. That is, it requires us to choose between one of two incommensurable, clashing values: religious freedom and gender equality. This dichotomous framework provides no common ground to resolve the conflict. Thus, it forces a choice that, in the end, is arbitrary: should freedom prevail over equality or vice versa? Second, by accepting this dichotomous framework, we are put in a position in which, by choosing equality over (religious) freedom, we conclude that prohibition of attire is an admissible policy. Needless to say, prohibiting voluntary conduct by others, which does not harm third parties, is always a difficult policy to support or accept. If, on the other hand, we choose (religious) freedom over equality we run the risk of legitimizing gender oppression yet again. Both alternatives disempower the actual women who choose to wear religious attire. Choosing equality over religious freedom makes gender inequality acceptable in the name of religion. Alternatively, choosing equality over religious freedom casts these women as either victims or collaborators of their oppressors and disqualifies their choice about how they want to live their life and express their gender roles.

I propose that we instead frame gender as a form of expression and attire as a form of gender expression. Using a specific understanding of both freedom of speech and gender, we can tackle the question of religious attire in a manner that will allow us to resolve the apparent tension between gender equality
and religious attire in a less arbitrary manner. Rather than simply choosing one value (gender equality) over another (religious freedom), we can take up the question in a manner that empowers—or at least refrains from disempowering—women who choose to wear religious attire. I suggest we see religious attire as a way women can express their adherence to specific aspects of gender roles they adopt and presumably value.

In Part II, I explain the analytic frameworks of both free speech and gender that I use to address these matters. In Part III, I propose that we understand gender as a form of speech and discuss the implications of this with regards to the tension between religious attire and gender discrimination. Finally, in Part IV, I argue that protecting religious attire as a form of gender speech better empowers women as opposed to denying religious attire constitutional protection in the name of gender equality.

One final word of warning: the platform for this proposal is taken from my previous work on the intersection of gender and freedom of speech, referenced above. It addressed the specific issues of gender and free speech—notably, same-sex marriage—in the context of Mexican constitutional debates. Then, I brought American authors to bear on Mexico’s development of free speech and gender debates, creating an enriching juxtaposition of constitutional traditions. Now, I bring Mexican constitutional cases and debates (along with the American authors on which I had previously relied) into the American forum, using direct translations from what I found pertinent from my previous work instead of refurbishing arguments. I do so because I wish to underline, not downplay, the origins of what I offer. Therefore, you will find references to Mexican law, authors, cases, and in particular, heavy reliance on one case (i.e., same-sex marriage as protected under the right to free speech). My hope is that this enriches the debate, but most important, I hope to avoid any pretention of discussing this issue within the confines of an American constitutional debate. Instead, I want to underscore that I am a Mexican constitutional scholar engaging with my peers from elsewhere on topics of common interest everywhere.
II. UNDERSTANDING GENDER AND SPEECH

This section seeks to explain the understandings of gender and free speech from which I approach the question of religious attire. It is structured in three parts. First, I define gender and flesh out its importance in the debate on the freedom of speech. Second, I provide the theoretical and doctrinal framework for the fundamental right to freedom of speech, from which I undertake my analysis. Finally, I present some of the clichés evident in recent academic discussions about freedom of speech in Mexico and try to either avoid or contest them.

A. Gender

It is a common trope to define gender in contrast to sex. In its simplest form, the distinction tells us that sex refers to the physiological differences between men and women, while gender refers to the roles or identities constructed, transmitted, and expected by society. These roles or identities are linked or associated with one sex over another. My sex is male because I have certain physiological characteristics that allow me to identify as such; my gender is masculine because as a child with the

Identifying these characteristics is actually much more difficult to answer than it initially seems. Laura Saldivia offers a synthesis that illustrates the complexity of the problem by pointing out at least eight medically distinguishable variables:

1) genetic or chromosomal sex, such as XY or XX; (2) gonadal sex determined by sexual reproductive glands, like the testes and ovaries; (3) internal morphologic sex that is determined after three months of gestation, such as seminal vesicles, prostate or vagina, uterus, or fallopian tubes; (4) external morphological sex, or genitals, such as penis, scrotum, clitoris, or labia; (5) the hormonal sex, such as androgens and estrogens; (6) phenotypic sex, or secondary sexual characteristics like facial or chest hair; (7) assigned sex and gender of rearing; (8) sexual identity.

referred physiological characteristics, I was taught to prefer football to dolls, the color blue to pink, and later, to sexually desire women, not men.

Therefore, in its simplest form, sex refers to the body, while gender refers to the social role, constitutive of an identity, and associated with the (sexed) body. The concept of gender emerges precisely to avoid the biological determinism of assigning social roles as a function of differences in reproductive physiology. In this regard, Professor Joan Scott tells us:

In its most recent usage, “gender” seems to have first appeared among American feminists who wanted to insist on the fundamentally social quality of distinctions based on sex. The word denoted a rejection of the biological determinism implicit in the use of such terms as “sex” or “sexual difference.” “Gender” also stressed the relational aspect of normative definitions of femininity. Those who worried that women’s studies scholarship focused too narrowly and separately on women used the term “gender” to introduce a relational notion into our analytic vocabulary. According to this view, women and men were defined in terms of one another, and no understanding of either could be achieved by entirely separate study.4

Consequently, discussing gender and not sex—as does the Mexican Constitution in the fifth paragraph of Article I5—emphasizes the social dimension, as opposed to the purely biological, of dividing people into men and women. The use of the term gender does not exclude biology; but rather, it encompasses issues that go beyond it. In addition, gender refers to the relational character (that is, that gender roles are defined in relation to one another) of assigned social roles based on

---


5 Constitución Política de los Estados Unidos Mexicanos [C.P.], as amended Diario Oficial de la Federación [DO], 5 de Febrero de 1917, art. I.
biological differences (mainly concerning reproductive capabilities). Gender and social roles are important for both men and women. Understanding the role of any one sex’s gender roles requires comprehension of both. Therefore, Professor Scott rightly points out that in the dominant social scientific discourse, [t]he term gender becomes a way of denoting “cultural constructions”–the entire social creation of ideas about appropriate roles for women and men. [T]he use of gender emphasizes an entire system of relationships that may include sex, but is not directly determined by sex or directly determining of sexuality.\(^6\)

However, the distinction between biology (sex) and social construction (gender) is not as sharp as it looks. The growing visibility of transsexuality and intersexuality directly controverts the distinction: sex has a strong component in social construction.\(^7\) We assign sex depending on how we interpret the body, sometimes literally *intervening in the body itself* and constructing one sex. For instance, when an infant has ambiguous sexual characteristics (a smaller penis than average, a clitoris larger than average, penis and labia, or a long list of possibilities), we intervene. Parents will then often decide which of two socially accepted options—male or female—the body of the infant will be adjusted to. This is done by removing—through surgery, hormone treatment, or some other means—the characteristics that are not of the chosen sex. The concepts of transsexuality and intersexuality controvert the discrete and binary frame in which all people can be classified as male or female. These phenomena demand that the conceptual apparatus

\(^6\) Scott, *supra* note 4, at 1056–57.

\(^7\) Transsexuality refers to a person changing his or her sex (from male to female or vice versa), who assumes the primary or secondary physiological sexual characteristics, conduct, and behaviors of the opposite sex. This does not necessarily question the binary distinction between the sexes, but rather questions whether the distinction is necessarily fixed. By contrast, intersexuality refers to people who do not completely fit into the physiological categories of male or female, and therefore resist the dominant binary classification of their physical bodies. *See* Saldivia, *supra* note 3, at 5.
of two discrete categories of male and female yield to either a gradual understanding in which there are multiple possibilities between these two poles, or else both poles are rejected for not adequately representing the reality of certain bodies and certain people.

For the purposes of this piece, I will not address in detail the implications of transsexuality and intersexuality on the analytical contraption through which we strive to understand the body and social relations. I will also not explore thoroughly the theoretical and normative implications—multiple and deep—of renouncing the use of discrete and binary categories that now prevail in our law. To the extent that transsexuality and intersexuality challenge the established categories of gender and sex, they should be considered as an expression of gender or as a gender expression. To be, or to be understood, as a transsexual expresses something in the same way that being or understanding one’s self as a heterosexual man expresses something. If someone refuses to be labeled explicitly as a specific gender, that refusal is an expression about gender. The most relevant aspect of transsexuality and intersexuality is what they tell us about the distinction between gender and sex for the purposes of free speech. Namely, that social construction is more important than it initially appears and that physiology is also a function of the cultural interpretations we make of the body. In this sense, transsexuality and intersexuality reinforce the importance of gender as a social construction and, thus, as an expression of what gender is or should be.

The social dimension—as opposed to the merely biological—and the relational character of gender are two important elements. A third important feature of the gender category is its necessarily political dimension. Foucault noted and analyzed the historical and discursive construction of sexuality. He argued that discourses that are generated around sexuality establish multiple and diverse power relations between people. The same happens when we talk about gender. Moreover, gender can be understood, among other things, as one such type of field, which has sprouted

---

around the study of sexuality, as well as one of the specific categories that have been incorporated into different disciplines (history, political science, medicine, law, etc.). The important thing is that gender—practices, symbols, ideas, customs, activities, artistic expressions, legal, and political or religious doctrines, regarding gender—produces power relationships between people.

B. The Theoretical Framework of Freedom of Speech

The future of constitutional interpretation of freedom of speech in the Mexican judiciary is uncertain. In a relatively short period, the Mexican Supreme Court has issued a number of opinions about the freedom of speech that are noteworthy in their theoretical dispersion and methodological inconsistency. Some opinions virtually extinguish freedom of speech, while

9 Santiago J. Vázquez Camacho, Introduction to LIBERTAD DE EXPRESIÓN: ANÁLISIS DE CASOS JUDICIALES XXVII (Santiago J. Vázquez Camacho ed., 2007).
10 See cases cited supra note 1.
11 The most famous case was the ruling of the Primera Sala de la Suprema Corte in the legal protection in review 2676/2003, better known as the case of “El Poeta Maldito” that Sergio Witz issued in October 2005. Primera Sala SCJN, amparo en revisión 2676/2003, sentencia de 5 de octubre de 2005. In that case, the majority concluded that the existence of a constitutionally protected entity (patriotic symbols) should be interpreted, ipso iure, as a limit to freedom of expression.

As rightly pointed by Francisa Pou, the Court deemed Witz’s poem punishable under criminal law, even though:

[T]here could not be a better example of what is often considered the core type of speech protected by the Constitution. That was a case of linguistic expression, not nonlinguistic or “symbolic” expressive behavior, as in the famous examples of burning American flags, books or crosses. The latter is generally analyzed as a regulation of expression, not a regulation of the conditions of the freedom of expression, as when discussing the influence of money in election campaigns, which is a regulation of expressive content. That is not simply a form and manner of expression, but rather one of indubitable political dimension. The expression had no individualized addressee, which excluded
others are progressive and demand that the State concern itself with improving the public debate about freedom of speech and its constitutional partner, the right to information. It is thus the need for complex weighing of judgments between freedom of expression and other fundamental rights of individuals (i.e. honor, privacy). The expression moved through an extremely classical channel, such as print media, and not a medium that stimulated discussion, such as television. Finally, the case concerned the speech of an individual, not a subject with a less defined constitutional status (interest groups, legal people, cultural communities).

Francisca Pou, *El precio de disentir, in Libertad de Expresión: Análisis de Casos Judiciales XXVII* 187–88 (Santiago J. Vázquez Camacho ed., 2007) (translated by author). If Witz’s poem does not find that that type of expression deserves constitutional protection under the freedom of expression, it is difficult to imagine what kind of expression does deserve it.

12 Known as the “Televisa Law,” case 26/2006, the Supreme Court, sitting *en banc*, issued the final portion of its Fifteenth *Considerando* on June 7th, 2007. See Comisión federal de telecomunicaciones. El artículo 9o.-A, fracción XVI, de la ley federal de telecomunicaciones, al otorgarle facultades exclusivas en material de radio y televisión, no viola los artículos 49 y 89, fracción I, de la constitución federal, Pleno de la Suprema Corte de Justicia [SCJN] [Supreme Court], Semanario Judicial de la Federación y su Gaceta, Novena Época, tomo XXVI, Diciembre de 2007, Tesis P. XXVII/2007, 26/2006, Página 963 (Mex.). The supermajority opinion explicitly discussed the function of the State as guarantor of the freedom of expression and the citizen’s right to information, which is a function that involves the obligation of the State to foster plurality and diversity in communication in order to achieve a society “more integrated, more educated and chiefly, more just.”

It is important to note that there are good reasons to be optimistic about the Court’s opinion on freedom of expression. It shows the underlying understanding of such an important fundamental right as having more weight in the evolution of the Court’s constitutional doctrine than the “damned Poet” precedent. First, the ruling is more recent. Second, it is a supermajority ruling by the Court sitting *en banc*, in contrast with a simple majority achieved in a Chamber. The ruling in Televisa Law was unanimous. Finally, the reaction and criticism from the legal community on the first ruling, and the overwhelming acceptance and celebration of the second one, should be read by the Court as an indicator of the quality of both rulings.

Further, the First Chamber seems to have honed its own criteria to issue a ruling in November 2006 in *amparo* 1595/2006. Libertades de expresión e imprenta y prohibición de la censura previa, Pleno de la Suprema Corte de Justicia [SCJN] [Supreme Court], Semanario Judicial de la Federación y su Gaceta, Novena Época, tomo XXV, Febrero de 2007, Tesis 1a. LVIII/2007,
impossible to predict or generalize the Court’s treatment of freedom of speech. Whether the State takes the role of a censor or as a protector of the diversity of expressions that reach the public forum, its role in relation to freedom of speech has not been understood by the Mexican Court to be a passive one. While much remains undefined, what is clear is that—at least in Mexico—the borders of the fundamental right to free speech are defined by the function of its political and instrumental role; that is, its role as an instrument for collective self-government.

Given the embryonic nature of a judicially generated constitutional doctrine around freedom of speech in Mexico, the theoretical framework that achieves the objective of this piece must be found elsewhere. Unfortunately, discussion of the constitutional doctrine on freedom of speech from Mexican academic circles is not particularly wide or rich. Undoubtedly, the recent decisions of the Mexican Court have generated academia’s interest in the subject, but there is no existing home grown theoretical framework sufficient to support the exploration of gender as an expression in the way this article contends.  

For the purposes of this piece, the theoretical framework articulated by Professor Owen Fiss is useful.

Fiss questions the assumption that state censorship is a threat
GENDER AND (RELIGIOUS) ATTIRE

against which the fundamental right of free speech is erected.\textsuperscript{14} Without denying the role of oppressor that a state can play against individuals who wish to express something, Fiss proposes that the state can also play the role of guarantor for the freedom of those same individuals. This is possible for two reasons. First, it is not only the state that can keep the individual from expressing herself; private power can also render individuals mute. Second, in Fiss’s understanding, the values that freedom of speech protects are eminently social, not individual. If a state values speech (and demonstrates as much in its Constitution), it is not because discourse is a form of self-expression or self-actualization, but rather because it is essential for collective self-determination, and therefore, to democracy.\textsuperscript{15}

Against the conception of freedom of speech that Fiss labels “libertarian”\textsuperscript{16}—freedom of speech protected as a form of self-expression, valuable in itself—Fiss proposes we adopt a democratic conception. Under such a conception, the purpose of free speech is to enrich and amplify the scope of public debate in order to allow ordinary citizens to know the issues that must be addressed and the arguments supporting the various positions around them.

If Fiss is right and a democratic conception of freedom of speech is the correct interpretation, then the state can have two distinct roles. First, it can play the role of censor, in which case the fundamental right to freedom of speech is a mechanism to prevent or stop certain abuse of political power. Second, it can play the role of promoter of vigorous public debate when powers different from the state are the ones censoring. In that case, the state must intervene to ensure that the weak are not silenced by the powerful.


\textsuperscript{15} “Speech is valued so importantly in the Constitution, I maintain, not because it is a form of self-expression or self-actualization but rather because it is essential for collective self-determination.” Id. at 3. In considering gender as an expression, I think the goal of this Article is to contribute to the discussion of Fiss’s position. Unfortunately, due to space restrictions, I will undertake this analysis in a future piece.

\textsuperscript{16} Id.
Fiss tells us\(^\text{17}\) that the U.S. Supreme Court initially decided cases involving freedom of speech by balancing the value of freedom (e.g., freedom of speech) against some counter-value (e.g., national security, the right to privacy and honor of the citizens, etc.). Under the libertarian model we would explain the conflict as a contest between two values that need to be balanced. If a value other than freedom prevails, a limit to freedom of speech exists that excludes certain types of speech from constitutional protection.

What are the consequences of adopting the democratic conception?\(^\text{18}\) The problem with an approach which balances value and counter-value is that, when the counter-value has the same constitutional status as the value, the balance between the values becomes sterile casuistry and impossible to resolve by application of general principles. It therefore becomes, to some extent, arbitrary. Such is the case, for example, when the counter value is equality in the form of the fundamental right to nondiscrimination. Under the libertarian conception, it would be necessary, at some point, to choose between the freedom of the discriminator who uses his or her freedom of speech to discriminate, and the discriminated subject’s right to equality. Fiss rightly proposes that under the democratic model, we can characterize the dichotomy in a more fruitful way: not as a conflict between freedom and equality, but rather as a conflict between freedom and freedom.\(^\text{19}\) The balancing then would take place between two competing manifestations of the same value (free speech), and thus can be resolved starting from a common ground and seeking to achieve a common purpose: the overall enhancement of that value.

Fiss observes that the problem of discriminatory speech, for example, is not only that such speech infringes on the value of equality—i.e., the fundamental right to nondiscrimination—but that it also has the consequence of “silencing” those who are discriminated against (or those who are excluded, slandered,

\(^{17}\) Id. at 5.
\(^{18}\) Id. at 3.
\(^{19}\) Id. at 15.
GENDER AND (RELIGIOUS) ATTIRE

etc.), impoverishing collective deliberation. Those who are discriminated against are effectively excluded from participating in the public debate, either because they are not heard or because, if they are heard, their voice is not valued because they have been previously disqualified. Fiss calls this the silencing effect of speech. But the silencing effect of speech does not only occur in cases in which the content of one person’s speech mutes the speech of others. It is also present in cases where, because the media through which competing discourses are expressed is asymmetric, the plurality of opinions is undermined. Asymmetric access to media has the effect of marginalizing one party’s speech making it effectively inaudible. Thus, plurality of opinions diminishes and public debate is rendered less robust.

The important aspect of the democratic model is that freedom of speech becomes an instrumental right with the immediate objective of ensuring inclusive public deliberation. This public

---

20 Id. at 16.
21 Id. at 16–18.
22 Mexico has recently witnessed this asymmetry in the case of Ley Televisa. See Comisión federal de telecomunicaciones. El artículo 90.-A, fracción XVI, de la ley federal de telecomunicaciones, al otorgarle facultades exclusivas en material de radio y televisión, no viola los artículos 49 y 89, fracción I, de la constitución federal, Pleno de la Suprema Corte de Justicia [SCJN] [Supreme Court], Semanario Judicial de la Federación y su Gaceta, Novena Época, tomo XXVI, Diciembre de 2007, Tesis P. XXVII/2007, 26/2006, Página 963 ( Mex.). When the Court deliberated this case, Televisa and TV Azteca, the two main national broadcast television companies who were the most interested in the outcome of the ruling, broadcasted many notes accusing the Supreme Court as being “Chavista” (referring to Venezuelan president Hugo Chavez) and totalitarian. During its public deliberations, the Supreme Court was already outlining the defeat of television companies’ interests. The two senators who led the challenge to the law were accused by the companies of being, in one case, a corrupt agent of foreign interests and, in the other, a murderer.

However, the court and senators are far from being vulnerable, voiceless actors. The court had media available, like Judicial Channel, and the Senate had extensive coverage in print media and some coverage on radio and cable television. But the difference between the ability of the court or senators to communicate compared with the ability of the two main national television broadcast networks was so abysmal that it had the effect of silencing the former from a large portion of the national population.
deliberation is political because its goal is to enable collective self-government. The state is constitutionally entitled to intervene, restricting a speaker in order to contain the silencing effect of his or her speech. The state is entitled to do so in the name of freedom of speech per se, not on behalf of another value or in spite of free speech. When the state intervenes in this way, it plays a role analogous to that of a parliamentary moderator: it removes someone from the podium so that others can now have access. This allows the plurality of speech to be enhanced and the robustness of public debate is aggrandized.

C. Common Tropes

Some of the tropes frequently used in discussions about freedom of speech are, I believe, counterproductive. The first trope establishes that there are different types of speech, and that the classification of speech under a particular category is central to determining if it is constitutionally protected. According to this notion, certain categories of speech are protected while others are not (or not as well) because some types of speech are more valuable than others. However, freedom of speech does not protect speech in and of itself. Instead, speech is protected by establishing a fundamental right because it has a specific function: ensuring that diverse issues and positions are not suppressed from public deliberation. The exercise of classifying speech into different types of speech has the effect of prejudging which discourses contribute to public deliberation.

For example, society tends to accept that religious discourse

23 For example, in Mexico, Juan Antonio Cruz Parcero argues that there are categories of privileged discourse: artistic, political and religious discourse: “[T]here are especially three aspects of these freedoms that are at all times crucial in a free society. Freedom to manifest religious beliefs and political ideas, and one generally ignored in the theoretical writings: freedom of artistic creation, that is, to manifest artistically.” Juan Antonio Cruz Parcero, De poemas, banderas, delitos y malas decisiones. La sentencia de la Suprema Corte sobre el caso Witz, 245 REVISTA DE LA FACULTAD DE DERECHO DE MEXICO 423, 430–31 (2006) (translated by author). He also argues that “freedom of artistic creation is a way to express ideas that deserves special protection, that a human being can express themselves artistically is considered something intrinsically valuable . . . .” Id. at 443.
deserves more protection than “obscene” speech (a category of speech which has historically been denied constitutional protection). At first glance, it is not clear why society should presume that a theological doctrine is of necessarily greater value than a pornographic image. Imagine the possibility that a theological proposition contributes little or nothing at all to cultural, political, social, or theological discussions. As a hypothetical, imagine a Roman Catholic individual arguing in favor of adopting the thesis of the immaculate conception of Mary, which has been part of Catholic Church dogma since the nineteenth century. The matter is quite settled for Catholics and quite irrelevant to almost all other groups, so positing the dogma contributes little to current public deliberation. In contrast, suppose a pornographic image provides a new perspective on how to enjoy healthy eroticism for thousands of people. For example, feminist pornography, or post-porn, both of which challenge the male-dominating discourse of commercial pornography without sacrificing the celebration of eroticism. Why would we hold that an argument in favor of the Immaculate Conception is inherently worthier than feminist pornography, a priori? In a case having to choose between guarding—by either protecting or promoting—one discourse over another, it would be rather more sensible to look at what each contributes to today’s individuals and/or today’s society in the existing historical and cultural context.

Moreover, prejudging based on the topic rather than the substance impedes an analysis of the effects of specific speech and consequently on what that speech brings to democratic deliberation. Of course, it is easier to have categories into which speech can be classified and then, depending on its classification, afford it greater or lesser protection. But this type of categorization contributes little to collective deliberation because it diminishes the potential for understanding and interpretation. It is important not to focus on what kind of speech is granted or denied protection. Instead, it is important to ask what that speech contributes to the public deliberation. The latter cannot be known

---

a priori and therefore it should not matter if the message falls into a particular category of speech. It should matter who and under what circumstances the message is offered.

A second trope is that speech is different from action. This is not commonly accepted in U.S. constitutional doctrine, but it is taken for granted elsewhere. Contrary to what this trope assumes, I hold that what is relevant is not the means by which we express ourselves—language, symbols, pictures, objects, actions, silence—but whether we are actually communicating something. Marching, burning a flag, boycotting a product—these are all forms of expression. Expressions, whatever form they may take, contribute to public deliberation.

The third trope is the notion that freedom of speech is a right enforceable against the state. Historically, freedom of speech may have originated as an effort to protect political dissidents from government violence. However, the genealogy of this fundamental right does not seem reason enough to limit its function. Today, oppression of private citizens by private power is more visible (and maybe more common). Private media can shut out a message; it can discredit a messenger; and it can project a specific message with a force unparalleled in the past. In the large political communities in which we live, private mass

---

25 For example, Francisa Pou quotes Paul Salvador saying that what is spoken and what is written is as different from the facts as “spirit is from matter.” Pou, supra note 11, at 188. However, the distinction between words and actions is becoming less relevant in the field of freedom of speech. The classic example is saying the word “fire” to a firing squad, which is no longer the exception to the rule. In an information society, words and actions are increasingly confused. For American constitutional doctrine that accepts freedom of speech protection for expressive conduct, see, for example, United States v. Eichman, 496 U.S. 310 (1990) and Texas v. Johnson, 491 U.S. 397 (1989).

26 American constitutional doctrine still holds that all fundamental rights are justiciable against the State, but another constitutional doctrine recognizes that such a right protects against private citizens as well. In considering free speech, however, the idea of the State as the only censor is still prevalent outside the U.S. Once again, Pou quotes Paul Salvador maintaining that a key premise of freedom of speech is that it has to protect those who “individually confront the established power, preferably the public, but also the private power.” Pou, supra note 11, at 188–89.
media is an example of a particularly important vehicle of communication and a particularly salient source of the silencing effect of speech. 27 It may be that private power can be deemed a more dangerous censor than public authority.

This last point is crucial. It has become a widely accepted thesis that large private powers are a potential threat to freedom of speech, but this is accepted by analogy with the State. 28 Namely that those who represent a threat to freedom of speech are the people or organizations who provide a public service (e.g., radio broadcasting), or are an economic power that has a disproportionate influence over the state, market, society, or all of the above. For example, a company or group of companies may monopolize basic services such as telephone services. Or else, a historic entity that having rivaled the state still holds sway over large portions of the populations—for instance, the Catholic Church. In all of these cases, the State as a paradigmatic censor remains near at hand in the imagination. We need to broaden our understanding of censorship by private actors in order to address some of the most ordinary forms of censorship at play in gender.

If we seriously consider that freedom of speech does not protect speech itself without qualification, but instead protects the plurality and diversity of speeches for inclusion in political deliberation, then we have to unmoor ourselves from the dominant paradigm. I propose broadening our perspective of what constitutes an agent that is capable of impinging upon freedom of speech to include not only private agents who have disproportionate power in absolute terms, but also to those private agents who have disproportionate power relative to the silenced speaker. 29 For example, a man who believes that women

27 Fiss, supra note 14, at 5–26.
28 For example, Pedro Salazar Ugarte and Rodrigo Gutierrez Rivas make this point, referring explicitly to the potential of actual private violators of freedom of speech in Mexico: the major economic powers, the media, multinational corporations, and criminal groups. They probably would not object to including noneconomic powers like churches, but they seem to have in mind a power similar to that of the State in some way. Ugarte & Rivas, supra note 13, 6–7.
29 For example, Salazar and Gutiérrez would easily coincide, since they rely in a relational conception of power and freedom when they explain how
are obligated to carry to term an unwanted pregnancy may not have a silencing effect on the candidate for public office that is running on a pro-choice platform, but may be able to silence his wife in a conversation with her pregnant daughter to decide whether she travels to Mexico City to terminate an unwanted pregnancy. This expanded understanding of the censor may be irrelevant when discussing the regulation of political propaganda or the use of the electromagnetic spectrum for broadcast, but it is important when addressing gender.

III. GENDER AS EXPRESSION AND GENDER EXPRESSIONS

Gender should be understood as a form of expression. There are two different perspectives underlying the policy implications for understanding gender as a form of expression: gender constitutionally protected as expression, and gender as a form of expression that limits another’s expressions about gender.

A. Gender as Expression

In order to understand in what sense gender is a form of expression, it is necessary to understand what is meant when the term “gender” is used on a daily basis. Professor Joan Scott’s work is helpful in this understanding. After analyzing the historical evolution of the use of the concept of gender, Scott presents a rich and complex conception of gender using two propositions that help to understand gender as an expression: “The core of the definition rests on an integral connection between two propositions: gender is a constitutive element of social relationships based on perceived differences between the sexes, and gender is a primary way of signifying relationships of power.”

---

30 I do not suggest that the State should intervene directly between private parties, but I want to illustrate the silencing effect. I will conclude that privacy interests outweigh freedom of speech in a case like this one, but it does not mean that a silencing effect is not present. I thank Estefanía Vela Bara for suggesting that I clarify this idea.

31 Scott, supra note 4, at 1067.
The first proposition—gender as a constitutive element of social relationships—is split into four elements, analytically distinct but closely related. The first of these elements is made of “available symbols that evoke multiple (and often contradictory) representations.” Scott uses Eve and Mary as examples of the Western Christian traditional gender symbols. The second element is the “normative concepts” that guide the interpretation of these symbols, checking and limiting their possible interpretations.

These [normative] concepts are expressed in religious, educational, scientific, legal, and political doctrines and typically take the form of a fixed binary opposition, categorically and unequivocally asserting the meaning of male and female, masculine and feminine. In fact, these normative statements depend on the refusal or repression of alternative possibilities, and, sometimes, overt contests about them take place . . . . The position that emerges as dominant, however, is stated as the only possible one. Subsequent history is written as if these normative positions were the product of social consensus rather than of conflict.

This second element is particularly relevant to freedom of speech. Specifically, with regard to the “normative concepts” or symbols, Scott proposes these concepts, which include legal, political, and religious doctrines, among others, that tell us how the interpretations of those symbols should be—or more precisely—how they can be interpreted. These dominant normative concepts contrast and suppress other possible interpretations of such symbols, naturalizing the interpretative possibilities that prevail. This means that the interpretation of symbols is forged by contrasting interpretative alternatives, which,

32 Id.
33 Id.
34 Id.
35 Id. at 1067–68.
36 Id. at 1067.
if one comes to be dominant over the others, can suppress the other symbols. What is at stake then is the interpretation of symbols that tells us what we are as men and women (and, in addition that we are men or women), and what we should be as men or women. Gender consists, in part, of an interpretation that seeks hegemony and suppresses different interpretations about what we are. In gender, we are in the field of discourse, and more specifically, a discourse to be imposed as a fixed fact that displaces alternatives.

The third element Scott describes is the social institutions and organizations that adopt and reproduce the interpretation of the symbols that are presented as fixed and as a product of consensus, when they really are not. Scott speaks of, at least, four institutions in which this takes place: kinship, work, education, and government.37

A fourth element is the subjective identity. The symbols, the indications of how we should interpret them (that is, the normative concepts), and the social institutions that adopt and reproduce these interpretations all have a direct impact on how we come to understand ourselves.

This understanding of gender enables one to see the intimate link between freedom of speech and gender. Gender is formed by a cluster of expressions: symbols, doctrines that tell us how to interpret these symbols, institutions, and organizations that require us to accept those symbols and ideas about ourselves, and reinforce the workings of the cluster as a whole. Gender is one of the forms of speech that permeates through us and connects us with each other; gender infuses our institutions, our doctrines, and our symbols with meaning, and constitutes our subjective identities. When one acts according to one’s gender role, one draws meaning from symbols and doctrines associated with that role. One uses that meaning in order to act within basic social organizations and institutions—such as family, school, religion, or government—and thereby confirms and reaffirms such meanings by understanding one’s self through the resulting interpretative framework. Gender, like expression, and like discourse in general, provides meaning and defines persons,

37 Id. at 1068.
institutions, relations, and symbols. Dressing a newborn in blue says something of what is expected of him, of what, starting then, he is. That act conveys a message to him, and to the rest of us. Professor Scott tells us: “[t]he sketch I have offered of the process of constructing gender relationships could be used to discuss class, race, ethnicity, or, for that matter, any social process.” Scott is right, but this does not diminish the discursive and expressive dimension of gender.

Scott also provides a second proposition that specifically explores the political profile of what she thinks is specific to gender (without actually describing it as exclusive): its ability to articulate power relationships. Scott explains:

[G]ender is a primary field within which or by means of which power is articulated. Gender is not the only field, but it seems to have been a persistent and recurrent way of enabling the signification of power in the West, in Judeo-Christian as well as Islamic traditions . . . . Established as an objective set of references, concepts of gender structure perception and the concrete and symbolic organization of all social life. To the extent that these references establish distributions of power (differential control over or access to material and symbolic resources) gender becomes implicated in the conception and construction of power itself.

Gender provides guidelines that naturalize and legitimize the distribution of power. It is a deeply (but not exclusively) political discourse.

The political use of gender is not limited to the perceived sexual differences between men and women. Citing anthropologist Maurice Godelier, Scott argues that differences between the sexes are often invoked in relation to social phenomena that have nothing to do with sexuality but by being attached to sex differences, become socially legitimate. That is,

38 Id. at 1069.
39 Id.
40 Id. (citing Maurice Godelier, The Origins of Male Domination, 127
gender serves as a key to interpreting social relations that have nothing to do with sexuality, and legitimizes them. As Scott explains, “Gender has been employed literally or analogically in political theory to justify or criticize the reign of monarchs and to express the relationship between ruler and ruled.”41 She goes on:

Gender is one of the recurrent references by which political power has been conceived, legitimated, and criticized. It refers to but also establishes the meaning of the male/female opposition. To vindicate political power, the reference must seem sure and fixed, outside human construction, part of the natural or divine order. In that way, the binary opposition and the social process of gender relationships both become part of the meaning of power itself; to question or alter any aspect threatens the entire system.42

Gender is so embedded within the symbolic language of power that the enterprise of problematizing gender is necessarily a political one. Gender is political in both the strict and expansive sense of the word: it configures power relationships between individuals—whether in the bedroom, at school, in the office, or in court—and power is frequently read in terms of gender. In short, gender is an expression and, significantly, a political expression in all its senses.

B. The Protection of Gender as Expression

There are valid reasons to protect gender as an expression. Whether gender is manifested linguistically or through behavior, it reflects and informs interpretations of what we are as men and women (or, neither one nor the other). Equally important, gender

---

41 Id.
42 Id. at 1073.
is an important political expression.

The richness of the analysis that stems from the premise that gender is an expression can be perceived from two perspectives: whether we are trying to reaffirm a dominant gender role, or if we are trying to question a dominant gender role. In both cases, gender as expression must be preliminarily protected. In the end, however, such protection can be curtailed or even defeated depending on various issues, such as the possible or actual silencing effect of that expression on others.

Some gender expressions reaffirm established gender roles. The controversy over the use of the headscarves by Muslim women in certain public spaces in Western Europe has been widely discussed, specifically the legal ban on the use of headscarves by Muslim students in French schools. Most frequently, the matter has been analyzed as a conflict between the apparent discrimination that young women are subjected to in wearing the veil and their freedom of expression. Salazar and Gutiérrez take such an approach: “Such practices have a community-religious thrust, and according to the report of the Commission [Stasi, which conducted the preliminary work leading to prohibitive legislation] they run counter to the principle of equality between men and women because they put the latter in a situation of marginalization.”

Thus, the conflict has been (partially) understood as one between the young Muslim’s right to nondiscrimination, and the right to freedom of religion and religious speech. This approach is problematic from the onset. Who is understood as the title-holder of the right to freedom of religion/religious speech? The Stasi Report, upon which the French legislation is justified, assumes that wearing the veil is most often not a girl’s voluntary decision, but an imposition by the girl’s parents and communities. The Legislature thus assumed that it is an expression or practice that is imposed, not chosen. As noted by Salazar and Gutierrez, this assumption dissolves the conflict:

---

43 Ugarte & Rivas, supra note 13, at 75.
there isn’t really a protected right to freedom of religion or speech, since the speech/practice is not free. Under this assumption, the veil should be banned because it violates two of the fundamental rights of Muslim girls: nondiscrimination and free speech.

The problem is that the presumption that the headscarf is an imposition needs to be proven. Denying, ex ante, these girls’ autonomy and attributing their religious expression not to them, but to their parents and communities, is a rhetorical and argumentative resource without empirical proof for such a supposition. Such proof can only be determined on a case-by-case basis, not through a general mandate. It may well be the case that such practices are imposed, coerced, or not voluntary, but the state should not presume so.

The controversy surrounding headscarves can be analyzed in a much more useful manner if the issue is rephrased as a conflict between two competing claims, both of which are grounded in freedom of speech: the expressive act of women who wear headscarves and the State’s interest in promoting gender equality.

To understand the first of these claims, we need first to ask ourselves if there is an important inaccuracy in how the expressive behavior of wearing the headscarf has been understood. Using a headscarf cannot be cast as an exclusively religious practice. Further empirical work is needed to better understand the phenomenon, but arguably, wearing a headscarf also expresses something about a woman’s gender role. The headscarf says something about the wearer as a woman as much as it expresses something about her as a Muslim. Women wear the headscarf because they are Muslim women, not just because they are Muslim, period. Therefore, the veil is linked with how (Muslim) women relate to others as women. That the veil has as much to do with the fact of being a woman than with the fact of being a Muslim is illustrated by some countries, such as Saudi Arabia, Southern Sudan, and Iran, in which all (post-pubescent) women must be covered in public, whether they are Muslim or not. Of course, in cases in which women are forced to wear the veil, it may not be deemed an expressive act to be protected by

---

45 Id. at 76.
free speech, but the point is that headgear is used by women because they are women, not exclusively or even necessarily because they are Muslim. For example, Muslim men do not wear headscarves, either at home or in public spaces. In addition, the profession of the Islamic faith is not represented directly by the headscarf: the headscarf does not have the same function among Muslims that the cross has among Catholics. The headscarf is not a symbol of the faith; non-Muslims are expected to wear it in certain contexts. The analogy is imprecise (and Christian-centric). The headscarf expresses something about what women are (or should be) and how they interact with men, with people outside their homes, and generally in public spaces, because they are women.

On the other hand, the State has a legitimate claim to promote both secularism and equal treatment between men and women and may deem that, in certain contexts, headscarves undermine both. Removing a practice that may hinder the achievement of a legitimate state interest in strategic contexts (such as schools where the young are in the process of defining their identity in the midst of their broader community) can be held to be legitimate. When banning headscarves from schools, the State is sending a powerful message: at school, a space in which the young acquire what are deemed to be necessary and shared abilities and knowledge, religious and gender-biased attire is out of place; what we all share in common cannot accept gender (or religious) cleaving. Women, the State is saying, should not present themselves to their peers as women first, and then as peers. Rather, they should be deemed peers on equal footing first and foremost. In a gender-biased world, literally covering women singles them out and skews the way their interactions are received. The State’s message can be seen as analogous to the policy of banning smoking from non-enclosed areas in educational facilities: such a ban has more to do with protecting the young from seeing—and potentially emulating—adults who are authority figures engage in destructive behavior than with protecting them from second-hand smoke. In banning all smoking from educational facilities the State conveys to the young a powerful message regarding smoking.

The conflict we are concerned with should thus be recast as
follows: on one hand, the state has an interest in communicating the importance of secularism and substantive equality between men and women; on the other hand, young Muslim women wearing a face veil\(^46\) are expressing speech that seeks to convey something about what they are, should be, or should appear as (i.e., something about their gender role). Under this framework, the State’s prohibition of headscarves is a message in favor of secularism and equality between men and women. Muslim women conceive of the headscarf as an expression of something about their identity as women, rather than (only as) exercising a religious practice or expressing (exclusively) something about their religion. Such tension between the state and the Muslim women can be resolved by working from a common platform: free speech.

The dispute is symbolic and discursive. To reframe the conflict in these terms does not require a particular solution. It could conceivably be argued that the state has a legitimate and compelling interest, or, indeed, an obligation, to promote secularism among its youth. One method the state could use is suppressing any symbol associated with (though not necessarily expressive of) a religion that distinguishes people by creed. For instance, Article III of the Mexican Constitution explicitly commands secular education.\(^47\) But one could also argue that young Muslim women in France have the right to express, through the use of the headscarf, whatever they consider they are or should be as women. In any case, it seems that young Muslim women wearing the headscarf in school are expressing a view in France when wearing the headscarf in accordance with the established gender role in their cultural and religious environment. This is gender as speech, and it should be protected.

In casting the dispute this way—as (state) speech versus (women’s) speech—the resolution can operate under the same principle and seek the same goal: safeguarding freedom of

---

\(^46\) I assume here that the expression of wearing a headscarf or a veil is voluntary. If not, there is no possible case for constitutional protection, at least not under the doctrine of freedom of speech.

\(^47\) Constitución Política de los Estados Unidos Mexicanos [C.P.], as amended Diario Oficial de la Federación [DO], 5 de Febrero de 1917, art. III.
speech, promoting diversity, and plurality of positions in matters of public concern. For example, the solution could be tailored to both protect Muslim girls’ gender expression from being suppressed, and authorize the state to manifest the importance of secular education and substantive equality between men and women. This solution could propose, not impose, that women—Muslim or not—need not accept a gender role that requires them to hide part of their body. Banning the headscarf would be unconstitutional, but the state could express—through other means—its desire for gender equality and secular public institutions. For instance, taking into consideration the age of the girls and their educational environment, the weighing of rights may favor the elimination of gender distinctions in their entirety, including those most accepted in French society. The State could, for instance, impose a policy in which both men and women use a standardized uniform consisting, for example, of shorts or long robes. 48 In any case, it is important to recognize that when the headscarf is freely taken and not imposed, Muslim women are expressing something about what they understand they are and something they understand women to be. While no fundamental right is absolute and the right to the free expression of gender can still be defeated depending on the circumstances, it is important to recognize gender’s expressive dimension.

Another example that helps to clarify the importance of understanding the problem of gender under the freedom of speech conceptual architecture is where gender roles are controverted, not reaffirmed, by expressive behaviors: notably the case for same-sex marriage. One of the dominant expectations deriving from gender roles is that in most societies, women must be attracted to men, and men to women. Homosexuality counters this aspect of gender roles in our societies. Women who are attracted to other women and men who are attracted to other men are still often regarded as deviants, both from what is expected, and from the accepted virtues of their gender.

In Mexico City, the Legislative Assembly established the legal institution of domestic partnership (*sociedad de convivencia*)

---

48 This proposal, of course, may have other serious constitutional problems itself.
in 2006. This institution gave partners most (some would argue all) of the rights and obligations that marriage gives to spouses. A few years later, in 2009, the same legislative body decided to go further and change the definition of marriage in order to eliminate the requirement that the two people marrying be a man and a woman, thus legalizing same-sex marriages (and adoption).

In Mexico, why were domestic partnerships not enough? In Mexico City, the rights linked to sustenance, successions, interdictions, and even adoption do not differ greatly between marriage and domestic partnerships. Adoption—which drew most energy of the ensuing debate—was not the reason because domestic partnership law was tailored to allow adoption by same-sex couples. Seemingly, what justified and, more important, motivated the legalization of same-sex marriage in Mexico City (where the legal functional equivalent to marriage already existed), was the expressive dimension of the institution of marriage. Marriage has an important expressive function as a symbolic role. Couples communicate their commitment through the act of getting married. They do so to each other, to their communities, and to the state. If marriage did not have a communicative, expressive, and celebratory function, most couples would marry before the Civil Registry (i.e., Town Hall) as if they were getting driver’s licenses (some do, certainly).

Most people get married for its symbolic value and because of what marriage represents. Few couples are primarily concerned with, or even aware of, the legal implications of getting married. In many cases, what matters—at least when you’re getting

49 Ley de Sociedades de Convivencia, GODF (Nov. 16, 2006).
50 Decreto por el que se reforman diversas disposiciones del Código Civil para el Distrito Federal y del Código de Procedimientos Civiles para el Distrito Federal, GODF, 525–26 (Dec. 29, 2009).
51 See Código Civil para el Distrito Federal [CC] [Federal Civil Code], Diario Oficial de la Federacion [DO], arts. 391, 392.
52 Article V of the Domestic Partnership Law for the Federal District equates partners with common law marriage for all legal purposes, while Articles 391 and 392 of the Civil Code for the Federal District equates common law spouses and formal spouses on the matter of adoption. Ley de Sociedades de Convivencia, GODF, art. V (Nov. 16, 2006); see also Código Civil para el Distrito Federal [CC] [Federal Civil Code], Diario Oficial de la Federacion [DO], arts. 391, 392.
married—is to communicate the existence of the union rather than to regulate it by law. Getting married and establishing a domestic partnership are acts which say different things to the people involved and to society. It is not a difference in importance, but a difference in kind.

Marriage as a speech act is protected under freedom of speech. The demand that same-sex marriages be recognized tells society something about the purported “deviant” character of homosexuality. Namely that if the law itself recognizes the equal legitimacy and status of a homosexual union in relation to a heterosexual union, it is saying that homosexuality is not or should not be understood as a deviation or variation. A gay couple that gets married is through that act saying something to society: our union is as legitimate, and in the same ways, that heterosexual unions are. Through marriage, homosexual couples have a vehicle to contest the gender roles they challenge and are still often imposed on them. In this regard, the state has a constitutional obligation to give homosexual couples access to the means of expression through marriage. This is due not only, or not even mainly, because of right to nondiscrimination, (after all, one could argue that in terms of personal and property rights, the domestic partnership equates or can equate homosexuals and heterosexuals), but because of the protection of the right to freedom of speech.

The state has an obligation to allow diversity of expressions linked to gender roles, and fulfill its obligation by extending the use of marriage as a form of expression, particularly for those who express gender roles that diverge from dominant ones. That is, especially to those who bring diversity to the “market place of ideas”53 about gender relations. Same-sex marriage should be constitutionally analyzed as expression through opposition of established gender roles, in addition to being analyzed under fundamental rights to equality, nondiscrimination, protection of the family, health, etc. Both wearing a headscarf and getting married are communicative acts that deserve constitutional protection under the right to freedom of speech.

C. Expressions about Gender

Starting from the democratic model of freedom of speech, the state’s function as moderator is particularly relevant. The state must seek to eliminate or mitigate the silencing effect of speech of some individuals in order to protect the speech of others.

Gender as an expression (particularly, but not exclusively, the behaviors and gender expressions that contradict established gender roles) must be protected when other expressions regarding gender threaten to silence it. The silencing effect is accentuated to the extent that the silencing expression disqualifies or intimidates others. Disqualifying a speaker (that is, labeling him or her as not apt for participation in the debate or expressive act)—to the extent that it says something about the speaker and not about the matter under discussion—has a particularly potent and harmful silencing effect, and contributes little or nothing to the general discussion.

For example, in the days following passage of the same-sex marriage law in Mexico City, Mexican Archbishop Norberto Rivera made several controversial declarations. He stated:

[The legislative reform that allows same-sex marriages] has opened the gates to a deviant possibility which allows these couples to adopt innocent children, whose right to have a family built by a mother and a father will not be respected, with the consequential psychological and moral damage that this injustice and arbitrariness will therefore cause . . . .

The Church considers an aberration to compare the union between same sex persons with marriage, because these are not able to reach the ends that gave origin to this essential institution that for Christians doesn’t just follow a form of social organization, but it is rather the order instituted by God since the creation of the world, and above the divine will that rules over the morality of marriage, no human law can be. 54

---

54 Gabriel Leon Zaragoza, Inmorales y aberrantes, las reformas aprobadas: Norberto Rivera, PERIÓDICO LA JORNADA, Dec. 22, 2009,
Rhetorically, same-sex couples are labeled as deviant and aberrant, and portrayed as a threat, not to the Archbishop, but to the innocent children who risk being adopted by them. Furthermore, same-sex couples are disqualified because of their fundamental betrayal of their gender roles: a person should be attracted to the opposite sex because the objective of sexual intercourse must be reproduction, which same sex couples can’t achieve (by themselves, the Archbishop should have qualified). The Archbishop attributes to marriage a necessary goal that by its (divinely ordained) nature excludes same sex couples.\(^\text{55}\)

The Archbishop’s statements exemplify normative doctrines to which Scott makes reference:\(^\text{56}\) they present the Archbishop’s interpretation of symbols that give value to institutions (marriages) as naturally truthful and therefore invulnerable to criticism. With this, the statements render impossible any interpretation of alternative choices. Thus, the Archbishop’s statements reinforce the monopoly of dominant ideas over the meaning of institutions; in this case, marriage. These statements create an interpretation of marriage that seeks to eliminate the opinions of the Archbishop’s rivals.

I doubt the Archbishop’s statements inhibited openly gay couples from getting married. I also doubt that homosexuals in Mexico ceased being homosexual because of what was said by the Church prelate. Nevertheless, I believe that, at least among the Catholic homosexual population of Mexico, the Archbishop’s statements will make some couples or persons refrain from expressing their intimate commitments to each other through marriage. This is due in no small part to the message itself. In his message, the Archbishop threatens homosexuals, at least those that are believers in his faith, to adopt the behavior that he expects from them: he states that same-sex persons’ marriages “have no future” because homosexuals that desire to get united under this scheme are “too few.”\(^\text{57}\)


\(^{55}\) Id.  

\(^{56}\) See supra Part III.A.  

\(^{57}\) See Zaragoza, supra note 54, at 29.
If we accept the democratic model of freedom of speech, in which the state has to intervene by limiting a dominant speaker so as to ensure that others are not excluded from collective deliberation, normative consequences follow: under this model, the state should protect expression (i.e., the questioning of the role of gender established through the celebration of a marriage between people of the same sex) by restricting or containing (not suppressing)\(^58\) the Archbishop’s message, and thereby attenuating his disqualifications and threats. The effect of the threatening and disqualifying tone of Archbishop Rivera’s statements on same-sex couples, especially Catholics, that may want to express themselves through the celebration of a civil marriage, is that of inhibition.\(^59\) To prevent this inhibition the state must provide some form of remedy to counter the Archbishop’s statements. This could take the form of a monetary fine—symbolically communicating that the Archbishop’s statements were reprehensible and impinged upon other’s rights—the demand of a public apology, or the promotion of the use of marriage by same-sex couples who wish to publicly express their commitments.

IV. CONCLUSION: PROTECTING GENDER SPEECH THAT REINFORCES GENDER ROLES AS A FORM OF EMPOWERMENT

Regardless of the analytic soundness of the offered framework, individuals interested in advancing gender equality and challenging gender roles should strategically favor protecting expressive practices (such as wearing a headscarf), even if such

\(^58\) By suppression, I mean the act of silencing or restricting expression. Intervention refers to actions seeking to regulate the manner and channel through which expressions are transmitted. Contention means actions geared toward countering the impact of the message, without affecting the message itself, or the manner in which it is transmitted. This may be a positive act, such as subsidizing rivaling speech. An example of suppression would be direct censorship. An example of intervention or restriction would be where and when the message can be transmitted (i.e., not in schools). And an example of contention would be, for instance, government subsidy to feminist porn (as an alternative to banning pornography deemed to be a form of violence against women).

\(^59\) Under Mexican law, only civil marriages confer legal status; thus, the Archbishop can only inhibit, not prohibit, gay Catholics from marrying.
practices reinforce traditional gender roles.

If we frame the issue of religious attire (e.g., headscarves) as one in which freedom of religion confronts gender inequality, then there are two possibilities. Either the title holder of the right in tension with gender equality is a religious community (wherever collective rights are ascribed to such groups); or else the title holder is a woman in so far as she is a member of that religious community. This means that the right is held by the community, as a community, and is protected as long as it conforms to that community’s preexistent internal rules. By contrast, if we cast the question as a matter of free speech on both sides of the equation, then the title holder of the fundamental right is the woman as an individual. Conceptually, this makes her the person entitled to decide whether or how to exercise the right.

If the fundamental right at stake is freedom of religion, then wearing a headscarf is a practice that is both rigid and reified, insofar as it is part of the paraphernalia or practices that are protected because it is embedded in tradition, or according to the religious groups’ rules and hierarchies. Instead, free speech is a practice that is far more ductile: an expressive act emitted by an individual who wishes to convey a message, but can choose to do so in a different manner, through different forms of expression.

As to the effect each framework has with regard to the woman herself, the two could not be more different. Framing the matter as one involving religious freedom requires that the woman conform to the religious practice of her community in order to enjoy constitutional protection. Instead, if the matter is framed from the perspective of free speech then the woman is empowered independently of her community. Discussing her actions as expressions of freedom of religion subsumes the woman into her religious community, making her an instantiation of a group practice and, thus, disempowers her vis-a-vis the group. Her actions are not hers, but the community’s. The community’s rights (practices, beliefs) are protected; the woman is not responsible for her actions but merely an object of the group’s traditions. However, casting her actions as an exercise in free speech, in contrast, simultaneously empowers her vis-a-vis both the state and any other entity—including her religious community—and holds her responsible for such actions. She
chooses how to express her gender role, so she is responsible for such expressions.

In the end, I believe that the most powerful reason to prefer freedom of speech is the same reason why freedom of speech is valuable in the first place: because it provides a minimum safeguard for diversity in collective interaction. When wearing a headscarf is cast as valuable or protected because it is the time-honored religious practice of a group, such action actually contributes to stifling diversity both within and without religious communities. It stifles diversity between religious communities because it requires that either a religious community entrench itself to defend a specific practice or else succumb to the majoritarian perspective (for example, secularism—as in France—or gender equality). Casting the choice to wear a headscarf as a religious practice also stifles diversity within religious communities because, in identifying the practice as belonging to a community, it fixes the practice and protects it only insofar as it is recognized as a collective practice. This view assumes that the message conveyed by the practice is inherently important to the religious community as a whole, disallowing the claim of dissenting messages within the community as legitimately expressing the community’s identity.

In contrast, framing the matter as a question of free speech, by establishing that individual women—not the religious communities themselves—are the title holders of the right in question, protects diversity both within and outside the religious community. It does within the community, because it empowers the individual women as the right holders, and thus, the actors capable of demanding state protection. Outside the community this view protects a specific message regarding gender roles—the roles according to the status quo within the religious community—from being stifled by the broader status quo, which sees the gender roles conveyed and sanctioned by that community as unacceptable. For these reasons, I argue, approaching the tension between gender equality and religious attire is best done through free speech, at least for those of us committed to empowering women.

It is important to keep in mind that the broad doctrinal structures through which we frame specific problems do not
determine specific outcomes. Regardless of how one frames the question, *legally,* the solution to the problem at hand can be constructed so that headscarves can or cannot be banned under law. However, choosing the framework does matter because it determines who the protagonist is—the individual women or the religious community—and what the value at stake is, freedom and diversity, or religion and tradition. While framing is not everything, it can determine much and is especially useful to describe and explain which issues are truly problematic and in what sense.

In the end, the strongest case for protecting women who wear religious attire in public spaces stems from the importance of allowing women themselves to say and do what they feel they should as women. It is a question of taking women at their word, through what they are saying and through their actions. Respecting and taking seriously what say, whatever form that speech takes, is something to which those of us who agree with the fundamental claims of feminism should always be committed.
RELIGIOUS FREEDOM:
EXPRESSING RELIGION, ATTIRE, AND
PUBLIC SPACES

Lucy Vickers*

INTRODUCTION

Religious expression has long been recognized as a fundamental element of the right to religious freedom. Article 18 of the International Covenant on Civil and Political Rights (“ICCPR”) and Article 9 European Convention on Human Rights (“ECHR”) both protect the right to freedom of thought, conscience, and religion, including the right to manifest religion in worship, observance, practice, and teaching. The expression of religion through religious attire is, in many cases, an important aspect of religious observance and practice, and thus comes within the protection of human rights law. As a result, the question of how to respond to religious attire in public spaces has traditionally been considered from a human rights perspective. Issues for debate have included the question of whether a dress

* Professor at Oxford Brookes University.

Everyone shall have the right to freedom of thought, conscience and religion. This right shall include freedom to have or to adopt a religion or belief of his choice, and freedom, either individually or in community with others and in public or private, to manifest his religion or belief in worship, observance, practice, and teaching.

Id. at 178.
code is religiously required, and when the rights of others may prevail over the right of an individual to express their religion. This Article revisits the debate over when religious attire may be restricted in the public space through the alternative, but complementary, perspective of equality. It suggests that viewing this issue through the lens of equality may provide additional insights for these debates, and argues that when assessing the proportionality of any restriction on religious expression, the interest in equality should be taken into account.

The Article focuses on the issue of religious dress as a form of religious expression. It begins by addressing the human rights approach to the protection of religious expression and discusses some of the difficulties which can arise from this approach, particularly with regard to the need to balance conflicting rights. It then discusses the ways in which an approach based on equality may provide additional insights into how to resolve some of those difficulties. It ends with a consideration of the factors that may be used when assessing the proportionality of restrictions on religious expression.

I. HUMAN RIGHTS APPROACHES TO RELIGIOUS EXPRESSION

Viewed from a human rights perspective, religious dress is generally understood as an example of the manifestation of religion. This manifestation of religion and belief is given qualified protection in most human rights regimes. Such a view allows plenty of scope for balancing human rights interests against other rights, such as rights to equality, and it is in this context that the interaction of rights related to religious expression via religious attire and the rights of others has

---

3 The debate over whether practices have to be religiously required or merely religiously motivated is discussed in Arrowsmith v. United Kingdom, App. No. 7050/75, 3 Eur. H.R. Rep. 218 (1978).

4 An example of how the “rights of others” may prevail over the right to religious expression can be seen in Sahin v. Turkey, 2005-XI Eur. Ct. H.R. 175, 206–07 and Dahlab v. Switzerland, 2001-V Eur. Ct. H.R 447, where the equality rights of others prevailed over Sahin’s religious rights.

5 See European Convention on Human Rights, supra note 2, art. 9; International Covenant on Civil and Political Rights, supra note 1, art. 18, § 3.
traditionally been discussed. For example, under Article 9 of the ECHR the protection of religious expression is limited where necessary to protect public safety, order, health or morals, or the fundamental rights and freedoms of others. In the context of religious dress, these limitations could, for example, be used to justify the removal of face veils in order to ensure proper identification at national borders, or to justify restrictions on flowing garments in order to limit infection control in hospitals. Such examples are relatively uncontroversial. More contentious is when religious expression contests the broader category of “the rights and freedoms of others.”

This raises a problem at the heart of religious freedom—whether the positive right to freedom of religion encompasses a corresponding negative right to have no religion or to be free from religion. Recognition of a negative right to religious freedom potentially allows for sweeping restrictions on religious expression. According to this view, religious attire may be restricted in public in order to protect the rights of those who wish to enjoy a public space free from religion and religious symbols.

In human rights law the method developed to manage the conflict between different rights is to undertake a balancing approach to determine where the correct boundaries of protection should lie. Human rights courts are experienced in seeking to resolve conflicting interests this way. Of course, such a balance

---


7 Health and safety reasons were accepted by the European Court of Human Rights as a legitimate aim for restrictions on religious dress in a hospital in Eweida v. United Kingdom, App. Nos. 48420/10, 59842/10, 51671/10, and 36516/10, 2013 Eur. Ct. H.R.

8 European Convention on Human Rights, supra note 2, art. 9.

9 A “negative right” is a right not to do something, which in this case, is a right to be free from religion, or a right not to practice a religion.

10 For examples of the European Court of Human Rights cases discussing an interest in being free from religious influence and referring to the fact that Muslims who wear the headscarf can put under pressure those Muslims who choose not to, see Sahin v. Turkey, 2005-XI Eur. Ct. H.R. 175, 206–07; Dahlab v. Switzerland, 2001-V Eur. Ct. H.R 447.
can be hard to find, especially when human rights conventions are operating on a transnational level. In response, the European Court of Human Rights (hereinafter ECtHR) has developed the notion of the “margin of appreciation,” which allows a degree of flexibility to member states in their observance of the ECHR. This mechanism provides the member states of the ECHR with a “margin of appreciation” in setting the parameters of their domestic law, and states that restrictions will only be found to breach human rights norms when they fall outside this margin.\(^{11}\)

The margin of appreciation has particular significance with respect to freedom of religion cases and a fairly wide margin operates with regard to these cases in Europe, reflecting the lack of consensus about how freedom of religion cases should be treated.\(^{12}\) Indeed, this flexible approach means that there is no uniform approach to religious attire in Europe, despite being governed by a common human rights code. Some member states of the ECHR such as Turkey, France, and Belgium impose significant restrictions on religious attire at work and in the public space more generally.\(^{13}\) However, most other signatory states of the ECHR\(^{14}\) allow religious attire in public, with the UK allowing


\(^{12}\) See Carolyn Evans, Freedom of Religion Under the European Convention on Human Rights 143–44 (2001). For example, in France the public sphere is strictly neutral, whereas the UK and Denmark have established churches.


\(^{14}\) The ECHR signatories include forty-seven member states, twenty-eight of which are members of the European Union. They are: Albania, Andorra, Armenia, Austria, Azerbaijan, Belgium, Bosnia and Herzegovina, Bulgaria, Croatia, Cyprus, Czech Republic, Denmark, Estonia, Finland, France, Georgia, Germany, Greece, Hungary, Iceland, Ireland, Italy, Latvia, Liechtenstein, Lithuania, Luxembourg, Malta, Republic of Moldova, Monaco, Montenegro, Netherlands, Norway, Poland, Portugal, Romania, Russian Federation, San Marino, Serbia, Slovak Republic, Slovenia, Spain, Sweden, Switzerland, the former Yugoslav Republic of Macedonia, Turkey, Ukraine,
perhaps the widest scope for when religious attire can be worn in public, including a range of public employment such as the police and judiciary.\(^{15}\) This broad range of responses to the question of what is the proper scope for restricting religious attire in the public space can serve as an illustration of how the “margin of appreciation” works in that they show that all these different responses, from the most restrictive of religious dress to the most liberal, can be lawful responses to the requirement to provide protection for religious freedom under the ECHR. In effect, as long as the protection for religion does not fall outside the range, or margin, it will remain lawful under the ECHR. These mechanisms enable the ECtHR to support a range of responses with regard to the issue of religious symbols, while maintaining a human rights-based approach to the issue.

An example of this balancing mechanism can be found in the 2005 Grand Chamber judgment in the case of *Sahin v. Turkey*.\(^ {16}\) In this case, a university student objected to the prohibition on religious attire being worn in her university and the ECHR\(^ {17}\) balanced the religious freedom of the student against the Turkish government’s interest in the protection of state neutrality.\(^ {18}\) In reaching the conclusion that the ban was compatible with the ECHR, the court relied on the mechanism of the margin of appreciation and gave a wide margin to the Turkish government to decide whether it was in fact “necessary” in the Turkish political and cultural context to prohibit the wearing of religious symbols in teaching institutions. In the *Sahin* case, the ECHR recognized that the state’s primary interest was the need to protect secularism in the public sphere. Similarly, with regard to the pending case regarding the “burqa ban,” *S.A.S v. France*,\(^ {19}\)

---

\(^{15}\) For an overview of the approaches to this matter in France, England, and the Netherlands, see *Van Ooijen*, supra note 13.


\(^{17}\) For a general introduction to the European Court of Human Rights, see *Theory and Practice of the European Convention on Human Rights* (Pieter van Dijk et al., eds., 4th ed. 2006) [hereinafter *Theory and Practice of ECHR*].


\(^{19}\) *S.A.S. v. France*, App. No. 43835/11; see also Press Release, Grand
the state interest is framed in terms of security, the need to uphold gender equality, and to avoid Muslim women being cut off from society. These interests will need to be balanced against those of the individual women whose freedom of religious expression is severely limited by the ban.

Although the approach to conflicting rights based on the balancing of interests within a margin of appreciation approach is well established, nonetheless profound questions remain about whether the resulting variety of practice across different countries should be acceptable within a single legal framework such as the ECHR. Greater clarity and consistency between different countries is desirable for a number of reasons. A key reason is because religious attire can form a significant element of religious identity, and it is important for individuals to have clarity about the extent to which this element of identity can be expressed in public. If religiously orthodox doctors and nurses, wishing to express religious faith through their attire, can do so in the U.K. but not in other parts of the E.U., then those same doctors and nurses are likely to not exercise the free movement of persons enjoyed by other citizens. Moreover, the need for consistency becomes particularly acute when the interference with religious freedom reaches beyond specific situations of work and education into the public sphere more generally. This can be seen through bans on the wearing of face coverings in public in France, Belgium, and the Netherlands.21

However, there is danger inherent in forced clarity and consistency. Clarity and consistency in the legal treatment of religious attire might seem inherently desirable; certainly clarity


21 See Van Ooijen, supra note 13.
will help both those subject to the law and those enforcing the law to be clear about what can and cannot be worn in public or at work. However, flexibility may be essential given the range of current practices across different jurisdictions and the lack of agreement about what should and should not be allowed in terms of religious dress. Moreover, the flexibility inherent in the balancing and “margin of appreciation” approach in human rights law remains attractive as it allows for a detailed examination of the facts of each case and for its context. One contextual issue that could be helpful in assessing how to treat religious attire is that of equality. In the next section of this article, I turn to consider the matter of religious attire from the perspective of equality, to consider whether arguments used in equality law can provide new ways to approach the debate.

II. AN EQUALITY PERSPECTIVE ON RELIGIOUS ATTIRE

Approaching the issue of religious attire from a perspective of equality will involve considering whether restrictions on attire have a differential impact on individuals on equality grounds. For example, it may be that restrictions impact differently on men and women, or on those of one religion more than another. Where this is the case, it can be seen that restrictions not only have implications for religious freedom but for gender and religious equality too.

One particular concern related to religious freedom which may be viewed differently from an equality perspective is the issue of state neutrality. The need for a neutral public sphere is often viewed as in competition with claims for religious freedom. However, it is important to note when considering restrictions on religious expression from an equality perspective that legal arguments based on the value of “state neutrality” appear in a different light from when they are based on human rights perspectives. In effect, strict state neutrality fails to achieve equality because our social organization can never be truly “neutral.” The reason for this is two-fold. First, a strict neutrality position tends to assume that we can make a clear separation

between the public and private spheres, and that it actually is “neutral” if the public sphere is “neutral.” However, if we consider lessons from equality jurisprudence, we can see that what may look “neutral” can in practice favor the dominant group. For example, in the U.S. it was recognized early on in discrimination jurisprudence that neutral rules can have a disparate impact on disadvantaged groups. The resulting concept of indirect discrimination was imported into UK law in the Sex Discrimination Act of 1975, and then into EU law in the 1980s. Thus, in the context of gender discrimination, it is well established that “neutral” norms tend in practice to be male norms, so that neutral rules requiring, for example, that workers be available to work full time, can be indirectly discriminatory on grounds of gender because fewer women than men can comply with them.

This issue with “neutrality” can clearly be seen in the context of religion and belief, where the social organization is largely “Christian” rather than neutral. This means that most workers enjoy a day off from work on Sundays, as well as time off at Christmas and Easter. Thus religiously observant Christians will rarely come across work rules and dress codes with which they cannot comply, unless they work in an area of work that requires staffing all week. This situation arises because of the historical dominance of Christianity in Europe and the U.S., creating a system where the wider society is organized in ways that are compatible with mainstream Christian practice.

Second, not only have the dominant social organizations adapted to accommodate mainstream Christian practice; but also Christianity itself has, arguably, adapted to the idea of the “secular” or at least “neutral” social model adopted in much of the EU and also in the U.S. This results, again, in Christians

24 Lord Lester of Herne Hill, Equality and United Kingdom Law: Past, Present, and Future, 2001 PUB. L. 77. See also Council Directive 76/207, O.J. L. 39/40 (1976) (“[T]he principle of equal treatment shall mean that there shall be no discrimination whatsoever on grounds of sex either directly or indirectly . . . .”). This directive is often referred to as the Equal Treatment Directive.
being able to comply more easily with workplace rules than individuals of other religions. A simple example illustrates the adaptability of Christianity: early Christians adapted the pagan winter festivals and turned those once pagan festivals into today what is known as Christmas. Arguably this helped the young religion gain acceptance in what was otherwise a hostile environment. Moreover, there is some Biblical authority which supports such an adaptive process: the command to “give to God what is God’s and to Caesar what is Caesar’s”25 provides many Christians a relatively easy method to reconcile civic duty with religious duty.

In addition to these examples of Christianity’s adaptability to secular power, Protestant Christianity in particular has developed a specific theology that makes it adaptive. This involves the idea of the separation between body and mind, with faith more a matter of the mind and its state of “righteousness” than a matter of the body.26 This, again, arguably allows for greater accommodation of secularism within Christianity itself. A full theological discussion of these ideas is beyond the scope of this Article, but put simply, Christian theology is largely based on orthodoxy or “correct belief.” Although debates of course continue within Christianity about the precise relationship between “faith” and “works,” nonetheless, the religion is based less on what the individual does and more on what he or she believes (in religious terms, his or her relationship with God). This means that the focus is on belief, and so dress codes and rules of attire, together with strict dietary rules or prayer rituals, tend to be less central as signifiers of religious observance. In contrast, religions such as Judaism and Islam have a greater focus on orthopraxy: a concern for correct religious practice as signifiers of religious adherence. In these traditions, codes of conduct related to attire, diet, and prayers have greater prominence as means to practice and observe religion. It is these religious practices that can cause conflicts with the secular world.

26 See generally THE OXFORD HANDBOOK OF GLOBAL RELIGIONS (Mark Juergensmeyer ed., 2006).
The greater focus on orthodoxy in Christianity, rather than orthopraxy, may provide an additional explanation of the relative lack of conflict between Christian practice and the secular world.

It is arguable, then, that the adaptive process between the secular Euro-American world and Christianity has been a two-way process. In part, the “world” has adapted to accommodate the dominant religion. One example is recognition of Sunday as the Sabbath day of Christianity and thus generally recognized as a day of rest. But in part, Christianity itself has adapted so that conflicts are reduced: there are few external requirements, such as head covering, for observant Christians to comply with, requirements which might otherwise conflict with secular practice.

If these theological understandings are brought into the debate over equality and religious attire, it becomes clear that so-called “neutral” rules, which prohibit the wearing of religious attire, are doubly non-neutral. First, it is more likely that adherents to minority faiths have rules of observance that are of such religious significance. Second, such rules can have a disparate impact upon adherents of minority faiths because the rules of those minority faiths are less likely to be compatible with mainstream social norms. This means that although rules which restrict religious attire can seem formally neutral, in practice they have a disparate impact on religious minorities.

The importance within religions of “right conduct” explains why some religiously observant individuals may appear to be what has been termed “obdurate believers,” or those who will not yield their religious beliefs or practices to other interests. For such individuals, a choice between removing religious attire and leaving the public space will not result in the removal of religious attire. Thus a seemingly neutral public space will be achieved at the expense of the religious believer, whose belief will remain hidden. In the context of religions based on orthodoxy, a split between private religious observance and public religious neutrality can be understood in terms of separating the personal

---

27 This phrase was coined by Anthony Bradney in his essay, *Faced by Faith*, in *Faith in Law: Essays in Legal Theory* 90, 91 (Peter Oliver et al. eds., 2000).
and the public: what is God’s is private, and can be kept private; what is public (“Caesar’s”) can remain in public. The split remains a reasonably neat solution to the problem of reconciling personal religion with the neutral public space. However, applied to religions of orthopraxy, attempts to exclude personal belief from the public sphere results in the exclusion of the person altogether.

The importance of “orthopraxy” in some faiths explains the inequality that can arise when certain religious practices are restricted in public, and lessons from an equality perspective may be instructive in this context. For example, it has long been recognized that if women’s participation at work is to be increased, then some accommodation of the family is needed. This can be through a workplace nursery provision, maternity leave, or other support for working mothers. If the argument were to be accepted that women’s family responsibilities are private matters, best left in the private sphere, then there would be no need for such workplace provisions. Yet within Europe it has been accepted that such an approach does not lead women to work on equal terms with men. If no accommodation of family life is offered, then many women will not leave their “personal matters” at home and head out to work; instead, they will stay at home. Many of the hard-won workplace rights, such as maternity leave, are predicated on the idea that such practical support is necessary if women are to be able to participate in the workplace. This lesson can be applied outside of the workplace, too: if we wish to include people in our societies, we need to provide some reasonable level of accommodation for their basic needs.

Applied to the ground of religion and belief, this equality-based reasoning suggests that outright bans on religious expression will lead to unequal results. Bans on religious expression will have a disparate impact on minority religious

---

29 See, for example, the extensive workplace protection for gender equality within the European Union, going back to 1957 when the principle of equal pay for equal work became part of the founding Treaty of Rome.
groups; they will lead to the removal of individuals holding these beliefs from the public sphere and will hamper attempts to include such individuals in mainstream society. Thus, it is clear that if inclusion of religious minorities is a society’s aim, then some accommodation of religious practice is essential. If headscarves are banned at work, many Muslim women will not remove the headscarf and go to work; they will stay at home. If turbans cannot be worn by public sector workers, Sikh men will not cut their hair; they will just not work in the public sector. Arguments based on this reasoning apply to the public space as well: if face coverings are banned in the public sphere, those women who wear them will in effect be excluded from the public sphere.

These arguments, based on the perspective of equality law, demonstrate that what can look like simple neutrality may not, in practice, be experienced as neutral. The public sphere is not as neutral as might at first be supposed and the religions themselves are not equally placed in relation to the public space, meaning that similar treatment of religious individuals will not result in all of those individuals having the same experience. Religious groups are not alike, and equal treatment by way of the neutral public space will not result in “like” experiences. It is clear that policies which do not accommodate religious differences ensure there is no equal participation in public life; instead there is exclusion. Thus, if we exclude the personal from the public sphere, we exclude the person as well.

These equality-based perspectives suggest that an absolute ban on religious attire in the public space fails to give sufficient recognition to the interests of religious individuals. They suggest that the better legal response is to ensure, instead, that there is consideration of the nuances and complexity involved in regulating religious practice. As has been argued above, responding to the plural religious landscape by creating a purely secular public space fails to recognize the deeply unequal way this would affect religious minorities. Instead, a more plural public space is required, with room for the religious and the secular to coexist, and even to engage in dialogue with one another.

To suggest a plural public space, however, is not to suggest
that restrictions on religious freedom can never be imposed. Human rights protection for religion does not require absolute protection for religious practice. It does require, however, that any restrictions on religion have a legitimate aim, and that the restrictions on religious practice remain proportionate to that aim. Proportionality requires that no more be done to restrict the religious practice than is needed to achieve the legitimate aim.\textsuperscript{31} The second part of this paper considers how and when restrictions on religious expression, through religious attire, may be justified, viewed through the equality context.

III. WHEN WILL RESTRICTIONS ON RELIGIOUS ATTIRE BE JUSTIFIED?

The ECtHR has heard a number of human rights cases (some referred to above),\textsuperscript{32} all involving challenges to the prohibition of religious dress at work. In these cases the prohibitions on religious dress at work have been upheld.\textsuperscript{33} In Dahlab \textit{v. Switzerland}, for example, the ECtHR held that the prohibition of the headscarf imposed on a Muslim teacher of young children was proportionate because the teacher had influence on the intellectual and emotional development of children.\textsuperscript{34} The court also took into account the fact that, as a public sector employee, the teacher was a “representative of the state.”\textsuperscript{35} The court also mentioned in its reasoning the fact that the headscarf is “hard to square with the principle of gender equality.”\textsuperscript{36} In \textit{Sahin v. Turkey}\textsuperscript{37} the court balanced the religious freedom of a student against the Turkish government’s interest in the protection of

\textsuperscript{35} \textit{Id. at} 462.
\textsuperscript{36} \textit{Id. at} 463.
state neutrality, in holding that the restriction on wearing a headscarf was justified. The court also noted the government’s argument that the wearing of a headscarf may put other students under pressure to adopt more fundamentalist approaches to their faith.

However, viewed from an equality perspective, and drawing on the insights discussed above, it can be strongly argued that these decisions fail to respect the equality interests of religious minorities. Instead, a more sensitive approach to justification is needed; one that takes into consideration the wide range of factors involved in cases involving religious expression through attire and other symbols.

An example of a more sensitive consideration of the factors that can be relevant when considering restrictions on religious attire can be seen in the case of Azmi v. Kirklees Metropolitan Borough Council. Azmi was a teaching assistant who wanted to wear the niqab when in the presence of male colleagues. She was dismissed for refusing the employer’s request to remove the niqab when assisting in class. Her initial claim to the court alleging direct and indirect discrimination was unsuccessful. The court did accept that there was prima facie indirect discrimination since the refusal to allow Azmi to wear the niqab put her at a particular disadvantage when compared with others, but the indirect discrimination was justified. The court held that the restriction on wearing the niqab was proportionate given the need to uphold the interests of the children in having the best possible education. What is interesting about the case is that the

---

39 A niqab is a face-covering for women that veils the face and hair down to the shoulders, with a small opening for the eyes.
41 Id. at 1161.
42 She also claimed victimization and was successful due to inadequacies on the part of the employer in dealing with her case. See id. at 1155.
court noted\textsuperscript{45} that the school had performed a thorough investigation before reaching the conclusion that the restriction was necessary. For example, it undertook a review of Azmi’s teaching to see if the quality of teaching was reduced when Azmi wore the face covering and came to the conclusion that it was; the school also investigated whether it was possible to rearrange Azmi’s timetable to enable her to assist only in classes with a female teacher and found that this was not possible.\textsuperscript{46} In relation to the question of justification, Azmi however argued that insufficient effort had been made to try to accommodate her religious requirements; for example, the school could have tried to assess alternative ways to improve her communication and performance when wearing the niqab.\textsuperscript{47} The court, however, held that the school had sufficiently shown that the restriction on wearing the niqab was proportionate to the school’s aim of providing effective education for the students.\textsuperscript{48}

The Azmi case illustrates how a proportionate response to what might otherwise be indirectly discriminatory can require a careful review of the facts and circumstances of the case. This “fact-sensitive” approach can enable a full analysis to be undertaken to determine whether any accommodation of the religious practice can be achieved without compromising the competing interests at stake. In Azmi, the court found that the competing interest in maximizing the children’s educational experience could not be achieved if the required accommodation was given.\textsuperscript{49} It is noteworthy that in Azmi the religious practice involved the covering of the face, which was found to impede communication in a context where non-verbal communication is essential. It is quite possible that in other cases, where the religious practice does not directly affect the purpose of employment, some accommodation of religious practice may be required in order to avoid the disparate impact that such restrictions can have on religious minorities.

A similar fact-sensitive approach can be seen in \textit{R (on the

\begin{itemize}
\item \textsuperscript{45} \textit{Id.} at 1172.
\item \textsuperscript{46} \textit{Id.} at 1169–70.
\item \textsuperscript{47} \textit{Id.} at 1160.
\item \textsuperscript{48} \textit{Id.} at 1169–70.
\item \textsuperscript{49} \textit{Id.} at 1172.
\end{itemize}
The case involved a pupil’s freedom of religious expression that was in conflict with the rights of others; here the right of the school to determine the dress code for the school. The school had a uniform which prohibited a female student from wearing the *jilbab*. The reason for imposing the uniform was to promote harmony between the different races, religions, and cultures represented in the school, and to foster a sense of cohesion and community within the school. There had been some history of conflict between pupils in the past, with pupils defining themselves along racial lines, and the school viewed the uniform as necessary as a way to combat these problems and to prevent the development of sub-groups identified by dress. In the case, the English House of Lords was asked to review a school’s decision not to allow Begum to attend school wearing the *jilbab*. As with Azmi, the court upheld the decision of the school, but only after a careful, fact-intensive review. The court recognized that the school had undertaken detailed discussion in reaching its decision. The school had consulted local religious leaders and had uniform requirements which met with common Islamic dress codes, in that it allowed for several uniform options, one of which was a salwar kameez which could be worn with the school tie and school jumper. The court recognized that the school’s decision creating the uniform requirements had been discussed fairly carefully beforehand. The House of Lords also noted that there was evidence that the school had previously suffered the ill effects of groups of pupils defining themselves along racial lines, with consequent conflict between them. Thus, based on the facts and due to the careful appraisal at the local

---


51 A loose fitting garment which hides the contours of the body, associated with Muslim women.

52 *Id.* at [18].

53 *Begum*, [2006] UKHL 15 at [40]-[41], [71], [91], [99].

54 *Id.* at [33].

55 A sleeveless smock-like dress with the loose trousers.

56 *Begum*, [2006] UKHL 15 at [34].

57 *Id.* at [18].
level, the House of Lords upheld the restrictions on religious attire imposed by the school since the restriction struck a proportionate balance between the conflicting interests at stake in the case.\textsuperscript{58}

Of course it will always be arguable that the court could have reached a different conclusion: it may be that Begum could have been accommodated without undue harm to others.\textsuperscript{59} However, although one might disagree with the conclusion reached in the case, it is clear that the court’s use of fact-based decision making allowed for a more contextual and sensitive decision. This type of decision making allows space in the legal framework for the complexity of the issue to be considered. This way the decision making process includes a full examination of religious freedom and equality concerns.

The benefit of submitting any prohibition on religious attire to the test of proportionality, assessed in light of a detailed factual examination, as was done in \textit{Azmi} and \textit{Begum}, is that a wide range of factors can be taken into account to decide the legality of a prohibition. This allows for a nuanced examination of the facts, which reflects the context of the prohibition, the rights of others such as pupils or colleagues, additional options that may be open to the individual, and the practical effect of any gender-based claims. For example, restrictions on headscarves on the basis that men impose them on women, or that headscarves create social pressure on others to conform, should be tested empirically. There is evidence that the courts’ assumptions that headscarves are worn by Muslim women because their male relatives force them to do so is incorrect in many cases.\textsuperscript{60} This is not to deny that in some instances this may happen, but, equally, legal policy should not be made on the assumption that this is usually the case. A model of legal protection based on a detailed review of

\textsuperscript{58} Id. at [68], [98].


the facts and circumstances surrounding each individual case enables courts and tribunals to reach reasoned decisions that are both flexible and responsive to the complexity of the issues involved.

While there are many benefits to a factually sensitive review as a model of protection for religious attire in the public sphere, such an approach does have some drawbacks. In particular, it can lead to uncertainty and inconsistency of approaches between different courts and different contexts. This can make it difficult to predict with certainty how any individual case will be resolved. For example, in Begum, the House of Lords stated:

> It is important to stress at the outset that this case concerns a particular pupil and a particular school in a particular place at a particular time. It must be resolved on facts which are now, for purposes of the appeal, agreed. The House is not, and could not be, invited to rule whether Islamic dress, or any feature of Islamic dress, should or should not be permitted in the schools of this country.\(^{61}\)

The court is extremely clear that it is not setting precedent for how other schools should determine the issue of uniform and Islamic dress. Instead, the judgment focuses on the process by which the issue of religious attire should be determined: that the decision should be made with due acknowledgment of the impact any decision may have on religious freedom.

An additional concern with such a fact sensitive review is that the many and varied factors identified as relevant in the consideration of proportionality on any restriction of religious expression may create a false sense of objectivity, masking the fact that the judgment is ultimately personal and subjective. This potentially runs the danger of perpetuating precisely the disadvantage that the creation of legal protection for religious interests should prevent, namely the dominance of minority religious interests by the majority.

The concern over undue subjectivity is a powerful argument, but while fact-sensitive review may never be fully objective, neither is it fully subjective. The factors to determine the proper

\(^{61}\) Begum, [2006] UKHL 15 at [2].
boundaries of religious expression when balanced against other concerns, are not drawn at random but are chosen as a result of careful consideration of the range of competing interests at stake. This includes the extent to which any claims are empirically valid, and the theoretical reasons for protecting the competing interests at all. Not every interest will be relevant. Thus, this fact-based proportionality approach relies on reasoned and principled analysis to determine which factors are relevant.

Moreover, this proportionality approach allows room for any decision relating to religious attire to be tested: any decisions reached must be open, and the factors which were relevant subject to review. Although ultimately courts may allow for some flexibility in the exercise of any discretion by decision makers such as schools or employers, this approach allows for challenges to be made if an important factor has been left out of the equation. Thus, a determination that an individual cannot wear religious attire at work or in school must be proportionate; it must take into account not only the needs of the business or school, but also the individual’s interest in freedom of religious expression. Where the religious interests of the employee have not been taken into account, this may mean that a decision can be challenged on the basis that it is disproportionate.

In sum, although there may be strong equality-based reasons to favor a plural public space, some restrictions on religious freedom will inevitably be necessary. Subjecting any proposed restriction on religion to a fact-sensitive proportionality review should mean that contextually sensitive decisions can be reached, with full account taken of relevant equality concerns.

**CONCLUSION**

An approach to the question of when and to what extent a person should enjoy freedom of religious expression via their choice of attire can be considered both from a human rights perspective and from an equality perspective. Of course these two perspectives are inherently linked, but nonetheless are different. The consideration of equality concerns in this context serves to highlight the need for sensitive responses to calls to restrict religious attire in the public sphere. Without a clear
understanding of the equality dimensions to the debate, questions about the role of religious attire may be resolved merely from the point of view of competing interests in religious freedom: the balance being between the right to freedom of religious expression and the rights of others to be free from religious symbols, particularly in the public sphere. Moreover, the use of the margin of appreciation in European human rights law means that the final standard of review on any restrictions of religious attire is weak: restrictions are effectively assessed against a “norm-reflecting” standard. This means that the case law under ECHR tends to accept current standards of protection for religion, even where standards are fairly low, rather than engaging in the setting of high standards of rights protection.

One of the reasons for weak protection for religious claims in the public space has been the competing interest in having a religiously neutral or secular public sphere. Yet while calls for a secular public space certainly have validity, when revisited in the light of the concerns that can be raised from an equality perspective, such claims lose some of their force. The recognition that secular or neutral public spaces lead to unequal outcomes for different, and usually minority, religious groups means that additional factors need to be taken into account when balancing competing rights. I propose that, when assessing the proportionality of any restrictions on religious attire, the interest in equality needs to be added to the balance, and an approach allowing for the setting of standards needs to be used, rather than the norm-reflecting margin of appreciation. With the recognition that unequal results can arise from a reliance on neutrality in the public sphere, it becomes clear that outright bans on the wearing of religious attire in the public sphere are unsustainable. This is not to say that more limited restrictions will never be allowed: as the discussion of the cases of Azmi and Begum illustrate, there remains scope for religious attire to be restricted, but only after careful review of the facts of the case. This more fact-sensitive review allows for the complexity of the issues surrounding religious equality and religious expression to have its proper space in the legal discourse.
(WHEN) CAN RELIGIOUS FREEDOM JUSTIFY DISCRIMINATION ON THE BASIS OF SEXUAL ORIENTATION?—A CANADIAN PERSPECTIVE

Noa Mendelsohn Aviv*

INTRODUCTION

Is it justifiable to ban LGBTQ-positive resources from a public school classroom because of the religious views of some parents?¹ Should bed and breakfast owners be permitted, on the basis of their religious beliefs, to cancel the room reservation of a gay couple?² What about a printer whose religious beliefs are in conflict with the material he’s being asked to produce for the Gay and Lesbian Archives?³ And how should we respond to marriage commissioners, acting on behalf of the province, who refuse to perform same-sex marriages as to do so would violate their religious beliefs?⁴ These are some of the questions that have faced Canadian courts and human rights tribunals in the past number of years.

The underlying question in these cases asks: in what circumstances, if ever, will a service provider’s beliefs justify exempting them from the duty to provide services without discrimination on the basis of sexual orientation or gender

* Noa Mendelsohn Aviv is Director of the Equality Program at the Canadian Civil Liberties Association (CCLA). The author thanks Cara Zwibel and Richard Moon for their time and insights. The opinions expressed in this paper are those of the author.

² See Eadie v. Riverbend Bed & Breakfast, 2012 BCHRT 247 (Can.).
⁴ See In re Marriage Comm’rs Appointed under the Marriage Act, 2011 SCKA 3 (Can.).
identity? The conflict in these “belief-based exemption cases,” as they will be referred to in this Article, arises in the clash between two fundamental and constitutionally protected rights, freedom of religion and equality—a clash that does not readily lend itself to reconciliation.

Conflicts between fundamental rights are never easy, in particular when they elicit highly emotive topics that touch on deeply held fundamental beliefs. And in Canadian law, there is a well-established principle that when it comes to the fundamental rights and freedoms protected in the *Canadian Charter of Rights and Freedoms* (the “Charter”), there is no hierarchy: rather than staking out a “trump” right that will always prevail, the courts are required to judge each case in its specific context. This has not been an easy task for the adjudicative bodies charged with deciding the belief-based exemption cases.

This Article provides a critical analysis of four belief-based exemption decisions in Canada and considers what lessons (and cautionary tales) can be learned from them to help resolve future such cases. These lessons include the following: the issues are complex, and as such, cannot be resolved in the abstract. Such cases must be resolved in context on a case-by-case basis in consideration of the evidence before the adjudicating body. A Canadian Charter section 1 analysis may be particularly helpful in this analysis. Solutions will likely be difficult, and one fundamental right or the other may be violated. In addition, adjudicators should take heed of their own prejudices and avoid perpetuating in the courtroom the kind of discrimination at play in society. Likewise, courts should be aware of their own biases and recognize the genuine issues and rights at stake on both sides of the conflict. Both equality and freedom of religion are fundamental rights in Canada. For many individuals and groups, their religious convictions underlie a strong belief in the inherent dignity, worth, and equality of all people. For others, their religion includes beliefs about proper conduct and practice and


how to interact with those who do not conform to these standards. This may take the form of denying service to LGBTQ individuals, for example, by denying them services. A secular legal system must continue to recognize the sincerity of religious beliefs, even if many in our society take issue with the content of these beliefs. On the other side, discrimination on the basis of sexual orientation and gender identity is a current as well as a historic reality, and strong equality protections are critically needed. Therefore, another principle that emerges from the case law is the danger of creating sweeping exemptions or ex ante policies that allow, legitimize, and perpetuate such discrimination. Finally, in light of the breadth of potential exemptions and the impact they would have, exemptions that allow service providers to discriminate against LGBTQ people, if allowed at all, should be strictly exceptional.

Given that analyses of belief-based exemptions must be made in context, the scope of this Article will be limited to discrimination on the basis of sexual orientation in the provision of services. In addition, given the case law, the analysis will focus on exemptions grounded in religious beliefs, as opposed to beliefs based on personal convictions and conscience.

Part I provides a brief overview of statutory and constitutional protections for equality and religious freedom in Canadian law. Part II discusses four belief-based exemption cases from Canada, offering a critical analysis of the central issues, while also drawing out useful discussions and conclusions, and pointing out dangers to be avoided in future cases of this nature. Part III summarizes some of the central principles discussed, that may prove helpful in considering belief-based exemptions.

I. LEGAL BACKGROUND: PROTECTION OF FREEDOM OF RELIGION AND EQUALITY IN CANADIAN LAW

Both the right to freedom of religion and the right to equality are constitutionally protected in Canada, as both are guaranteed in the Canadian Charter of Rights and Freedoms. The Charter sets out the fundamental rights and freedoms protected under the Canadian Constitution (of which the Charter is a part).

Section 1 of the Charter simultaneously guarantees Charter
rights whilst also providing for reasonable limits on those rights in limited circumstances, as follows: “The Canadian Charter of Rights and Freedoms guarantees the rights and freedoms set out in it subject only to such reasonable limits prescribed by law as can be demonstrably justified in a free and democratic society.”

This reasonable limits clause allows legislation and government action to limit Charter rights, but only if the limit is for a pressing and substantial objective, the means chosen by the law or action are rationally connected to this objective, the limit is minimally impairing, and the limit is proportional in that its deleterious effects do not outweigh its salutary ones. Whether or not a limit is reasonable must be judged in its context.

As to the substantive rights at issue in the belief-based exemption cases: Section 2 of the Charter establishes the right to freedom of religion and conscience (among others) as follows:

(2) Everyone has the following fundamental freedoms:
(a) freedom of conscience and religion;
(b) freedom of thought, belief, opinion and expression, including freedom of the press and other media of communication;
(c) freedom of peaceful assembly; and
(d) freedom of association.

Section 15 of the Charter sets out the equality guarantee. Section 15(1) provides: “Every individual is equal before and under the law and has the right to the equal protection and equal benefit of the law without discrimination and, in particular, without discrimination based on race, national or ethnic origin, colour, religion, sex, age or mental or physical disability.”

While the Charter protects a full array of fundamental rights, it also provides for reasonable limits on those rights in limited circumstances, as follows: “The Canadian Charter of Rights and Freedoms guarantees the rights and freedoms set out in it subject only to such reasonable limits prescribed by law as can be demonstrably justified in a free and democratic society.”

This reasonable limits clause allows legislation and government action to limit Charter rights, but only if the limit is for a pressing and substantial objective, the means chosen by the law or action are rationally connected to this objective, the limit is minimally impairing, and the limit is proportional in that its deleterious effects do not outweigh its salutary ones. Whether or not a limit is reasonable must be judged in its context.

As to the substantive rights at issue in the belief-based exemption cases: Section 2 of the Charter establishes the right to freedom of religion and conscience (among others) as follows:

(2) Everyone has the following fundamental freedoms:
(a) freedom of conscience and religion;
(b) freedom of thought, belief, opinion and expression, including freedom of the press and other media of communication;
(c) freedom of peaceful assembly; and
(d) freedom of association.

Section 15 of the Charter sets out the equality guarantee. Section 15(1) provides: “Every individual is equal before and under the law and has the right to the equal protection and equal benefit of the law without discrimination and, in particular, without discrimination based on race, national or ethnic origin, colour, religion, sex, age or mental or physical disability.”

While the Charter protects a full array of fundamental

---

8 See, e.g., Hutterian Brethren of Wilson Colony v. Alberta, 2009 SCC 37, para. 186 (Can.).
9 See, e.g., Toronto Star Newspapers v. R., 2010 SCC 21, para. 3 (Can.).
11 Id. § 15(1).
rights—such as the rights to life and liberty, freedom of expression, and freedom of association—across Canada, there are also quasi-constitutional provincial, territorial, and federal “human rights” statutes whose focus is the prohibition against discrimination in such areas as housing, employment—and significantly for this Article—the provision of services. In other words, and to avoid semantic confusion, “human rights” in many Canadian jurisdictions is sometimes understood in its legal meaning as the specific right to be free from discrimination. And human rights tribunals are for the most part established pursuant to the aforementioned human rights statutes (not the Charter) to adjudicate complaints of discrimination under these statutes.

Thus, belief-based exemption cases may be decided under the Charter and resolved through a reasonable limits test under section 1, or they may be decided under the human rights laws.

II. CANADIAN JURISPRUDENCE—FOUR BELIEF-BASED EXEMPTION CASES AND WHAT THEY CAN TEACH US

A. Ontario Human Rights Commission v. Brockie.\textsuperscript{12} Can a Commercial Printer Refuse a Printing Job On the Basis That Its LGBTQ-Positive Content Violates His Belief?

In April 1996, Mr. Ray Brillinger went into a commercial print shop on behalf of the Canadian Gay and Lesbian Archives (“Archives”) and asked the printer to print blank letterhead and envelopes for the Archives, as well as some business cards for its officers.\textsuperscript{13} The text on the materials noted that the Archives “represented [the] interests of ‘gays’ and ‘lesbians’ but said nothing of [its] objects, activities or membership.”\textsuperscript{14} Without inquiring into these matters, Mr. Brockie, the president of the print shop (the printer), would not provide this service and later attempted to justify his refusal on the basis of the Charter right to freedom of religion.

The evidence before the Ontario Human Rights Board (the

\textsuperscript{13} Id. at para. 6.
\textsuperscript{14} Id.
“Board”) included Mr. Brockie’s testimony as to his religious beliefs, including a belief that “homosexuality is detestable” and that “providing printing services to [LGBTQ] organizations would be in direct opposition to his belief.” The printer had previously done work for LGBTQ customers and for a company which “produces underwear marketed to the gay male population,” but argued that this was different since, in his view, the Archives were promoting the “homosexual lifestyle.” The Board decided against Mr. Brockie on the basis of the significant social and historical discrimination faced by LGBTQ individuals, the economic and psychological impact of this discrimination, and the fact that Canadian society had decided to protect LGBTQ people from discrimination. The Board found that Mr. Brockie would still be free to hold and practice his beliefs within his home and Christian community, just not by denying service to one group in the public marketplace. In the result, the Board made two orders against the printer. It ordered him and his company to pay damages to Mr. Brillinger and the Archives. And it ordered the printer in the future to provide printing services to LGBTQ people and to organizations that exist for their benefit.

On appeal, the printer claimed that this decision by the Board violated his right to freedom of religion under section 2(a) of the Charter and under section 15 as a violation of his right to be free from discrimination on the basis of religion. He argued that his dignity would be demeaned by being “conscripted to support a cause with which he disagree[d]” on the basis of a sincere religious belief. This, in his view, should confer a “defence to discrimination” and a “right of dissent.”

The court hearing the appeal presented this case as a “conflict of dignities,” citing from the preamble to the Ontario Human Rights Code (the “Code”), a statute dedicated to promoting

---

15 Id. at para. 15.
16 Id. (internal quotation marks omitted).
17 Id. at para. 17.
18 Id. at para. 37.
19 Id. at para. 19.
20 Id.
21 Id. at para. 20.
22 Ontario Human Rights Code, R.S.O. 1990, c. H.19, pmbl (Can.).
equality and prohibiting discrimination, as follows:

(a) recognition of the inherent dignity and the equal and inalienable rights of all members of the human family is the foundation of freedom, justice and peace in the world and is in accord with the Universal Declaration of Human Rights as proclaimed by the United Nations; and

(b) it is public policy in Ontario to recognize the dignity and work [worth] of every person and to provide for equal rights and opportunities without discrimination that is contrary to law, and having as its aim the creation of a climate of understanding and mutual respect for the dignity and worth of each person so that each person feels a part of the community and able to contribute fully to the development and well being of the community and the Province.23

The interesting question raised in this case asked: should there be an exemption for a service provider who did not refuse to serve LGBTQ individuals, but rather refused to produce content that ran directly counter to the service provider’s own beliefs? The court on appeal answered in the affirmative, while still finding against Mr. Brockie with respect to the particular facts at issue. The court upheld the Board’s specific order against the printer requiring him to pay damages for refusing to print the requested letterhead, envelopes and business cards. However, the court modified the Board’s general order that would have required him and his company to serve LGBTQ people and LGBTQ-positive organizations in the future. Instead, the court held that in the future, the printer would not be required “to print material of a nature which could reasonably be considered to be in direct conflict with the core elements of his religious beliefs or creed.”24 Unfortunately, the court’s formulation is unworkable, and it opens the door to many forms of unacceptable discrimination.

23 Id.
The Issues: Who is a Person, Producing Content that Conflicts with One’s Beliefs, and the Slippery Slope of Exemptions

As a preliminary matter, the court briefly considered the question of who is a “person” under the Code—for the purpose of bringing a discrimination claim, being the subject of a claim, or raising the right to religious freedom—and whether these would include organizations and corporations.\textsuperscript{25} The court found that the term “person” could include a corporation responsible for discriminating.\textsuperscript{26} Likewise, organizations and corporations were able to claim that they are the object of discrimination, as this is consistent with the Code’s purpose, and would allow those suffering from discrimination to act in association with others.\textsuperscript{27} However, when it comes to the discriminator, the Court found that a corporate entity could not assert a Charter right, such as the right to freedom of religion.\textsuperscript{28} This finding may be helpful in other exemption-based belief cases. It was, however, of no practical import in Brockie, as Mr. Brockie was able to raise his own individual Charter rights.

As to the main issue concerning the content of the requested service, Mr. Brockie argued that there should be a distinction between a refusal to provide service because of the customer’s human characteristic, here his sexual orientation, and a refusal aimed at a person engaged in the political act of promoting the cause of those with such characteristics. The court rejected this argument as specious and irrational.\textsuperscript{29} The court stated that no authorities had been cited to support such a distinction, and concluded that “efforts to promote an understanding and respect for those possessing any specified characteristic should not be regarded as separate from the characteristic itself.”\textsuperscript{30} This conclusion is similar to the Canadian courts’ consistent rejection

\textsuperscript{25} Id. at para. 24.
\textsuperscript{26} Id.
\textsuperscript{27} Id. at para. 26.
\textsuperscript{28} Id. at para. 39.
\textsuperscript{29} Id. at para. 29.
\textsuperscript{30} Id. at para. 31.
of attempts to distinguish between the identity and behaviour of LGBTQ people\(^{31}\) (as discussed below) but goes even further. The Court held that not only are individuals protected from discrimination in relation to who they are and what they do, but they are also protected in their endeavour to seek understanding and respect for themselves.

The court then considered whether the Board, in making its order against the printer, had exercised its discretion in a manner consistent with the \textit{Charter}.\(^{32}\) This order had not only required Mr. Brockie to pay damages to Mr. Brillinger and the Archives, but had also required that in the future, the printer would have to provide printing services to LGBTQ people and to organizations promoting their interests. All of the parties (and two of the intervenors) conceded that the Board’s decision infringed Mr. Brockie’s freedom of religion as it would force him to act in a manner contrary to his beliefs.\(^{33}\) The question at issue was, therefore, whether this infringement was justified under section 1 of the \textit{Charter}.\(^{34}\)

In its section 1 analysis, the court considered whether the Board’s order was rationally connected to its objective.\(^{35}\) The court distinguished between the activity in issue—the printing of materials such as letterhead and business cards—and a hypothetical situation involving the printing of materials with more editorial content.\(^{36}\) The latter materials, in the court’s view, could espouse “causes or activities clearly repugnant to the religious tenets of the printer.” Since the objective of the Code is to prohibit discrimination on the basis of certain characteristics, and to encourage equality, the court held that an order prohibiting more than discrimination may not be rationally connected to its objective, and even if so, would be unconstitutional.\(^{37}\)


\(^{32}\) Brockie, 161 O.A.C. 324 at para. 36.

\(^{33}\) \textit{Id.} at para. 40.

\(^{34}\) See supra notes 7–9 and accompanying text.

\(^{35}\) Brockie, 161 O.A.C. 324 at paras. 45–56.

\(^{36}\) \textit{Id.}

\(^{37}\) \textit{Id.} at para. 49.
With respect to the minimal impairment branch of the section 1 analysis, the court found that: “[s]ervice of the public in a commercial service must be considered at the periphery of activities protected by freedom of religion,” and that limits to this freedom may be justified where the exercise of this freedom causes harm to others. Nonetheless, the court held that the general order was not minimally impairing, as the Board could have achieved its goals without intruding to the extent it did on Mr. Brockie’s freedom.  

Finally, the court upheld the damages order against Mr. Brockie for his refusal to print the letterhead, business cards and envelopes at issue. However, the court modified the Board’s general order concerning future print jobs, creating a new standard and order according to which, the printer and shop would not be required “to print material of a nature which could reasonably be considered to be in direct conflict with the core elements of his religious beliefs or creed.” The court offered two contrasting examples to illustrate how, in its view, this would work: (1) if the printing project contained material that proselytized and promoted the “gay and lesbian lifestyle” or that mocked Mr. Brockie’s religious beliefs, this material may be found “in direct conflict with the core elements of his religious beliefs,” and (2) if the material to be printed contained a directory of goods and services of interest to the LGBTQ community, this material may be held as not “in direct conflict with the core elements of Mr. Brockie’s religious beliefs.”  

Presumably, then, the Court viewed the letterhead, envelopes and business cards as falling into the second category.

The court in Brockie faced a difficult issue—how to uphold the duty to provide services without discrimination, while also recognizing the position of a service provider whose beliefs run counter to the material they are being asked to produce. This conundrum could also, in some respects, be stated in reverse (although the Court did not do so): what if it had been Mr.

---

38 Id. at paras. 51–52.
39 Id. at para. 58.
40 Id. at para. 56.
41 Id.
Brockie who had walked into the print shop of Mr. Brillinger, asking on behalf of Mr. Brockie’s Church to print a brochure containing anti-LGBTQ Biblical passages and a call-out to LGBTQ people to attend this Church? Does the right to equality in the public domain always require a service provider to produce material regardless of its content? The court’s response attempted to create an objective standard according to which the duty to provide services to the public without discrimination would generally be upheld, while exempting the printer if the material to be printed was in direct conflict with the core elements of his religious beliefs or creed. This standard is problematic on a number of levels.

First, the idea of an objective standard to assess belief systems is unfeasible. In Brockie, the court’s conclusion—and the “objective standard” it relied upon—was that the printing of the letterhead, business cards and envelopes was not in conflict with Mr. Brockie’s core beliefs. This was based on a legal fiction. Not only was no evidence produced to support this conclusion, but it appeared to contradict the facts that were established in the case. The court cited evidence showing that Mr. Brockie had been willing to do business with LGBTQ people, as well as with a company whose underwear was marketed towards gay men. Mr. Brockie’s refusal to do business with Mr. Brillinger, then, appeared to have been based precisely on the content of the materials. Given that Mr. Brockie chose to turn down business, and potentially alienate Mr. Brillinger, the Gay and Lesbian Archives, and possibly other customers as a result of his refusal, it seems at least plausible, if not likely, that Mr. Brockie refused to print the material because it was in direct conflict with his core beliefs. The court’s “objective” standard is not helpful in

---

42 If the discrimination had been against a church, the analysis would be different. As discussed supra Part I, an analysis concerning discrimination on the basis of sexual orientation must be viewed in light of its current and historical contexts. Discrimination on the basis of religion and creed will raise its own issues, for example: whether creed includes nonreligious beliefs based on a person’s conscience; and whether disapproval of certain religious views or practices ought to be considered discrimination. These issues are beyond the scope of this Article.

43 See supra notes 36–41 and accompanying text.
clarifying this situation.

Second, belief and practice are highly personal, a principle well established by the Canadian courts. While Canadian courts do utilize certain objective standards with respect to religious beliefs—such as whether the infringement of these beliefs is trivial or insubstantial—it should be difficult for a court in some circumstances to insert an “objective” standard without supporting evidence to establish that a belief is not a core element. This is especially true if a savvy service provider has testified that avoiding the promotion of certain behaviours or ideas is central to their religious and spiritual integrity. Supporting evidence on whether or not a particular element or belief is important to a person may come in many forms, such as the individual’s testimony as to what impact would result from a rights violation, and evidence concerning the consistency of their behaviour, though these examples raise their own challenges.

Third, the court’s “objective” standard—that would exempt a service provider from the duty to provide service without discrimination if the product is in direct conflict with the core elements of their religious beliefs—could also create a slippery slope leading to countless additional denials of service. This point was effectively made by the concurrence in *Marriage Commissioners.* In the context of discrimination on the basis of sexual orientation, belief-based exemptions could be claimed by a wedding planner asked to organize a same-sex wedding, or anyone associated with the wedding industry from the caterer to the receptionist working for the dress-maker. An architect asked to design a family home or bedroom could refuse, as could anyone else in the building industry. An individual involved in service or hospitality, such as a room service waiter or a concierge asked for the location of a romantic restaurant might feel the same urge to refuse. And the same may be said for any person providing services to support the couple or family’s life as a couple or family.

---

46 See *infra* Part II.D for further discussion.
Indeed, the logic behind the court’s examples, suggesting that Mr. Brockie may not refuse to print neutral, LGBTQ material, but may refuse to print material that promotes the “gay and lesbian lifestyle,” could justify many refusals as described in the above paragraph, all of which involve the service provider arguably promoting or contributing to said “lifestyle.” In addition, the court’s attempted distinction between LGBTQ value-neutral content and LGBTQ promotional material bears a striking resemblance to the distinction that the court had earlier rejected between discrimination on the basis of sexual identity and discrimination on the basis of sexual behaviour.

A more useful standard may nonetheless be derived from one of the examples provided by the court. The court had suggested that it may be permissible to exempt Mr. Brockie if the brochure mocked his religious beliefs. Given the danger of creating a slippery slope and overly broad exemptions, this Article would narrow the court’s example still further and consider permitting an exemption for a service provider who refuses to produce material that directly fosters hate towards the service provider (and/or towards a group protected under the antidiscrimination laws). Thus, if Mr. Brillinger had been the service provider and had refused to print the above-mentioned hypothetical brochure containing anti-LGBTQ Biblical passages, he might have been justified in this refusal. Likewise, Mr. Brockie might be justified if he refused to print a brochure stating, for example, that any church not recognizing LGBTQ rights is Satanic.

It should be noted that the issue of refusals on the basis of content is limited in scope. Human rights laws prohibit only discrimination on the basis of particular grounds (such as race, gender, creed, and sexual orientation). In all other contexts, a service provider is free to refuse to produce material that has a message with which they disagree, as long as the message is not a proxy for the protected group.

To conclude, the facts in Brockie present a useful basis for considering discriminatory refusals involving content that violates

---

47 See supra note 35–41 and accompanying text.
48 See supra note 31 and accompanying text.
49 Brockie, 161 O.A.C. 324, at para. 31.
a service provider’s beliefs. While the Court’s attempt to define an objective standard based on the core beliefs of the service provider is not helpful, not feasible, and in fact demonstrates how such standards could lead to a multitude of exemptions that would undermine the purpose of the human rights laws, a narrowed solution may exist for those situations in which a service provider is asked to produce material that directly fosters hate towards the service provider (and/or towards a group protected under the antidiscrimination laws).

B. Chamberlain v. Surrey School District No. 36: Can a School Board Refuse to Allow into its Classrooms Books Depicting Same-Sex Parents?

Chamberlain v. Surrey School District No. 36 is one of the most recent and relevant decisions from the Supreme Court of Canada dealing with discrimination on the basis of sexual orientation in the provision of services, addressing both freedom of religion and LGBTQ rights. Chamberlain involved a kindergarten teacher who asked the local school board to approve three books as supplementary learning resources for use in teaching the family life education curriculum. The books—Asha’s Mums, Belinda’s Bouquet, and One Dad, Two Dads, Brown Dad, Blue Dads—depicted families with same-sex parents. The school board responded by passing a resolution refusing to allow these books into the schools. As a result of the board’s resolution, in some schools in the district, certain resources were removed, including library books, posters, and pamphlets.

While Chamberlain did not involve a private actor in the role of service provider, it concerns discrimination in the provision of “services.” Those being discriminated against or otherwise negatively impacted may have included: children, parents,

---

51 Id. at para. 44.
52 Id. at para. 50.
53 Id.
54 Id. at para. 46.
teachers, and the general community. As to the service-provider, while the discriminating body in this case was an elected school board, its decision was based in large part on the views of “parents” in the community who objected to the books, and a concern that having the books at school would create controversy in the children’s homes because of their parents’ views.

The majority rejected the school board’s resolution as unreasonable for having violated the board’s obligations under its governing statute and the relevant regulation, which should have included secularism, nonsectarianism, tolerance, and respect for diversity. By resolving the case on the basis of administrative law principles in this manner, the majority declared it unnecessary to address Charter issues.\(^55\) It was the dissent who raised the difficult questions about the right to dissent and disapprove, and who demonstrated the clash between freedom of religion and equality and their underlying values in this case.\(^56\) A complete analysis and response to the dissent should address these issues through an expanded understanding of secularism and a contextual balancing of the interests at stake, and then conclude, as the majority did, that the school board’s decision to ban the books was impermissible.

1. The Majority: Diversity, the Meaning of Secularism, and Tolerance vs. Freedom of Religion

The majority opinion, delivered by Chief Justice McLachlin, decided against the school board on the basis of administrative law, thus attempting to avoid the difficult issues by avoiding a Charter analysis.\(^57\) However, the dissent tackled these issues head-on, often in very problematic ways, as will be discussed.\(^58\) Perhaps in response to this, the majority relied on the relevant administrative law to take strong stands on diversity, secularism,

\(^{55}\) Id. at para. 76.
\(^{56}\) See id. at paras. 146–52.
\(^{57}\) The majority concluded that the school board’s decision must fail because the board acted outside its statutory mandate by failing to apply both statutory criteria and the board’s own procedures. Id. at para 59.
\(^{58}\) Id. at paras. 75–187 (Gonthier, J., dissenting); see also infra Part II.E.
and tolerance: It made a compelling case for respecting diversity. It considered the meaning of secularism, whether religious views may be included in public debate, and how these views may and may not be used in decision-making. And the majority responded to the dissent’s position on cognitive dissonance—the experience of parents whose children may be forced to learn values contradictory to those at home.

On the issues of secularism and tolerance, the majority held that these were part of the school board’s statutory obligations which the board had failed to meet:

The Board’s first error was to violate the principles of secularism and tolerance in [section] 76 of the School Act. Instead of proceeding on the basis of respect for all types of families, the Superintendent and the Board proceeded on an exclusionary philosophy. They acted on the concern of certain parents about the morality of same-sex relationships, without considering the interest of same-sex parented families and children who belong to them in receiving equal recognition and respect in the school system.

Similarly, the majority noted the requirement that the board recognize diverse communities within the school district and approach the needs of each with “respect and tolerance.” The majority found that the board had not considered families with same-sex parents and had relied instead on the views of a particular group who opposed any depiction of same-sex relationships.

The recognition that there may be different kinds of families

59 Chamberlain, 4 S.C.R. 710 at para. 33 (majority opinion).
60 Id. at para. 49.
61 Id. at para. 33.
62 Id. at para. 59.
63 Id. at paras. 62–66. The majority also found that the school board had failed to follow its regulation, and that the criteria it relied on were the wrong ones. Id. at para. 71.
64 Id. at para. 58.
65 Id. at para. 25.
66 Id. at para 71.
in the school—some who oppose the book, some with same-sex parents, some with an LGBTQ-positive approach—is so obvious it should not need to be stated. But as will be discussed below, this was a point the dissent missed almost entirely.

The majority’s administrative law analysis also involved a discussion about secularism and freedom of religion, and whether the school board was permitted to take into account the views of parents who objected to the books on the basis of religious concerns. The majority concluded that the principle of secularism required by the law did not preclude parents from objecting to the books on the basis of religious considerations. What secularism did require, they found, was that the religious views of one part of the community could not be used to exclude minority voices, that educational decisions and policies must respect the “multiplicity of religious and moral views” held by parents and families in the community, and that the board’s decision must be reasonable in the context of the statutory scheme. In his concurring opinion, Justice LeBel agreed with the majority’s conclusion that parents’ decisions can be based on religious or other views. However, for Justice LeBel, the idea of secularism would rule out “policy based on beliefs that are intolerant of others . . . whether those beliefs are religious, moral or philosophical.” Translating such beliefs into policy is prohibited, he continued, to the extent that these beliefs deny the validity of other points of view:

There is no difficulty in reconciling the School Act’s commitment to secularism with freedom of religion. Freedom of religion is not diminished, but is safeguarded, by the state’s abstention from favouring or promoting any specific religious creed . . . . Disagreement with the practices and beliefs of others, while certainly permissible and perhaps

---

67 There will also be families with LGBTQ children and members of the community with a deeply held, fundamental belief in equality.
68 See discussion infra Part II.E.
69 Chamberlain, 4 S.C.R. 710 at para. 59.
70 Id.
71 Id. at para. 188 (Lebel, J., concurring).
72 Id. at para. 210.
inevitable in a pluralist society, does not justify denying others the opportunity for their views to be represented, or refusing to acknowledge their existence.\textsuperscript{73}

Accordingly, the majority seemed to imply what the concurrence stated explicitly—that the constitutional right to freedom of religion, coupled with the board’s statutory duty to uphold the principle of secularism, required that there be room for all manner of belief and opinion. Given the inevitable conflicts that may arise between two or more belief systems, intolerance would not be tolerated. It would be interesting to consider expanding the meaning of secularism still further, such that in a situation involving a fundamental rights violation, community standards (and prejudices) could not prevail over evidence-based decision making.

Another interesting aspect of the majority opinion was its response to the school board’s reliance on the concept of “cognitive dissonance”\textsuperscript{74} in order to exclude the books. The board had used this term to mean that children should not be exposed to ideas with which their parents disagreed.\textsuperscript{75} The majority found this argument antithetical to the curriculum’s objective of promoting tolerance and an understanding of all types of families. The majority provided a number of examples of differences (based on religion or morals) that may be found in a diverse community—including differences in what classmates were permitted to eat or wear or how they behaved—and stated:

[S]uch dissonance is neither avoidable nor noxious. Children encounter it every day in the public school system as members of a diverse student body . . . . The cognitive dissonance that results from such encounters is simply a part of living in a diverse society. It is also a part of growing up. Through such experiences, children come to realize that not all of their values are shared by others. Exposure to some cognitive dissonance is

\textsuperscript{73} Id. at paras. 211–12.
\textsuperscript{74} Id. at para. 64 (majority opinion).
\textsuperscript{75} Id. at para. 58.
arguably necessary if children are to be taught what tolerance itself involves. When we ask people to be tolerant of others, we do not ask them to abandon their personal convictions. We merely ask them to respect the rights, values and ways of being of those who may not share those convictions. The belief that others are entitled to equal respect depends, not on the belief that their values are right, but on the belief that they have a claim to equal respect regardless of whether they are right. Learning about tolerance is therefore learning that other people’s entitlement to respect from us does not depend on whether their views accord with our own. Children cannot learn this unless they are exposed to views that differ from those they are taught at home.

The emphasis on tolerance is critical in a multicultural society, and it is true that members of this society—children and adults—will be exposed to diverse customs, families, and values. However, the majority’s analysis would have been far better if it had reached this conclusion without minimizing the genuine harm that may have been suffered by some parents who objected to the books and whose children may have been in the classrooms at issue. Once images have been viewed, words read, or ideas shared among the children, they cannot be unviewed and unlearned. Many parents would recoil at the thought of their children being coercively taught values that directly contravene their own—whether such values espouse militarism, sexism, or a particular telling of history. Indeed, as will be discussed, such coercion may amount to a violation of the parents’ dignity. And while the possibility of private school or home-schooling may allow certain parents to opt out of the public system, such an option is beyond the means of many families due to the cost of private school and the financial needs of families with two working parents. Leaving the public school system could also result in a loss of other benefits, such as academic standards,

---

76 Id. at paras. 65–66.
77 Id. at para. 30.
social and community engagement, and sports and art programs. These benefits should, of course, also be available to LGBTQ students, parents, teachers, and equality seekers. Therefore, on balance, openness, inclusion, and diversity would need to prevail.

To conclude, while the majority reached the correct conclusion, it would have been preferable if it had done so with greater sensitivity to the religious freedom of the objecting parents. Such sensitivity would have required the majority to directly engage the real rights infringement faced by these parents and the difficult issues presented by the dissent. The resulting discussion would have more accurately depicted the interests at stake, and would have been richer as a result.

2. The Dissent: Heterosexist Assumptions, Sincere Discriminatory Beliefs, and Dignity

The dissenting opinion, delivered by Justice Gonthier, was indeed sensitive to the harms suffered by objecting parents. As such, the dissent raised the difficult issues in this case concerning freedom of religion, freedom of conscience, equality, and freedom of expression; the tension when a dissenting opinion is discriminatory; and the collision of dignities between those of differing views. In other respects, however, the dissenting opinion should serve as a cautionary tale of how not to adjudicate cases involving discrimination on the basis of sexual orientation. With respect, significant parts of this opinion were based on anachronistic ideas and a heteronormative perspective that quite simply failed to recognize the claims and, in some cases, the existence of LGBTQ children, parents, and educators, as well as other equality seekers in the community.

First, the dissent relied on the distinction (already then discredited in the case law) between the right to equality of all persons, which it said was “consonant with their inherent human dignity,” and the “conduct of persons,” which it implied may not be deserving of equal respect, concern, and consideration. The sexual orientation/sexual behaviour distinction (alternatively

78 Id. at paras. 75–187 (Gonthier, J., dissenting).
79 Id. at para. 77.
referred to as the status/conduct or identity/practice distinction) is one that recurs throughout the belief-based exemption cases. Canadian tribunals and courts have consistently rejected this distinction when it comes to LGBTQ rights, affirming instead that: “Human rights law states that certain practices cannot be separated from identity, such that condemnation of the practice is a condemnation of the person.”

Second, the dissenting opinion attempted to distinguish between the rights of “homosexual persons” to be free from discrimination and “parental rights to make the decisions they deem necessary to ensure the well-being and moral education of their children.” In other words, the dissent’s analysis proceeded as if there were two distinct categories: (1) parents who have a right to educate their children, raise them in their faith, and decide their “best interests”; and (2) “homosexual persons” seeking the inclusion of LGBTQ-positive materials in schools. The values and rights of LGBTQ or LGBTQ-positive parents or students were largely excluded from the analysis and did not seem to play a significant role in the dissenting opinion’s heteronormative worldview.

Third, unfortunately and with respect, things went from bad to worse when the dissent tried to determine whether the books under discussion were about nondiscrimination, or whether they contained LGBTQ-positive messaging. The low point in a less-than-flawless opinion occurred when the dissenting judges, two members of the Supreme Court of Canada, expressed this
distinction with no apparent shame or apology as follows:

The experts basically present two competing views of these Three Books. One view is that they are simply books aimed at the dominant theme of nondiscrimination, with the presence of parents in a same-sex relationship simply being tangential context. The books are therefore about acceptance. The other view is that regardless of the valid and present acceptance theme, a different message is also present: parents in same-sex relationships are being portrayed as “normal” by being portrayed in a positive sense.  

There is good reason to take issue with the above aspects of the dissenting opinion. Nonetheless, the dissent should be credited with bringing to the surface one of the fundamental challenges in this case: in a liberal, pluralistic society, how should the right to equality interact with dissenting beliefs. The dissent expressed this challenge as follows:

It is a feeble notion of pluralism that transforms “tolerance” into “mandated approval or acceptance.” In my view, the inherent dignity of the individual not only survives such moral disapproval, but to insist on the alternative risks treating another person in a manner inconsistent with their human dignity: there is a potential for a collision of dignitie

There are a number of interesting threads in this reasoning. First, there is the assertion that people can and must be able to

86 Chamberlain, 4 S.C.R. 710 at para. 174 (Gonthier, J., dissenting) (emphasis added).
87 Id. at para. 132.
hold a diversity of views to agree with, but also to disapprove of each other’s conduct. For this reason, the dissent asserted, equality cannot simply trump freedom of religion and conscience—there will be a need for balancing. Second, it is not clear if the dissenting judges required that only beliefs based on religion must be protected. For the dissenting judges, it may have been that one person disapproving of another’s conduct constituted a protected belief. They did not explicitly require that such beliefs be religiously grounded. The third thread is the idea that dignity may in some cases be offended when one’s freedom of religion is violated.

While the reasoning in these threads is correct, they fail to paint a full picture, and deserve further attention. With respect to the first thread, protections for dissenting and pluralistic views are critical. It is true that such views may protect offensive beliefs such as those of the objecting parents, but they have also protected minority and marginalized views—including LGBTQ-positive positions. However, what was missing from the dissent’s analysis was the deeply inequitable context in which these two dissenting views were competing. At the time of the Chamberlain challenge, majoritarian privilege rested with the community that banned the books and accepted anti-LGBTQ discrimination. In addition, in the conflict between the two dissenting views, one (the anti-LGBTQ perspective) was aimed at singling out, excluding, and removing from the classroom any resources that mentioned the other (LGBTQ parents).

With regard to the second thread concerning the basis of the beliefs, it is important that deeply held fundamental beliefs of conscience be respected and protected. In the Canadian Constitution, freedom of conscience is protected under the Charter alongside freedom of religion. Such freedom of conscience may protect a person’s right to hold anti-LGBTQ beliefs even if not based on a religious worldview, though the content of such beliefs may lose validity or credibility in the

---

88 Id. at paras. 132–35.
public perception, absent a religious connection. Indeed, discriminatory acts and beliefs that lack the sanction of a religious worldview may better demonstrate the intolerable nature of such discrimination. However, freedom of conscience should also protect a person’s right to the belief that all people are born equal in dignity and rights. This belief can and frequently is grounded in religious roots. It could also be based on secular humanism or other deeply held convictions. The practical result, with respect to the belief-based exemption cases, is that if a person experiences discrimination on the basis of their sexual orientation, it is not just their equality rights that have been violated. There may well also have been a violation of their and others’ deeply held fundamental belief in equality and dignity. In *Chamberlain*, for example, parents with deeply held beliefs in the equality of all people may have felt that the school board’s ban violated their freedom of conscience or religion and caused cognitive dissonance in their homes. The same may be said of Mr. Brillinger when he was denied printing services, or any person for whom equality is a fundamental value, when forced to participate in a discriminatory situation, whether the discrimination is against themselves or others. For this reason, invoking freedom of religion and conscience may not be determinative in resolving such cases.

Finally, on the issue of dignity, the dissent provided an important reminder that a violation of religious freedom could offend one’s dignity.\(^90\) “Dignity” is most commonly associated with equality rights, and discrimination will in many cases result in injury to a person’s dignity. What the dissent establishes is that dignity is not just the purview of equality. If, for example, one is coerced to act against one’s deeply held fundamental beliefs, such as being forced to convert to another religion, or perhaps to violate one’s laws of purity, or as here, to have one’s children taught to believe in a value that contradicts one’s own beliefs, such coercion could amount to a violation of dignity. In the result, resort to the notion of “dignity” could apply to equality or freedom of religion, and as such, this concept may also not be determinative in resolving tensions between these rights in the

\(^{90}\) *Chamberlain*, 4 S.C.R. 710 at para. 134 (Gonthier, J., dissenting).
belief-based exemption cases. A “collision of dignities,” as mentioned in the dissent, may require another form of resolution.91

In conclusion, the court may have done well to rely on an expanded understanding of secularism. Given a situation in which fundamental rights were at stake, the decision about allowing resources into the classroom should perhaps have been based on evidence as to the material’s educational value, ability to engage, age appropriateness, or harmfulness. A proportionality analysis should require that parents’ objections on the basis of their religious freedom would need to be weighed in context against the impact on LGBTQ parents, children, teachers, and others that would be singled out for exclusion, and the fact that all children in the relevant grades would be deprived of exposure to the diversity at issue. While the dissent raised compelling questions about a collision of dignities and the right to disapprove and dissent, on balance in this case, the objecting parents seeking to single out a group for discrimination and exclusion in a public school should not be able to rely on rights such as freedom of religion and equality, whose very purpose is to avoid discrimination and exclusion.

C. Eadie v. Riverbend Bed & Breakfast:92 Can Bed and Breakfast Owners Rely on Their Religious Convictions to Deny a Room to a Gay Couple?

The case of Eadie v. Riverbend Bed & Breakfast involved a couple (the “complainants”) who booked a room at the Riverbend Bed and Breakfast (the “Riverbend”).93 When the Riverbend owners (the “owners”) learned that the complainants were gay, they cancelled the reservation.94 The complainants filed a human rights complaint with the British Columbia Human Rights Tribunal (“Tribunal”), on the basis of section 8 of the British

91 Id. at para. 132.
92 2012 BCHRT 247 (Can.).
93 Id. at para 1.
94 Id. at para. 2.
Columbia Human Rights Code (the “B.C. Code”). The owners denied that their conduct was discriminatory, arguing that the cancellation was justified on the basis of their constitutionally protected right to freedom of religion.

The Tribunal found for the complainants, holding that the owners had refused to provide service because of the complainants’ sexual orientation, that the owners had not proven a *bona fide* and reasonable justification for the discrimination, and could not rely on any of the other exemptions in the B.C. Code. Some of the remedies included: a declaration that the owners’ conduct was discriminatory, an order for the owners to cease and desist from this and similar conduct, and an order that the owners pay each complainant a modest sum for the indignity and humiliation each had suffered.

Although the *Eadie* Tribunal reached the correct result, its analysis was strained and flawed. Based as it was on contemporary human rights (antidiscrimination) law, it did not use the appropriate tools to adequately address a conflict of rights.

The Tribunal’s analysis began well, carefully evaluating evidence to establish the context, including the beliefs of the

---

95 The B.C. Code is legislation aimed at prohibiting discrimination in such areas as employment, housing, and the provision of services, absent a *bona fide* and reasonable justification for the discrimination. *Id.* at para. 95 (quoting British Columbia Human Rights Code, R.S.B.C. 1996, c. 210, § 8(1)).

Section 8(1) of the British Columbia Human Rights Code provides as follows:

A person must not, without a *bona fide* and reasonable justification,

(a) deny to a person or class of persons any accommodation, service or facility customarily available to the public, or

(b) discriminate against a person or class of persons regarding any accommodation, service or facility customarily available to the public . . . .


96 *Eadie*, 2012 BCHRT 247 at para. 2. The Tribunal did not have jurisdiction over constitutional questions *per se*. However the question of equality versus freedom of religion was properly before the Tribunal, as it had jurisdiction to interpret and apply the antidiscrimination provisions of the B.C. Code using normal principles of statutory interpretation, including an interpretation informed by Charter values.

97 *Id.* at para. 173.
owners and the harms caused to the complainants resulting from the discrimination. The Tribunal also properly considered and rejected the sexual identity/behaviour distinction and reached the correct conclusion that the complainants had made out a prima facie case of discrimination. The challenges in this case arose in the next phase of the analysis, when the Tribunal followed the prescribed steps to evaluate whether the owners had a bona fide and reasonable justification (“BFRJ”) for their discriminatory conduct. The BFRJ test proved unhelpful in resolving the issues, and did not allow for nuance or a balancing of the conflicting rights. In its BFRJ analysis, the Tribunal defined the function of the service provider in an absolutist manner that effectively determined the outcome of the complaint, applied a spectrum analysis to evaluate the religiosity of the service provider, and in this as well, came to a conclusion lacking in nuance; and conducted a superficial analysis of two other statutory exemptions that it found to be inapplicable to the complaint. A proportionality analysis of the kind employed in constitutional cases under section 1 of the Charter would have been more direct in raising and assessing the relevant issues, and it would have allowed for the kinds of nuance and balancing necessary in a conflict of rights situation.

1. Establishing Prima Facie Discrimination; Evidence of Religious Beliefs and the Impact of Discrimination; and Sexual Orientation vs. Sexual Behaviour

The decision in Eadie began with the Tribunal taking the time to consider evidence concerning not only the events that occurred, but also the beliefs of the owners and the impact of the events on the complainants. All of these were important for an

---

98 Id. at paras. 1–95.
99 Id. at paras. 96–115.
100 Id. at para. 116.
101 Id. at para. 144.
103 Id. at para. 125.
104 Id. at paras. 1–80.
in-depth contextual analysis. The owners Susan Molnar and Les Molnar were a religious couple, active members of a Church, who hosted religious activities in their home which was also the Riverbend bed and breakfast.\textsuperscript{105} The Riverbend itself had no direct connection to the Church.\textsuperscript{106} The owners’ beliefs about sex and sin, the role of their home religiously, and their belief in God were all recounted by the Tribunal.\textsuperscript{107} These included a belief that all sex outside of a committed, heterosexual marriage is a sin, and that the owners are responsible for what takes place in their home.\textsuperscript{108}

The complainants Shaun Eadie and Brian Thomas were a gay couple.\textsuperscript{109} The Tribunal took the time to describe the emotional and psychological impact of the cancellation on the complainants. This included a description of the bullying, demeaning conduct, and bigotry the complainants had faced since childhood, and how the incident with the Riverbend caused one complainant to “return” to a childhood in which he was shunned and excluded. The Tribunal described how the one complainant had since established his self-confidence, but the refusal shocked and devastated him. The Tribunal also detailed how angry, emotional, and disturbed both complainants felt as a result of the cancellation, and how they had experienced this as an affront to their dignity.\textsuperscript{110}

In this case—as in many belief-based exemption cases—there was an attempt by the owners to distinguish between sexual behaviour and sexual orientation. Mr. Molnar argued that his concern was with sexual conduct in his home and therefore he might have considered an “amicable” arrangement, such as providing two rooms and receiving assurances from the complainants that they would do nothing offensive to the owners’ beliefs (i.e. no sexual conduct).\textsuperscript{111} The Tribunal in \textit{Eadie} rejected

\begin{itemize}
\item\textsuperscript{105} \textit{Id.} at paras. 11–17.
\item\textsuperscript{106} \textit{Id.} at para. 21.
\item\textsuperscript{107} \textit{Id.} at paras. 11–17.
\item\textsuperscript{108} \textit{Id.} at paras. 15–17.
\item\textsuperscript{109} \textit{Id.} at para. 1.
\item\textsuperscript{110} \textit{Id.} at paras. 77–80.
\item\textsuperscript{111} \textit{Id.} at para. 58.
\end{itemize}
this orientation/conduct distinction both as a matter of fact and of law. In support of the latter conclusion, the Tribunal cited, among others, the 2005 B.C. Human Rights Tribunal decision of *Hayes v. Barker*:

[T]he ground of sexual orientation is not exclusively status or identity based, but also protects against discrimination on the basis of behaviours engaged in as a result of a person’s orientation. If it were otherwise, the prohibition on discrimination on the basis of sexual orientation would offer scant protection indeed. Such an interpretation would prohibit a person from being fired for “being” gay, while doing nothing to prohibit a gay man being fired for having sex with his male partner . . .

The Tribunal also rejected the owners’ hypothetical “amicable” arrangement on the grounds that the complainants should not be required to make assurances in order to access a service. One of the complainants expressed his concerns with such an arrangement as follows: “[I]t would have been the same as asking a person of colour to enter from a separate door.”

In conclusion, the Tribunal was satisfied that the complainants had made out the first part of the complaint, having established a prima facie case of discrimination. It was now open to the owners, under the B.C. Code, to attempt to prove a bona fide and reasonable justification for the discrimination. The owners tried

---

112 The Tribunal found that Mr. Molnar had cancelled the complainants’ reservation because they were a gay couple. *Id.* at para. 115.
113 2005 BCHRT 590 (Can.).
115 *Id.* at para. 144.
116 *Id.* at para. 66.
117 *Id.* at para. 115.
118 The elements of a bona fide and reasonable justification are: (1) the respondents adopted a standard, rule, or goal that is rationally connected to the function; (2) they adopted the rule in good faith, in the belief that it was necessary to the fulfillment of the purpose or goal; and (3) the standard was reasonably necessary to accomplish the purpose or goal, in that the
to do so on the basis of their religious beliefs.\textsuperscript{119}

2. *The Bona Fide and Reasonable Justification Defence; Defining the Service; the Spectrum Analysis; Intimacy of the Service; and a Balancing Test*

Both the complainants and the owners in *Eadie* relied on a 2005 B.C. Human Rights Tribunal belief-based exemption case called *Smith v. Knights of Columbus*,\textsuperscript{120} which also focused on the BFRJ analysis. The Knights of Columbus was a Catholic men’s organization that rented out a Church-owned and Church-affiliated banquet hall to Parish church groups, as well as to the general public. While there were no restrictions publicized, the Parish priest had the final word on which activities were permissible in the hall.\textsuperscript{121} In *Knights*, a couple had rented the hall for their wedding reception, but when the organization learned that the rental was for a reception following a same-sex wedding contrary to the Church’s teachings, they cancelled the reservation.\textsuperscript{122}

The *Knights* decision is notable for certain problematic aspects of its analysis. The Tribunal in *Knights* did declare, correctly, that “while everyone is entitled to hold and manifest their own sincerely held religious beliefs and to declare those beliefs, . . . [this] right is not absolute.”\textsuperscript{123} In effect, however, the reasoning of the *Knights* Tribunal provided near-absolute protection for the organization’s freedom of religion in the public domain.

Following the prescribed steps for a BFRJ analysis,\textsuperscript{124} the *Knights* Tribunal had to determine certain concepts to be applied in the test, namely: the rule or standard that led to the prima facie discrimination; and the function of the service at issue. The respondents could not accommodate the individual without incurring undue hardship. *Id.* at paras. 116–17.

\textsuperscript{119} *Id.* at paras. 128–30.
\textsuperscript{120} 2005 BCHRT 544 (Can.).
\textsuperscript{121} *Id.* paras. 1, 6.
\textsuperscript{122} *Id.*
\textsuperscript{123} *Id.* at para. 93.
\textsuperscript{124} *See supra* note 116
Knights Tribunal made these determinations in a manner that incorporated religious belief, effectively deciding the outcome of the analysis through these determinations.\textsuperscript{125} The Knights Tribunal determined that the rule adopted by the Knights organization was: the organization does not rent out the hall for purposes “contrary to its core [Catholic] beliefs.”\textsuperscript{126} The function of the service was determined to be: renting the hall in ways that would not undermine the organization’s relationship with the Catholic Church or conflict with the beliefs of the members of the organization.\textsuperscript{127} Given that both the rule that led to the denial of service and the function of the service were defined in connection with the beliefs of the Church and/or of the organization, it is no wonder that the Knights Tribunal concluded that this rule was rationally connected to this function.\textsuperscript{128}

As to the question of whether the Knights could have accommodated the complainants without undue hardship,\textsuperscript{129} the Tribunal conducted a “spectrum analysis” to evaluate where the case fell on the spectrum between upholding the service providers’ freedom of religion, and the equality rights of the complainants. The Tribunal held that the further the act of prima facie discrimination from the service provider’s core religious beliefs, the less it would be likely to be justified. The Tribunal found, further, that in the case at bar, the hall fell somewhere on the continuum between a parish church, that would not have been required to act against its religious beliefs, and a purely commercial space with no religious affiliation, in which case the complainants would have been entitled to rent the space. Interestingly, the Knights Tribunal referred to Brockie as an example of just such a commercial enterprise.\textsuperscript{130} The Tribunal concluded that:

\begin{quote}
a person, with a sincerely held religious belief cannot be compelled to act in a manner that
\end{quote}

\begin{footnotes}
\textsuperscript{125} Knights of Columbus, 2005 BCHRT 544 at paras. 108–09.
\textsuperscript{126} Id. at para. 108 (emphasis added).
\textsuperscript{127} Id. at para. 88.
\textsuperscript{128} Id. at para. 89.
\textsuperscript{129} Id. at paras 91–92.
\textsuperscript{130} Id. at paras. 106–10.
\end{footnotes}
conflict[s] with that belief, even if that act is in the public domain . . . . [T]he Knights are entitled to this constitutional protection and therefore cannot be compelled to act in a manner that is contrary to their core religious beliefs.  

Even though the respondents were not being asked to participate in the solemnization of a same-sex marriage, the Tribunal decided, renting the hall for its celebration would have required the organization to indirectly condone the celebration of a same-sex marriage, contrary to the members’ core religious beliefs.  

This absolutist position, whereby the organization’s core beliefs should not be violated, may be more extreme than the problematic “objective” standard adopted by the Court in Brockie.  

Despite this, the Knights Tribunal found for the complainants on the narrow ground that the Knights had not accommodated the complainants in a manner that did not violate the respondents’ beliefs. Such accommodation could have included, according to the Tribunal: meeting with the complainants, explaining the situation to them, formally apologizing, reimbursing them immediately, and possibly offering them assistance to find another venue.  

The BFRJ analysis in Knights raises a number of difficulties, demonstrating certain analytical positions that should be avoided in belief-based exemption cases. First, in Knights, the religious nature of the organization appears to have trumped other factors, such as the extremely tenuous connection between the service provider and the service. In contrast, for example, to Mr. Brockie the printer, who presumably would have had to be personally involved to some extent in producing the requested material, the organization in Knights did not appear to have had a connection to the event other than through the rather impersonal act of renting out the hall; and even that was for the wedding reception, not the ceremony itself. Indeed, the identity of the

---

131 Id. at para. 113 (emphasis added).
132 Id.
133 See supra notes 32–48 and accompanying text.
organization as service provider appears to have been determinative for the Tribunal, as it stated in its spectrum analysis that it would not require the Catholic Church to rent its Parish Church space for the reception against its core religious beliefs, but that it would have no difficulty compelling a commercial enterprise, such as that in *Brockie*, to rent its hall.\footnote{Id. at para. 109.} It is not clear how the Tribunal reconciled this reasoning with its conclusion that it would not force any person with a sincerely held religious belief to act against that belief. After all, Mr. Brockie was the owner of just such a business, and a commercial marriage hall might be owned by someone with views similar to those of the Knights. An additional difficulty arises in this regard. The premise of the spectrum analysis as applied in this case seems to create an exemption based on the religious identity of the service provider. However, the B.C. Code already has an exemption for religious organizations, an exemption that did not apply in this case. At the least, the Tribunal should have attempted to reconcile its spectrum-analysis exemption with the existing statutory exemption.

Second, the Tribunal adopted an absolutist position—that it would not have forced the organization to do *anything* that violated its members’ core beliefs, even in the public domain—which is wrong both in principle and in law. The logic in *Knights* could lead to even more severe and absurd results, as it would seem to justify any discrimination as long as there was a sincerely held core religious belief. What if, for example, the Knights organization, on the basis of a sincerely held religious belief, refused to work with LGBTQ couples, would not refer them to another venue, and perhaps even felt it immoral to be near with them (and therefore put up signs in the window indicating that LGBTQ people would not be served)? On the logic of the *Knights* Tribunal, this discriminatory conduct could be justified and may be protected. The Tribunal’s absolutist reasoning undermines the very basis of the human rights antidiscrimination laws, which were designed specifically to compel people to act against their convictions if those convictions would lead to discriminatory results. To be fair, it is hard to know whether the *Knights*
Tribunal intended to take such an extreme position, given its conclusion that the organization had in fact discriminated.

The Eadie Tribunal, in the first part of its BFRJ analysis, used similar reasoning, but reached a different conclusion as to the function of the service being provided by the Riverbend. This Tribunal accepted that the Riverbend owners sincerely believed allowing a same-sex couple to share a bed in their home would harm the owners’ relationship to their Lord. Nonetheless, the Tribunal found that the owners had not established a bona fide and reasonable justification for their discriminatory conduct. Here, as in Knights, the conclusion hinged on the definition of the function of the service. In Eadie, the Tribunal defined the function of the bed and breakfast without reference to the owners’ religious views, finding that the Riverbend’s function was to provide temporary accommodation to the general public. Therefore, the Tribunal concluded, the rule excluding couples who were not a married man and woman was not rationally connected to this purpose of providing temporary accommodation.

The two cases together provide a clear illustration of how the definition of the service’s function effectively determines the rest of the analysis, but is not helpful in resolving the belief-based exemption cases. If the function is defined, as in Knights, in relation to the service providers’ religious beliefs, they will be granted near-absolute protection for these beliefs, subject to the duty to accommodate. If the function is defined without reference to the service providers’ religious beliefs, as in Eadie, they will be left without any protection, despite the fact that they appear to have believed just as fervently as the Knights did that their business should be run without harming their relationship with their Church or with their Lord. Such absolutist conclusions in both cases do not leave room for nuance or balancing.

For the sake of caution, the Eadie Tribunal did not stop at the

---

136 Eadie v. Riverbend Bed & Breakfast, 2012 BCHRT 247, para. 139. (Can.).
137 Id. at para. 145.
138 Id. at para. 141.
139 Id. at para. 144.
first part of the BFRJ analysis. In its decision on the third part of the analysis assessing the duty to accommodate, the Tribunal addressed a number of issues. It considered the argument that the Riverbend was in the owners’ home.\textsuperscript{140} The Tribunal held that there was no statutory exception available for “services” in a situation of shared sleeping, bathroom, or cooking facilities, to parallel the statutory exception for tenancy in such circumstances.\textsuperscript{141} The Tribunal concluded that those parts of the Riverbend occupied by guests were properly characterized as business premises.\textsuperscript{142} What underlies this formalistic discussion may be the idea that intimacy between service provider and the recipients of the service could in very exceptional cases justify a belief-based exemption, or minimize the duty to accommodate. No such level of intimacy was reached by renting a room in a bed and breakfast.

Next, the \textit{Eadie} Tribunal considered the religiosity of the Riverbend. It conducted a spectrum analysis, concluding that the Riverbend was “\textit{more toward} the commercial end of the spectrum . . . . While the business was operated by individuals with sincere religious beliefs respecting same-sex couples, and out of a portion of their personal residence, it was still a commercial activity.”\textsuperscript{143}

The Tribunal also considered the Riverbend’s clientele, concluding that they were not restricted to the Christian community.\textsuperscript{144} The Tribunal refrained from deciding whether this would have made a difference to the decision.

The focus on religiosity seems to imply that religious institutions who wish to discriminate in providing services to the public could enjoy a lower duty to accommodate. As discussed above, an exemption or diminished duty to accommodate based on the religious identity of the service provider is rife with issues, as it lacks nuance and does not leave room for consideration of factors such as the impact of the harm. A test hinging on the

\textsuperscript{140} \textit{Id.} at paras. 151–53.
\textsuperscript{141} \textit{Id.} at para. 160.
\textsuperscript{142} \textit{Id.} at para. 161.
\textsuperscript{143} \textit{Id.} at para. 165 (emphasis added).
\textsuperscript{144} \textit{Id.} at para. 166.
religiosity of the service provider may also lend itself to absolutist conclusions, as appears to have been the case in *Eadie*.

The Tribunal also held that as it was the owners’ decision to run a business in their home, in this they were not compelled by the state to act contrary to their religious beliefs.\(^\text{145}\) The Tribunal acknowledged that being religious practitioners in the public domain may carry a cost (in money, tradition, or inconvenience), but stated that such costs are less serious than a limit that effectively deprives the adherent of any meaningful choice with respect to their practice.\(^\text{146}\) Having decided that the owners were not deprived of a meaningful choice with respect to the exercise of their religion, the Tribunal concluded:

> [T]heir choice or mode of business operation may be limited by their religious practice. Having entered into the commercial sphere, the Molnars (owners), like other business people, were required to comply with the laws of the Province, including the *Code*, which is quasi-constitutional legislation that prohibits discrimination on the basis of sexual orientation.\(^\text{147}\)

Thus the *Eadie* Tribunal appeared to be saying that business people who choose to enter the commercial sphere may not discriminate. This too is near-absolutist reasoning, as the Tribunal appeared to be ruling out religious freedom exemptions for any person providing services in the public domain who is not acting in furtherance of a religious goal. This included the owners, despite their strong personal religious views. It could also include a female massage therapist who, for religious reasons, does not accept adult male customers.

The Tribunal’s reasoning on this point is not persuasive. Harms to religious freedom are constitutional infringements, which must be acknowledged as such and balanced against the relevant countervailing interests, as discussed below. It is also unpersuasive to assert that “simply” asking people to change their mode of business does not constitute an interference with their

\(^{145}\) *Id.* at para. 165.

\(^{146}\) *Id.* at para. 168

\(^{147}\) *Id.* at para. 169.
freedom of religion. Asking people to change or move their business because they are not in compliance with the law is coercive, whether the owner chooses to shut down their business (as the owners did subsequently in Eadie) or to comply with the law against their own convictions. Such coercion may be justified, as it was in Eadie, but the impact on the owners can nonetheless still be acknowledged.

What the analyses in Knights and Eadie demonstrate are the challenges created when trying to balance equality with freedom of religion within a bona fide and reasonable justification test as applied in these cases. The Tribunals engaged in awkward discussions and reached unlikely conclusions, inserting religious beliefs into the function of a wedding hall and determining that a bed and breakfast was “more” on the commercial end of the spectrum. A straightforward balancing exercise, like that under section 1 of the Charter, would be preferable,148 similar to that conducted by the courts in Marriage Commissioners and in Brockie.149 Though the court reached the wrong conclusion in the latter case, these courts were able to ask themselves the correct questions and consider nuanced solutions.

On the facts in Eadie, a balancing exercise could consider the equality rights of the complainants, assess the actual harms they suffered, as well as the social context and greater harms that may occur if the owners could single out LGBTQ people for discrimination. It could weigh this against the religious freedom of the owners, the actual harms they would suffer if prohibited from discriminating, and the greater social context. In a belief-based exemption case such as this, one party or the other may end up feeling forced out of the public domain. Indeed, some might consider this to have been the fate of the Riverbend owners, who did in fact leave the public domain and shut down their bed and

148 This appears to be the method proposed by the Ontario Human Rights Commission for those situations where reconciliation is not possible. See ONTARIO HUMAN RIGHTS COMM’N, POLICY ON COMPETING HUMAN RIGHTS (2012), available at http://www.ohrc.on.ca/sites/default/files/policy%20on%20competing%20human%20rights_accessible_2.pdf.

breakfast following the Tribunal’s decision prohibiting them from discriminating. If anyone has to leave the public domain, in most instances, it likely should be those who want to be kept apart, to exclude or discriminate.

Moreover, in the *Eadie* and *Knights* cases, as well as the others discussed in this Article, it is likely no coincidence that those who created the exclusionary rule were those with greater social capital, while those who would have been excluded belonged to a group that has experienced, and continues to experience, discrimination and marginalization. This too may be relevant context in determining the outcome of these cases. Had the *Eadie* Tribunal engaged in such a balancing exercise, it could have concluded, as this Article would, that the individual and social benefits of preventing discrimination against a couple on the basis of their sexual orientation in the specific, social, and historical context of this case outweigh the deleterious impact on the owners.

This conclusion, that the owners in *Eadie* should not be permitted to discriminate, is simple albeit coercive. However, there is nothing earth-shattering in the proposition that law coerces. The human rights laws work ex ante by prohibiting discrimination, and ex post facto by enforcing coercive measures where discrimination has taken place. The coercive nature of these laws has not changed since they were first established with the purpose of forcing individuals and businesses to serve, employ, and house people of different religions and races, against the sometimes deeply held convictions of those who would have otherwise discriminated. The importance of the antidiscrimination measures justifies the creation of these coercive human rights laws and continues to justify their implementation. This is particularly true for anti-LGBTQ discrimination in the current Canadian context. That said, it is still necessary for a court or Tribunal to examine each belief-based exemption case on its facts.
D. Reference re Constitutional Act, 1978 (Saskatchewan): Should Civil Marriage Commissioners Be Exempt From Solemnizing Marriages Contrary to Their Religious Beliefs?

In 2004 and 2005, same-sex marriage was legally recognized across Canada following a series of constitutional challenges and ultimately, new federal legislation. This tremendous change raised a new legal and constitutional question with respect to belief-based exemptions. In the province of Saskatchewan, eight marriage commissioners resigned after being informed that they were required to perform same-sex marriage ceremonies, while others filed human rights complaints claiming that their freedom of religion and their right to carry on an occupation without religious discrimination had been violated. Other human rights complaints and litigation followed.

Eventually, the Lieutenant Governor in Council asked the Saskatchewan Court of Appeal to provide an advisory opinion, known as a reference, on the constitutionality of two alternative possible amendments to the province’s Marriage Act. The court’s opinion, although not legally binding, was provided in the form of a judicial decision: this is the Marriage Commissioners decision. The two possible amendments under consideration in this case, if passed into law, would have created a belief-based exemption for marriage commissioners by allowing them to decline to solemnize a marriage if it would be contrary to their

---

150 In re Marriage Comm’rs, 2011 SCKA 3.
152 Civil Marriage Act, S.C. 2005, c. 33 (Can.).
154 2011 SKCA 3 (Can.).
religious beliefs. The first option would have made the exemption available only to those marriage commissioners appointed on or before November 5, 2004—the date on which the courts in Saskatchewan recognized same-sex marriage. The second option would have made the exemption available to all marriage commissioners in the province, regardless of the date of their appointment. The court’s analysis addressed both possible exemptions together.

The court held that these exemptions would, if enacted, be unconstitutional and invalid: they would violate the equality rights of LGBTQ individuals guaranteed in section 15 of the Charter; this infringement of section 15 would be unreasonable and unjustifiable under the Charter’s section 1 reasonable limits test, as the proposed exemptions would not be minimally impairing, and their harms would far outweigh their benefits.

This case, similar to Chamberlain, did not involve a private actor in the role of service-provider. Indeed, the majority’s decision was based in large part on the fact that marriage commissioners act as government officials. Nonetheless, there is a great deal in the Marriage Commissioners decision that is helpful for exploring the issue of belief-based exemptions—for the most part because of its contribution to this debate, in particular its section 1 analysis, but also, with respect, because of its flaws. The issues raised include the following: whether the purpose of the proposed exemptions was to protect religious

---

155 The operative part of the first option reads as follows:

28.1(1) Notwithstanding The Saskatchewan Human Rights Code, a marriage commissioner who was appointed on or before November 5, 2004 is not required to solemnize a marriage if:

(a) to do so would be contrary to the marriage commissioner’s religious beliefs; and
(b) the marriage commissioner has filed the notice mentioned in subsection (2) within the period mentioned in that subsection.

Id. at para. 17.

156 The second option reads: “28.1 Notwithstanding The Saskatchewan Human Rights Code, a marriage commissioner is not required to solemnize a marriage if to do so would be contrary to the marriage commissioner’s religious beliefs.” Id.
freedom or to facilitate discrimination on the basis of sexual orientation,\textsuperscript{157} who may define religious beliefs, and the oft-cited yet problematic distinction between religious beliefs and acts; the deleterious effects of the exemptions, and their impact on individuals and society, particularly if institutionalized and legitimized through official policy; and again, the slippery slope of exemptions that could be justified if the proposed exemptions were permitted.\textsuperscript{158} This Article shares many of the court’s conclusions, but differs on the question of religious freedom and the right of individuals to define their religious views and priorities for themselves.

1. Factual Background, Charter Analysis, Equality, the LGBTQ-Specific Context, and the Purpose of the Exemptions

The court’s factual findings established the background to this case, as follows: marriage commissioners in Saskatchewan are appointed by the Minister and provide the only route to marriage for individuals who want a nonreligious ceremony.\textsuperscript{159} Indeed, the Marriage Act specifies the requirements and the wording for a civil ceremony—and these are strictly nonreligious. Individuals wanting a civil marriage may receive contact information through the provincial government, following which they can contact a commissioner directly.\textsuperscript{160} According to the court, this route would be the only one available to many gay and lesbian couples who want to get married.\textsuperscript{161}

The two proposed amendments to the province’s Marriage Act at issue\textsuperscript{162} would have exempted all or some of these civil marriage commissioners from the duty to solemnize a marriage, if doing so would be contrary to their religious beliefs.\textsuperscript{163} The court was asked to provide its opinion on the constitutional

\textsuperscript{157} Id. at paras. 74, 78–79, 115.
\textsuperscript{158} Id. at para. 90.
\textsuperscript{159} Id. at para. 9.
\textsuperscript{160} Id. at para. 8.
\textsuperscript{161} Id. at para. 9.
\textsuperscript{162} The Marriage Act, 1995, S.S. 1995, c. M-4.1 (Can.); see also supra notes 155–56.
\textsuperscript{163} See supra notes 151–52 and accompanying text.
validity of these proposed exemptions. Applying the established test, the court began its analysis by considering whether the proposed exemptions infringed a Charter right or freedom. It concluded that the exemptions did in fact violate the right to equality under section 15 of the Charter. It then fell to the court to determine whether the exemptions constituted a reasonable limit on this right such that they could be justified under section 1 of the Charter.  

Applying the section 1 analysis to the exemptions, the court first considered the purpose of the exemptions. The majority (though not the concurrence) held that the purpose of the exemptions—to protect freedom of religion—was pressing and substantial, thus satisfying this branch of the test. Second, the majority held that the proposed amendments were rationally connected to this purpose, in that the exemptions would indeed protect marriage commissioners’ religious freedom. However, on the third—minimal impairment branch of the section 1 analysis—the majority found that the exemptions were more restrictive than necessary to achieve their objective. Given the possibility of an alternative method for matching couples with marriage commissioners—a method that would have harmed equality rights less than the proposed exemptions—the majority concluded that the exemptions were not minimally impairing, not a reasonable limit on the right to equality, and as such, they were unconstitutional. The court decided to provide its opinion as well on the fourth, and final, branch of the section 1 analysis and concluded that the deleterious effects of the exemptions far outweighed their salutary effects. For this reason as well, the exemptions did not constitute a reasonable limit and were unconstitutional.

Aspects of the court’s Charter analysis raise interesting, insightful, helpful, at times controversial, and even troubling elements, all of which are illuminating in the context of a larger

---

164 For the reasonable limits test, see supra notes 7–9 and accompanying text.
165 Marriage Comm’rs, 2011 SKCA 3 at para. 82.
166 Id.
167 Id. at para. 88.
168 Id. at para. 101.
discussion about belief-based exemptions. It is these elements that will be drawn out for a more detailed discussion below.

In the initial stage of its Charter analysis, the majority considered whether the proposed exemptions infringed a Charter right. It concluded that the exemptions would curtail the Charter’s right to equality under section 15. Its conclusion was based on the following findings: an exemption could lead to any number of commissioners refusing to perform same-sex marriages; the impact of such refusals on an LGBTQ couple could be very significant and genuinely offensive; and even if a few commissioners opted out of performing same-sex marriages, LGBTQ couples looking for a commissioner might face some inconvenience, could have to deal with numerous refusals, and they may encounter real difficulty in small or remote locations. Also, in light of historical discrimination and mistreatment of LGBTQ individuals, allowing marriage commissioners to refuse a same-sex couple service “would clearly be a retrograde step—a step that would perpetuate disadvantage and involve stereotypes about the worthiness of same-sex unions.”

In the next stage, it was necessary to conduct a Charter section 1 analysis to determine whether the infringement of the right to equality could be justified as a reasonable limit on this right. If so, the exemptions would be constitutional. To begin the section 1 analysis, the court needed to establish the purpose of the proposed exemptions. These exemptions were drafted broadly and would have allowed a marriage commissioner to refuse to perform any kind of marriage, such as an inter-faith union. And indeed, the majority found that the purpose of the exemptions was to protect the religious freedom of marriage commissioners by relieving them of their duty to perform any marriage contrary to their religious beliefs. That said, Justice Richards, writing for the majority, focused his analysis specifically on one kind of situation—a refusal to solemnize the marriage of a same-sex couple—as there was no evidence of any

169 In re Marriage Comm’rs Appointed under the Marriage Act, 2011 SKCA 3, para 45 (Can.).
170 Id. at para. 24.
171 Id. at para. 76.
other kind of refusal and because same-sex marriage was in fact the issue underlying the debate.\textsuperscript{172}

The concurrence delivered by Justice Smith went further, considering, clarifying, and redefining the purpose of the proposed exemptions.\textsuperscript{173} According to the concurrence, the \textit{objective} of the exemptions was not simply to accommodate the religious freedom of marriage commissioners, but to permit marriage commissioners to refuse to perform \textit{same-sex marriage ceremonies} when doing so would conflict with their religious beliefs.\textsuperscript{174} The facts underlying this conclusion were not difficult to demonstrate, particularly given that one of the two proposed exemptions was drafted specifically as a grandfathering option for those marriage commissioners appointed on or before the date that same-sex marriage was recognized in Saskatchewan. The concurring opinion concluded that this objective was not “pressing and substantial,” as required by section 1 of the \textit{Charter}, or at least it was doubtful whether this objective met the required threshold.\textsuperscript{175}

What is particularly useful here is the insistence (by the concurrence, and perhaps in its wake, the majority), that the analysis be situated in its specific context. In this case, the issue was not simply one of religious freedom versus equality. It was squarely about discrimination on the basis of sexual orientation. And those who would be most harmed by the exemptions’ discriminatory effect were LGBTQ individuals. Providing detailed context and considering the identity, circumstances, and history of those who would be impacted by the denial of service will be critical for various stages of the analysis in any belief-based exemption case.

\textit{2. Minimal Impairment and the Single Entry Point System}

As to whether such a limit on equality rights was justifiable, the court found that the exemptions failed the minimal

\textsuperscript{172} \textit{Id.} at paras. 104–09.
\textsuperscript{173} \textit{Id.} at paras. 115–30.
\textsuperscript{174} \textit{Id.} at para. 154.
\textsuperscript{175} \textit{Id.} at para. 152.
impairment test,\textsuperscript{176} as there was at least one alternative system that could harm individuals’ equality rights less than the proposed exemptions.\textsuperscript{177} This alternative, known as the “single entry point” system,\textsuperscript{178} would allow couples seeking a marriage commissioner to apply through a central office, at which point they would provide information about themselves (i.e., their genders). After this, the Director of the Marriage Unit would provide them with a list of available commissioners. The list provided to the couple would exclude those commissioners not prepared to officiate—all of which could have been established privately and “behind the scenes.”\textsuperscript{179} Such a system, the majority held, would be less harmful than the proposed exemptions, as it would accommodate marriage commissioners’ beliefs; the accommodation would not be readily apparent to an LGBTQ couple; and the couple would not risk being refused service because of their sexual orientation. The court, however, was careful to explain that the discussion about the single entry point system did not prove the system’s constitutionality. It only served to prove the lack of constitutionality of the proposed exemptions which would be even more harmful to individuals’ equality rights.\textsuperscript{180}

The single entry point system is indeed flawed. While it could mercifully shield LGBTQ couples arranging their weddings from the indignity and pain of a refusal, it would still include a request for information about a person’s gender, which should be irrelevant once same-sex marriage is recognized, and which is always problematic for transgender people. This system would also do nothing to shield the people “behind the scenes” from having to work with and implement this policy. Marriage commissioners and their associates, the Director of the Marriage Unit, and clerical and technical staff may themselves be LGBTQ, have a loved one who is, and/or have a deeply held belief in the equality, dignity, and worth of all people. Being required to work

\textsuperscript{176} The minimal impairment test requires that for a limit to be reasonable, it must “limit rights no more than necessary.” \textit{Id.} at para. 83.

\textsuperscript{177} Indeed, there was a suggestion that a system of this kind operated in Toronto. \textit{Id.} at para. 87.

\textsuperscript{178} \textit{Id.} at para. 85.

\textsuperscript{179} \textit{Id.}

\textsuperscript{180} \textit{Id.} at para. 89.
with a single entry point system would require such individuals to fill out forms, enter data, manage lists, and so forth in a context that would facilitate the singling out of LGBTQ people for exclusion. As will be discussed below, the very creation of an official ex ante policy allowing people to opt out of performing same-sex marriages sends a problematic message legitimizing this refusal.181

3. Salutary Effects and Defining Religious Beliefs

On the final branch of the section 1 analysis, the court engaged in a balancing exercise to weigh the deleterious effects of the exemptions against their salutary ones.182 The exemptions’ benefits, the majority stated, were intended to protect marriage commissioners from having to do certain actions contrary to their religious beliefs. While these beliefs may be significant for some commissioners, the majority held that the benefits of the exemptions were less significant than they appear because:

the freedom of religion interests [that the exemptions] accommodate do not lie at the heart of [section] 2(a) of the Charter. [The exemptions] are concerned only with the ability of marriage commissioners to act on their beliefs in the world at large. They do not in any way concern the freedom of commissioners to hold the religious beliefs they choose or to worship as they wish.183

While agreeing with the majority’s conclusion that the exemptions’ harms far outweigh their benefits, this Article does not accept the distinction, invoked as well by the concurrence, between acting on beliefs and holding them as an appropriate method of determining the significance of a restriction. More generally, and with respect, the proposition that in defining a person’s religious freedom, courts can rely on the distinction

181 See infra Part II.D.4. This point was made by the concurrence about the exemptions themselves. Marriage Comm’rs, 2011 SKCA 3 at para. 107.
182 Id. at para. 90.
183 Id. at para. 93.
between “belief and conduct”—a distinction frequently cited in freedom of religion and equality cases—may be factually inaccurate and is philosophically unsound. This distinction privileges one subjective understanding of religion over others. While it is true that some religions are based primarily on faith and worship, there is a diversity of religious and spiritual systems, expressions, and practices that should not be overlooked. For some adherents and religions, faith or attendance at a house of worship may be less religiously and spiritually significant than, for example: acts of charity; ethical behaviour (whether others agree or not with aspects of these ethical systems); or ways of being in the world (including dietary regimes, modes of dress, and laws around sexual behaviour). In other words, for many adherents, their core religious freedom may be dependent on the freedom to conduct themselves according to a system of ethics and practices prescribed by their religion. This is not to say that one can never limit religious practices—such limits can and should take place in various situations, including the case under discussion. However, a meaningful analysis should rely on the actual religious worldview of the individual in question, not on the court’s subjective beliefs about what constitutes religion.

The concurring opinion also dealt with the questions of how to define—and who should define—religious beliefs deserving of section 2(a) freedom. For instance, given the nonreligious nature of civil marriage and the importance of the civil scheme, the concurrence asked “in precisely what respect being compelled to perform a same-sex marriage can offend the religious freedom of a marriage commissioner.” Justice Smith’s intention, she explained, was not to question the sincerity of the belief, but rather to examine the significance of the societal harm the exemptions are intended to remedy and to what extent freedom of religion is offended by requiring marriage commissioners to perform same-sex marriages.

---

185 Marriage Comm’rs, 2011 SKCA 3 at para. 129.
186 Id. at para. 130.
This is, with respect, a strange question. If a person were asked to participate in a ritual that was offensive to their sincere moral or ethical core, surely this would be a violation of section 2(a). Like equality, freedom of religion is a value. And this value would be harmed if people were required to act against their beliefs (including beliefs as to how they may and may not conduct themselves). It will nonetheless be open to a court to determine whether another conflicting value, such as equality, will receive greater protection in a particular context.

The concurrence also turned to the evidence and analysed the beliefs expressed by various affiants, including the following statement: “[M]onogamous, non-polygamous, heterosexual marriage is . . . a uniquely Christian doctrine. A Christian must always recognize marriage as such, and understand that any attempt on the part of society to define it in any other way is disobedience to the Covenant and incurs the righteous judgment of God.”\(^{187}\) The concurrence attempted to demonstrate that refusals to perform same-sex marriage are not reasonable, plausible, or compliant with the law. Justice Smith also held that performing a same-sex marriage does not necessarily imply approval of the union. And she asserted that “[t]he performance of a civil marriage by a marriage commissioner under the Act is not a religious rite or practice. Nor does the requirement to do so limit or restrict religious belief.”\(^{188}\)

Justice Smith’s analysis with respect to defining religious beliefs and freedom was erroneous. First, the concurrence was asking itself the wrong questions when it tried to assess the reasonableness, plausibility, or legal coherence of the refusals. Religious beliefs do not become less religiously true just because they may be unreasonable or implausible. Second, Justice Smith held that performing a marriage does not imply approval of the marriage. While this assertion may be true for the respected Justice, it directly contradicts the evidence of the affiants who expressed concerns about condoning or approving of these unions by virtue of solemnizing the marriages, as cited by Justice Smith herself. Finally, the concurrence seemed to suggest that only

\(^{187}\) *Id.* at para. 135 (emphasis added).

\(^{188}\) *Id.* at para. 147.
religious rites can be relevant to an infringement of a person’s religious belief. This suggestion is unfounded. If a person is forced to violate a religious prohibition (such as eating pork, having a blood transfusion,\textsuperscript{189} or removing a religiously mandated ritual object),\textsuperscript{190} this may well be a violation of their religious freedom even where no religious rite is involved.

The question of how to define “core” religiosity or significant harms to religious freedom is one that recurs in the equality and freedom of religion cases. With respect, the concurrence’s analysis in this regard seemed to demonstrate a conceptual difficulty in grasping the nature of religious or conscientious belief and practice. This was particularly evident from the fact that Justice Smith found only a weak to nonexistent interference with religious freedom, despite having cited to an affidavit expressing an individual’s fear about incurring “the righteous judgment of God.”\textsuperscript{191}

Religious beliefs (with respect to worship, conduct, and practice) are subjective and personal. Their range and expression may be diverse, their content may be irrational and idiosyncratic, and they may contain values that are anathema to others. Courts and tribunals adjudicating belief-based exemption cases should engage in a serious contemplation of freedom of religion that allows for the possibility that such beliefs are nonetheless real for the adherent.

It is not the role of the judge to define others’ beliefs based on their own logic and understanding. The role of the court, when freedom of religion is claimed, is to test the evidence with respect to sincerity of belief and the scope of the purported harm to the individual. It is also the courts’ role to limit these sincerely held core beliefs if this is justified under section 1 of the \textit{Charter}. 


\textsuperscript{190} See, e.g., Multani v. Comm’n scolaire Marguerite-Bourgeoys, [2006] 1 S.C.R. 256 (Can.).

\textsuperscript{191} Marriage Comm’rs, 2011 SKCA 3 at para. 135.

Having established a conflict between religious freedom and equality, the critical phase of the analysis in belief-based exemption cases may well take place in the test that weighs the benefits of a measure against its harms. The majority in Marriage Commissioners found that the first deleterious effect of the proposed exemptions was the fact that they would undermine the struggle for equality generally, and perpetuate a brand of discrimination which our national community has only recently begun to successfully overcome. It would be a significant step backward if, having won the difficult fight for the right to same-sex civil marriages, gay and lesbian couples could be shunned by the very people charged by the Province with solemnizing such unions.192

The exemptions’ second deleterious effect, according to the majority, was in their harmful impact on individuals. To demonstrate this, the majority cited the testimony—from a different case in which a marriage commissioner denied service to an LGBTQ couple—of an individual who was denied service, giving voice to his experience and reaction. The man, M.J., had testified as follows:

It was actually pretty devastating . . . . So when this happened I was quite devastated. I rehashed this I don’t know how much when I couldn’t sleep because I actually wound up sleeping very little. I was just crushed about it. I couldn’t believe that as a human being I wasn’t going to be treated as a real person.193

The majority found that the negative and harmful effects of this kind of denial would affect not just those LGBTQ individuals

192 Id. at para. 94.
denied services, but the LGBTQ community, their friends and family, and the public as a whole—as many members of the public would be hurt and offended by the idea that a governmental official would deny services to LGBTQ couples.\textsuperscript{194}

The third and “in some ways most important” deleterious effect of the exemptions, in the majority’s view, was that they would undermine the principle that the government serves everyone equally without discrimination.\textsuperscript{195} Marriage commissioners do their jobs as agents of the province. Individual public office-holders cannot expect to change the way the office interacts with the public to conform to their own beliefs, as this would be inconsistent with the principle of the rule of law.\textsuperscript{196}

Concurring Justice Smith discussed additional deleterious effects that the proposed exemptions would cause. She provided a detailed overview of the marriage solemnization regime. This included the fact that, according to her, a significant number of religious organizations disapprove of same-sex marriage, and thus, civil marriage may be the only route to marriage available for same-sex couples.\textsuperscript{197} For example, there were 138 religious bodies whose clergy may marry according to their rites and usages in the province, in contrast to the single prescribed form (including a set script) for a nonreligious marriage.\textsuperscript{198} And the number of clergy ever registered with the marriage unit (5,713) was contrasted with the number of marriage commissioners (578).\textsuperscript{199} These facts told a compelling story about the importance of maintaining an open, accessible, and impartial civil marriage option.

The concurrence also described the impact of discrimination on the basis of sexual orientation and took the time to set out the broader context responsible for the “extreme vulnerability” of LGBTQ people to hatred and discrimination.\textsuperscript{200} She cited a well-known and painful passage that described the historic and current

\begin{flushright}
\textsuperscript{194}\textit{Id.} at para. 45. \\
\textsuperscript{195}\textit{Id.} at para. 97. \\
\textsuperscript{196}\textit{Id.} at para. 98. \\
\textsuperscript{197}\textit{Id.} at para 106. \\
\textsuperscript{198}\textit{Id.} at para. 119. \\
\textsuperscript{199}\textit{Id.} at paras. 119–28. \\
\textsuperscript{200}\textit{Id.} at para. 107. 
\end{flushright}
disadvantages faced by this group, including: public harassment, verbal abuse, violence, exclusion from public life, a need to conceal identities and orientation, rejection by families, and, as a result of these, higher rates of suicide and attempted suicide.\textsuperscript{201}

Next, the concurrence explained, the harm goes further than the individual exemptions: “[E]ven if the risk of actual refusal were minimal, knowing that legislation would legitimize such discrimination is in itself an affront to the dignity and worth of homosexual individuals.”\textsuperscript{202} Thus, she concluded, what is at stake is not just the right of same-sex couples to marry, but the right of this vulnerable group to be free from discrimination in the provision of a public service, which is provided without discrimination to every other person in society.\textsuperscript{203} Justice Smith reinforced this point by demonstrating that there was no other legislative provision in the province explicitly operating in conflict with the provincial Human Rights Code.\textsuperscript{204} Her insight provides a coherent and persuasive message, demonstrating why an official policy whose effect is to permit exclusion against a particular group is so problematic:

Astonishingly, this clause [the exemptions] would grant to a public official, charged with the delivery of a public service, an immunity to the antidiscrimination provisions of the Code not enjoyed by any other person in this Province. Moreover, in practice, it would deny to gays and lesbians the protection from discrimination that the Code provides to others. In the words of the Supreme Court’s decision in \textit{Vriend}, . . . this clause would send “a strong and sinister message” that “gays and lesbians are less worthy of protection as individuals in Canada’s society.”\textsuperscript{205}

The existence of an official, ex ante policy allowing discrimination has serious consequences, whether the

\begin{itemize}
\item \textsuperscript{201} \textit{Id.}
\item \textsuperscript{202} \textit{Id.}
\item \textsuperscript{203} \textit{Id.} at para. 108.
\item \textsuperscript{204} \textit{Id.} at para. 161.
\item \textsuperscript{205} \textit{Id.} at para. 158.
\end{itemize}
discriminatory policy at issue is the exemption permitting a marriage commissioner to refuse to marry a couple, or the single entry point system discussed above permitting marriage commissioners to opt out of performing these marriages behind the scenes. Such a policy can be implemented and discussed with co-workers, staff, and supervisors, and effectively conveys the message that this form of discrimination on the basis of sexual orientation is understood, expected, tolerated, and legitimized. Consider, by contrast, whether society would tolerate an official system allowing people, directly or behind the scenes, to discriminate on the basis of race or religion. The institutionalization of discrimination may serve to perpetuate, magnify, and increase it.

Finally, the concurrence asserted, “if the proposed exemptions were constitutionally acceptable, [then] so too would be virtually any legislative [exemption] allowing service providers to discriminate against same-sex couples, in the public or private sphere, on the basis of religious disapproval of the “same-sex lifestyle.” The logic underlying the exemptions, according to Justice Smith, would be to permit marriage commissioners to refuse to solemnize same-sex marriages, since in their view, performing such a marriage would connote approval of same-sex relationships, and this conflicts with their religious beliefs.

On this logic, Justice Smith stated, a wide range of service providers who disapprove of same-sex relationships could also try to justify discriminating against LGBTQ people if the disapproval was on religious grounds, which she found it frequently is. These other service providers could include persons who rent halls for marriage celebrations, sell marriage licenses, rent living accommodation to married couples, and provide restaurant meals or entertainment, as mentioned above.

It is arguable whether solemnizing a marriage is akin to

206 Id. at para. 103.
207 Id.
208 Id. at para. 144.
209 Id. at para. 145.
210 Id.
selling popcorn to a couple at a romantic movie, given the closer nexus of the marriage commissioner and greater degree of personal involvement in the union. However, there is no question that the logic of disapproval on its own could apply in far too many situations of discrimination such as the examples discussed by the court in Brockie. Both a printer who refused to print editorial content and a printer who refused to print a business directory aimed at LGBTQ interests would likely have asserted their refusal on the basis of their disapproval.

The conclusion reached by the majority and concurrence in Marriage Commissioners was the correct one, and a great deal of their analysis is extremely helpful in advancing the law about belief-based exemptions in the context of discrimination on the basis of sexual orientation. The proposed exemptions, if enacted, would not have been constitutional due to their deleterious effects, the particular harms of official and ex ante policies permitting discrimination, and the role of the marriage commissioners as agents of the province. For all that, it should be recognized that for some individuals, solemnizing a same-sex marriage could go against their sincerely held religious beliefs, and they may choose to leave their position as marriage commissioner rather than be compelled to create the union. Being forced out of work is a significant and coercive result. However, when weighed against the proposed exemptions’ serious individual and social harms, these exemptions would not be constitutional, and the court was correct in finding that the right to equality in this case should prevail.

III. CONCLUSION: CAN FREEDOM OF RELIGION AND EQUALITY BE RECONCILED? GUIDING PRINCIPLES AND CONCLUSIONS

A serious and dedicated approach to equality, freedom of religion, and dignity acknowledges that people hold diverse views which will come into conflict from time to time. In addressing belief-based exemption cases, a few themes emerge that may help guide future cases. These can be derived from existing Canadian case law and from a critical analysis of some of the decisions.

First, the issues at play are complex and cannot be resolved in
the abstract. Freedom of religion and equality do not lend themselves to reconciliation in many of the belief-based exemption cases, and resorting to higher-order principles will generally not provide a solution, as these principles frequently apply to both rights.

Religion is a subset of equality, as it is one of the prohibited grounds of discrimination. Freedom of religion also embodies elements of freedom of association and liberty. Equality likewise incorporates these elements. Both equality and freedom of religion may relate to a value system, which is sincerely believed in and deeply cherished by many individuals. Indeed, the two are not mutually exclusive. For many religious individuals, a deeply held and cherished belief in the inherent dignity, equality, and worth of all people comes from their religious faith. And numerous religious individuals and groups have been involved in various antidiscrimination causes, including efforts to recognize same-sex marriage in Canada. Finally, the concept of dignity—generally recognized as being at the heart of the right to equality—also underlies the protection of religious freedom, as pointed out by the dissent in Chamberlain and the court in Brockie. Thus, for example, if one is forced to pray to a foreign deity, touch an impure object, use one’s artistry or talents to tell a false story, or have one’s children taught an ideology that runs counter to one’s deeply held values—this may amount to a violation of dignity.

In the result, where an individual or entity seeks an exemption based on their religious beliefs from the duty to provide services to the public without discrimination on the basis of sexual orientation, a theoretical reconciliation between equality and freedom of religion will often not be possible. Instead, individual exemptions should be examined on the basis of their concrete facts in context and on a case-by-case basis. Courts should consider evidence as to the specific individuals’ beliefs, the actual harms that would result from infringing a person’s religious freedom, as well as the actual harms from the discrimination. The

---

examination should include not only the impact on individuals, but also the broader social, legal and historical discrimination and context at issue, as considered by the Tribunal in Eadie and the court in Marriage Commissioners. The section 1 Charter test provides a helpful framework for such an analysis and balancing exercise.

Given that reconciliation between the rights is not likely, the second theme that emerges is that one party’s fundamental right may be violated, and a coercive solution may be necessary. This conclusion, while uncomfortable, would result regardless of which way the court decided. And coercive solutions are consistent with many other laws that force individuals to act in a manner they might not otherwise have chosen, including the human rights laws that prohibit discrimination.

Third, adjudicators would do well to reflect on their own prejudices. Courts should avoid the kinds of hetero-normative assertions made by the dissent in Chamberlain, and focus instead on a respect for diversity on the basis of sexual orientation and gender identity.

The fourth guiding principle, similarly, addresses a particular preconception that may preclude adjudicators from engaging seriously with the rights on both sides of the conflict. Some people sincerely feel that their religion prohibits them from participating in certain events or activities. Courts should not summarily dismiss, as the concurrence did in Marriage Commissioners, the difficulties faced by someone required to choose between their work, and their conscience and dignity, if forced to act against their beliefs. Nor should an adjudicative body impose its own understanding of religion on the person seeking religious freedom. What the court can and should do is examine the evidence concerning both the sincerity of the belief and the possible harms that could occur if a person’s religious freedom is infringed. Courts must take seriously the possibility that individuals may be concerned about harming their relationship with their Lord or about the “righteous judgment of God.” Such sensitivity and consideration is not the end of the process, but it is necessary to examining and balancing the real

212 See supra notes 184–90 and accompanying text.
issues at stake.

The fifth principle that emerges from the case law is that official, sweeping, or ex ante exemptions permitting discrimination on the basis of sexual orientation in the provision of services should be avoided. Such exemptions may send a message that discrimination is to be expected, tolerated, and legitimized, as reasoned by the concurrence in *Marriage Commissioners*, and could lead to the institutionalization and perpetuation of the discrimination.

Finally, belief-based exemptions, if allowed at all, must be extremely rare and exceptional. None of the cases discussed in this Article presented a justified belief-based exemption. However, it is possible that such situations may occur. For example, as discussed, if a person is asked to participate in creating a product with which they disagree, an exemption might be justified if the requested product would be derogatory or hateful towards the service provider and/or hateful towards a group protected under the antidiscrimination laws.

New situations may also bring to light the possibility of other exemptions. For instance, services that require the exceptionally intimate and personal involvement of a service provider in a relationship with which they disagree may generate a different conclusion than services requiring a more tenuous connection. For example, a sex therapist who refuses to work with a same-sex couple likely has a stronger argument, on the basis of intimacy, than an electrician who refuses to rewire that couple’s home. Even in the case of the sex therapist, however, and assuming that bona fide occupational requirements were not at issue, a case-specific, contextual analysis is required that would consider the reason for the refusal, the question of minimal impairment, and the deleterious and salutary effects of granting the requested exemption. Such an analysis may conclude in favour of the sex therapist because of the exceptional intimacy required, or it may reach the conclusion that the therapist should find a less intimate occupation.

The danger of the slippery slope is significant in belief-based exemption cases, and balancing exercises to assess these cases give adjudicators wide discretion. Absent a strict exceptionality standard, there could be too many discriminatory refusals. Thus,
for example, even if one were to take the standard from *Brockie*, in which a service provider may be exempt if the material conflicts with their core beliefs, and make it stricter by adding a requirement that the service provider must have a direct and personal involvement in the work, in its outcome, or with the customer, the list of possible exemptions would still be long. After all, many service providers are personally involved in their work and/or with customers, and some individuals do hold beliefs that would lead them to discriminate on the basis of sexual orientation or gender identity. Refusals that meet this description could include (to use recent Canadian and American examples): a florist asked to enhance the beauty of a wedding; bed and breakfast owners renting out a room in their home; or a wedding photographer whose job it is to create a lasting image of love and romance.

It is not difficult to imagine any number of other situations that could also meet this standard, including: an architect asked to design a family home for a family they believe should not exist—or a builder or interior designer with similar views; a lawyer, banker, or investment adviser asked to protect the property of children or a spouse in a family arrangement the service-provider considers invalid; a wedding planner, dress maker, or barber who objects to the wedding, or perhaps a hairdresser or manicurist who objects to the romance; a teacher whose beliefs run counter to parts or all of the curriculum; a police officer whose job may involve risking their own safety to protect an event (a Pride parade), ceremony (a same-sex marriage), or person (a politician active in promoting LGBTQ rights) that the police officer does not believe warrants such protection; a health provider or hospital worker asked to convey information or provide assistance to the loved one of a patient—or asked to deliver a baby in a family arrangement of which they disapprove.

It is difficult to accept the idea that doctors or police officers might refuse to save or protect individuals in any circumstances. It is also difficult to contemplate a society in which refusals to provide service that single out customers based on their sexual orientation could be commonly tolerated and institutionalized. It is for this reason that exemptions, if any, should be extremely rare and exceptional. After all, the human rights laws were
passed because of people’s refusal to countenance a society in which discrimination would be tolerated. To countenance it now is deeply concerning in the face of pervasive and often socially accepted discrimination against people on the basis of their sexual orientation or gender identity. Accepting exemptions in any but the most exceptional of circumstances could lead to many acts of refusal and exclusion, and could legitimize and perpetuate this discrimination, undermining the very purpose of our human rights laws. Such a result would be out of balance and would not justify violating the dignity, equality, and fundamental rights of people in Canada.
RELIGIOUS FREEDOM AND EQUAL TREATMENT: A UNITED KINGDOM PERSPECTIVE

Karon Monaghan QC*

I. INTRODUCTION

Religious discrimination is institutionalized in the United Kingdom. The Church of England occupies a privileged space in the United Kingdom’s constitutional arrangements. The Queen, as a constitutional monarch, holds the title “defender of the Faith and Supreme Governor of the Church of England.” Twenty-six Bishops of the Church of England sit in the legislature’s upper house, the House of Lords. The position of these “Lords

* Barrister, Matrix Chambers, Griffin Building, Gray’s Inn, London, United Kingdom WC1R 5LN, +44 (0)20 7404 3447, karonmonaghan@matrixlaw.co.uk. Author of KARON MONAGHAN QC, MONAGHAN ON EQUALITY LAW (2d ed. 2013).

1 New constitutional settlements with Wales, Scotland, and Northern Ireland have devolved specified matters to regional assemblies but Parliament retains responsibility for certain matters and the fundamental constitutional arrangements remain in place (subject to referenda in the case of Scotland and Northern Ireland). Government of Wales Act, 1998, c. 38; Government of Wales Act, 2006, c. 26; Scotland Act, 1998, c. 46, §1, sch. 5; Northern Ireland Act, 1998, c. 47, §§ 1–4, sch. 3.


3 All Bishops are men as the Church of England does not allow women in the episcopate (that is, to become Bishops).

Spiritual” is unique: no representatives from other religious organizations are entitled as of right to membership of the House of Lords.\(^{5}\)

These facts establish the United Kingdom as an essentially Christian State,\(^{6}\) and this is reflected in a number of legal measures. For example, education law in the United Kingdom requires that every pupil take part in a daily act of collective worship, which must be wholly or mainly of a broadly Christian character.\(^{7}\) It is difficult to understand why this state of affairs is tolerated in an apparently modern, pluralistic, liberal democracy. However, notwithstanding the wide-ranging constitutional reforms put into place by the last Labour Government,\(^{8}\) the right

lords-appointment/ (last visited Feb. 22, 2014). Presently, there are only twenty-three Lords Spiritual. Lords by Party, Type of Peerage and Gender, PARLIAMENT.\(^{5}\) uk, http://www.parliament.uk/mps-and-offices/lords/composition-of-the-lords/ (last visited Feb. 22, 2014). It is customary for one of the Lords Spiritual to read prayers in the House at the beginning of each day’s proceedings, and for this purpose a rota is furnished to the House in which certain of the Lords Spiritual are allotted periods for which they are responsible for this duty. See 78 HALSURY’S LAWS OF ENGLAND, para. 833 (5th ed. 2010).

\(^{5}\) Bishops in the House of Lords, supra note 4.

\(^{6}\) David Cameron, Prime Minister, has declared of the United Kingdom: “We are a Christian country[,] and we should not be afraid to say so.” David Cameron Says the UK is a Christian Country, BBC NEWS (Dec. 16, 2011), http://www.bbc.co.uk/news/uk-politics-16224394?.

\(^{7}\) School Standards and Framework Act, 1998, c. 31, § 70 (UK) (stating that all students in attendance must take part in collective worship). “Collective worship is of a broadly Christian character if it reflects the broad traditions of Christian belief without being distinctive of any particular Christian denomination.” Id. sch. 20. A parent may request that a pupil be wholly or partly excused from receiving religious education and the pupil may be so excused. Id. § 71(1). The 1998 Act does not apply to Scotland. Id. § 145. For Scotland, see SCOTTISH GOVERNMENT CIRCULAR, PROVISION OF RELIGIOUS OBSERVANCE IN SCOTTISH SCHOOLS (2005).

\(^{8}\) See House of Lords Act, 1999, c. 34 (restricting membership of the House of Lords by virtue of a hereditary peerage); Constitutional Reform Act, 2005, c.4 (formally separating the state’s judicial and legislative functions by the creation of a Supreme Court and making consequential changes). See also Government of Wales Act, 1998, c. 38; Government of Wales Act, 2006, c. 26; Scotland Act, 1998, c. 46; Northern Ireland Act, 1998, c. 47 (redefining the relationship between the three constituent countries of the United Kingdom
of senior Church of England Bishops to sit in the House of Lords was retained and no changes were made upsetting the primacy of Christianity, and the Church of England in particular, in public life.  

This privileging of Christianity over other religious or nonreligious beliefs has, until recently, been aggravated by the absence of any legal protections against religious and belief-based discrimination. Certain “religious” groups have for some time been legally categorized as “ethnic groups.” By this route, certain religious minorities discriminated against because of their religious beliefs have enjoyed the protection of laws against race discrimination, but by and large discrimination connected to religion and belief remained outside of the law. In school and at work, therefore, working hours, holidays, dress codes, and the like were generally constructed around a Christian norm free

9 See generally DEPT FOR CONSTITUTIONAL AFFAIRS, CONSULTATION PAPER, CONSTITUTIONAL REFORM: NEXT STEPS FOR THE HOUSE OF LORDS (2003).

10 The position was somewhat different in Northern Ireland because of its particular political context. See Fair Employment and Treatment (Northern Ireland) Order, 1998, SI 1998/3162 (N. Ir. 21), arts. 2–4, 74 (addressing discrimination connected to religious belief and political opinion, and affirmative action).

11 Seide v. Gillette Indus. Ltd., [1980] I.R.L.R. 427 (Jews); Mandla v. Dowell Lee, [1983] 2 A.C. 548 (Sikhs). These cases recognize the relationship between culture and religion: many people belonging to particular racial groups see their religion as occupying more of a cultural or political space in their lives, and “faith” as less significant to their identity. In Northern Ireland, therefore, where there has been compelling protection against discrimination connected with religion and belief in employment and related fields for some time, it is well understood that these protections were passed to address the political and cultural, rather than the theological, divides between the Catholic and Protestant communities. See generally Fair Employment and Treatment (Northern Ireland) Order, 1998, SI 1998/3162 (N. Ir. 21).

12 This caused a good deal of controversy. When promoting new laws addressing religious discrimination, the Lord Chancellor observed that this “remedie[d] the anomaly whereby members of some religions are protected against discrimination in the provision of goods, facilities and services, but members of other religion or belief groups are not.” Michael Rubenstein, Equality Act 2006: A Guide, 151 EQUAL OPPORTUNITY REV. 21, 25 (2006).
from the scrutiny of equality law, sometimes causing real disadvantage to those holding non-Christian beliefs.\(^{13}\)

The distinction between those religious groups deemed ethnic groups and those not, meant that laws outlawing the incitement of racial hatred protected some religious groups but not others.\(^{14}\) At the same time, the (now abolished)\(^{15}\) common law offense of blasphemy protected Christians only against certain forms of insult. Blasphemy was an indictable offense at common law consisting of the publication of any “contemptuous, reviling, scurrilous or ludicrous matter relating to God, Jesus Christ, or the Bible or the formularies of the Church of England.”\(^{16}\) No comparable protection was afforded to other religious groups.

This disparity was ameliorated by a series of legislative measures. First, the Employment Equality (Religion or Belief) Regulations 2003 made discrimination connected with a person’s religion or belief unlawful in employment and related fields.\(^{17}\) Second, Part 2 of the Equality Act 2006 outlawed discrimination connected to religion and belief in the provision of goods, facilities, and services; in the disposal and management of premises; in education; and by public authorities.\(^{18}\) While these laws have now been revoked and repealed, respectively, they have been largely consolidated and are now reflected in near-identical provisions in the Equality Act 2010.\(^{19}\) Third, the Racial and Religious Hatred Act 2006 now criminalizes “threatening words or behavior, or displays [of] any written material which is

---


\(^{14}\) Public Order Act, 1986, c. 64, § 18 (UK).

\(^{15}\) See Criminal Justice and Immigration Act 1998, c. 4, § 79 (1) (UK).


\(^{17}\) The Employment Equality (Religion or Belief) Regulations, 2003, S.I. 2003/1660, §§ 3, 6 (UK). These regulations covered employees, contract workers, officer holders, the police, barristers, advocates, partnerships, trade organizations, qualifications bodies, providers of vocational training, employment agencies, career guidance services and further and higher education institutions.


\(^{19}\) Equality Act, 2010, c. 15, pts. 2–14 (UK) and associated Schedules.
threatening” if the person so acting “intends thereby to stir up religious hatred.” These measures were directed at securing greater equality for minority religious belief-holders, and for those without any religious belief at all. However, all of them have proved controversial.

The extension of the hate speech provisions to religious hatred was highly contentious. It led to a campaign by well-known comedians and others concerned about the censoring of religious criticism, whether that criticism was through satire or otherwise. However, the difficulty faced by those resisting this change was that some forms of religious criticism were already outlawed by race hate and blasphemy laws, but only in the case of certain religions. This discriminatory distinction had a pernicious effect on nonprotected religious groups who suffered the public ignominy of apparently legally sanctioned second class status and was plainly difficult to justify. In an environment of increased hostility towards Muslims, in particular, a position where the law protected Christians but left Muslims unprotected was simply unsustainable. The comedians and their protagonists did not succeed in their attempts to have these new laws criminalizing religious hate speech blocked, and the discrimination inherent in hate speech laws that protected some religious groups but not others was eliminated.

The most controversial protections have proved to be those conferring nondiscrimination rights on those who have been disadvantaged because of their religion or belief. This is because such protection, perhaps inevitably, extends beyond disadvantages associated with the holding of a particular belief, to religiously motivated acts. The problem is not that a Christian, a Muslim, or a Jew, as the case may be, can call on the protection of antidiscrimination law when discriminated against because they adhere to particular religious beliefs. Instead, it is that they may rely on those same laws for license to act in ways said to be in pursuance of those beliefs. Sometimes this is uncontroversial. A

20 Racial and Religious Hatred Act, 2006, c. 1, § 29B(1), sch. 1 (UK) (amending the Public Order Act 1986, c. 64 (UK)).
21 Id. § 29A.
woman who wishes to wear a crucifix at work, absent any specific and compelling requirements of the job, can easily be accommodated, as can a turban-wearing Sikh, or a Muslim seeking to observe prayers on certain days of the week. More problematic, and increasingly so given changes in United Kingdom and regional antidiscrimination and human rights law, are those cases where the assertion of a religious belief through practice impinges on the rights of others, particularly women and sexual minorities. It is this issue that is the focus of this Article.

Section II of this Article will identify the various legal measures operating in the United Kingdom that confer the right to religious freedom and the right to freedom of religious expression. Section III examines the way in which competing rights, particularly those affecting women and sexual minorities, are managed within those legal frameworks. Section IV concludes that both statutory law and case law afford considerable protection to religious groups and individuals within them, sometimes at the expense of otherwise highly protected classes, women and sexual minorities in particular. This occurs notwithstanding that in practice those religious groups, specifically Christians, who most commonly claim such protections against the interests of women and sexual minorities, can in no sense be regarded as forming a “minority” or as being socially or structurally disadvantaged.

II. FREEDOM OF RELIGION: THE LEGAL FRAMEWORK

There are four main sources of legal protection against discrimination connected to religious belief in the United Kingdom. These derive from the European Convention on Human Rights and Fundamental Freedoms, and its effective transposition into United Kingdom domestic law through the Human Rights Act 1998, European Union law, and the United Kingdom’s Equality Act 2010. These legal protections originate from varied historical and political imperatives and, though they

inform each other, offer discrete routes to protection, and address religious freedoms and nondiscrimination rights differently.

These measures guarantee the right to freedom of religion and provide compelling protection against religious discrimination. The right to freedom of religion, as protected by these laws, is not restricted to the enjoyment of the right to freedom of conscience (the so-called “forum internum”), but also to the right to manifest religious belief (the so-called “forum externum”). These same laws also provide protection to women and sexual minorities against discrimination. The rights guaranteeing religious freedoms and the dignity and equality rights of women and sexual minorities, in particular, can on occasions conflict. The method by which the coexistence of these rights is managed in an increasingly diverse and pluralistic society varies, but in the case of each of the legal schemes addressed below, resolving such conflicts is complex and controversial.

A. European Convention on Human Rights and Fundamental Freedoms

1. Religious Freedom

The European Convention on Human Rights and Fundamental Freedoms (“ECHR”), to which all Member States of the Council of Europe are party, contains provisions addressing religious freedom, and nondiscrimination. This reflects the origins of the ECHR, which is firmly rooted in the atrocities of the Second World War. The ECHR was a response to the horrors perpetrated against the Jews, and the commitment to preventing the repeat of any such genocide. Such commitment


\[26\] New members are expected to ratify the ECHR at the earliest opportunity. See EUR. PARL. ASS., Resolution 1031 (1994).

\[27\] For more information, see generally DAVID HARRIS ET AL., LAW OF THE EUROPEAN CONVENTION ON HUMAN RIGHTS (2d ed. 2009).
provided the impetus for embedding the rights and freedoms addressing both religion and nondiscrimination in a legally binding instrument.

As history demonstrates, there can be a close relationship between religion and race. This is especially true in Europe where religion has often been used as a proxy for race. This is evident from the commission of crimes against humanity in Europe, including the atrocities of the Second World War and those after, specifically the “ethnic cleansing” and genocide of Bosnian Muslims under the leadership of Slobodan Milošević. It is also apparent from the rise of political parties across Europe promoting an anti-Islamic discourse.28 The need for robust protections against religious discrimination cannot, therefore, be doubted.

The specific provision in the ECHR directed at protecting religious freedom confers the right to freedom of thought, conscience, and religion as well as the right to manifest religion or belief, in worship, teaching, practice, and observance.29 In the latter case, that freedom is qualified. An interference with the right to manifest one’s religion or beliefs may be justified where

28 “In the last two decades, parties promoting an anti-Islam discourse have had sufficient electoral success to be represented in the national parliaments of a considerable number of European countries including Austria, Belgium, Denmark, France, Italy, the Netherlands, Norway and Switzerland.” AMNESTY INT’L, CHOICE AND PREJUDICE: DISCRIMINATION AGAINST MUSLIMS IN EUROPE 15 (2012). The Council of Europe’s Commissioner for Human Rights has noted that: “[O]pinion polls in several European Countries reflect fear, suspicion and negative opinions of Muslims and Islamic culture. These Islamophobic prejudices are combined with racist attitudes—directed not least against people originating from Turkey, Arab countries and South Asia.” Thomas Hammarberg, European Muslims are Stigmatised by Populist Rhetoric, HUMAN RIGHTS COMMENT (Oct. 28, 2010, 9:19 AM), http://commissioner.cws.coe.int/tiki-view_blog_post.php?postId=99; see also ISLAMOPHOBIA Watch: Documenting Anti Muslim Bigotry, ISLAMOPHOBIA-WATCH.COM, http://www.islamophobia-watch.com/islamophobia-watch/category/uk (last visited Feb. 22, 2014).

29 ECHR, supra note 25, art. 9. This is the only Convention provision that may be relied upon by an organization, as well as an individual. See Church of Scientology Moscow v. Russia, App. No. 18147/08, 46 Eur. Ct. H.R. 16, para. 81 (2008); Supreme Holy Council of the Muslim Cmty. v. Bulgaria, 41 Eur. Ct. H.R. 3, para. 74 (2005).
certain criteria are met, in particular where the interference pursues one of a list of enumerated aims and is proportionate.\(^3\)

Unsurprisingly given its provenance, the European Court of Human Rights ("ECtHR"), the ultimate arbiter of disputes under the ECHR, regards the guarantee of freedom of thought, conscience, and religion, enshrined in the ECHR, as "one of the foundations of a ‘democratic society.’"\(^3\)

The religious freedoms protected by the ECHR are broad because, for good reason, the concept of “religion” is not prescribed. Subject to certain minimum criteria being met, whether a belief is protected under the ECHR is generally not a question for the secular courts. To permit one branch or another of the State to delineate those beliefs worthy of respect and those not, would be to raise just the dangers that recent history shows still exist. Instead, the court will generally ask only whether the belief in question is sincerely held. Similarly, whether any act is a “manifestation” of such a sincerely held religious belief is, in the usual case, to be adjudged by the believer herself. The courts do not engage in any assessment of the validity of the belief that is said to drive the actions (or “manifestation”) in issue; they ask only whether an individual genuinely holds that belief.\(^3\)

\(^3\) ECHR, supra note 25, art. 9(2) ("Freedom to manifest one’s religion or beliefs shall be subject only to such limitations as are prescribed by law and are necessary in a democratic society in the interests of public safety, for the protection of public order, health or morals, or for the protection of the rights and freedoms of others."); see also R (SB) v. Governors of Denbigh High Sch., [2006] UKHL 15, [20]\(^3\)-[32].

\(^3\) Sahin v. Turkey, 2005-XI Eur. Ct. H.R. 175. There are hints, however, of a privileging here of Christianity over Islam. Compare Karaduman v. Turkey, 74 Eur. Comm’n H.R. Dec. & Rep 93 (1993) (holding that there was no interference with the rights guaranteed by Article 9 when a University failed to award the applicant, a Muslim, a diploma because she refused to produce a photograph of herself without her headscarf), with Lautsi v. Italy, 54 Eur. H.R. Rep. 3 (2012) (holding that the presence of crucifixes in State-school class rooms did not violate Article 9), with Eweida v. United Kingdom, App. Nos. 48420/10, 59842/10, 51671/10, and 3516/10, 2013 Eur. Ct. H.R., available at http://hudoc.echr.coe.int/sites/eng/pages/search.aspx?i=001-115881 (holding that the applicant’s rights under Article 9 were violated when she was prohibited from wearing a crucifix at work).

\(^3\) R (Williamson) v. Sec’y of State for Educ. & Emp’t, [2005] 2 A.C. 246 (H.L.) [22]-[23] (appeal taken from Eng.).
a low threshold for the purposes of determining whether a belief is protected by the ECHR. According to the ECtHR:

\[
\text{[t]he right to freedom of thought, conscience and religion denotes views that attain a certain level of cogency, seriousness, cohesion and importance . . . . Provided this is satisfied, the State’s duty of neutrality and impartiality is incompatible with any power on the State’s part to assess the legitimacy of religious beliefs or the ways in which those beliefs are expressed.}^{33}\]

In particular, it is not necessary or appropriate for the court to question the extent to which other members of the religious group to which a person belongs, and to membership of which they attribute the belief in question, subscribe to that belief or engage in that expression. The answers to any such questions are not relevant in deciding whether that person herself has such a belief, nor are they relevant to whether the belief can be categorized as “religious.” This approach ensures that the religious freedoms guaranteed by the Convention are given very wide reach.

The ECtHR has acknowledged that even in cases where the religious belief asserted meets the necessary threshold, not “every act which is in some way inspired, motivated, or influenced by it constitutes a ‘manifestation’ of the belief. Thus, for example, acts or omissions which do not directly express the belief concerned or which are only remotely connected to a precept of faith” are not protected.\(^{34}\) Nevertheless, the right to manifest religious belief

---


is broad: there need only be “the existence of a sufficiently close and direct nexus between the act and the underlying belief . . . . In particular, there is no requirement on the applicant to establish that he or she acted in fulfillment of a duty mandated by the religion in question.”  

This has particular ramifications for acts done in pursuance of religious beliefs that discriminate against others, usually women and sexual minorities.  

In addition to explicit guarantees relating to freedom of religion, the ECHR contains an open-textured equality clause, triggered when any complaint falls within the scope of one or another of the substantive Convention rights. These include, for example, the privacy provision, which has historically been used to protect the rights of sexual minorities, and the family

35 Id.

36 Children are also often affected, though consideration of this impact is outside the scope of this Article.

37 ECHR, supra note 25, art. 14 (“The enjoyment of the rights and freedoms set forth in this Convention shall be secured without discrimination on any ground such as sex, race, colour, language, religion, political or other opinion, national or social origin, association with a national minority, property, birth or other status.”).

38 See id. arts. 2–13; protocol no. 1, arts. 1–3; protocol no. 4, arts. 1–4; protocol no. 6, arts. 1–2; protocol no. 7, arts. 1–5; protocol no. 7, arts. 1–5; protocol no. 12, art. 1 (which contains a free standing nondiscrimination right but which is neither signed nor ratified by the United Kingdom); protocol no. 13, art. 1.

39 Id. art. 8 (“Everyone has the right to respect for his private . . . life.”).

40 See, e.g., Dudgeon v. United Kingdom, 4 Eur. Ct. H.R. 126 (1981) (holding that the criminalizing of homosexual acts between consenting adults was a breach of Article 8); Lustig-Prean v. United Kingdom, 29 Eur. Ct. H.R. 449 (1999); Smith v. United Kingdom, 29 Eur. Ct. H.R. 493 (1999) (holding that the investigation into and subsequent discharge of personnel from the Armed Forces on the basis that they were homosexual was a breach of Article 8); A.D.T. v. United Kingdom, 31 Eur. Ct. H.R. 33 (2001) (holding that legislation criminalizing homosexual acts between men in private, and a fortiori prosecution and conviction, was a breach of Article 8); Goodwin v. United Kingdom, 35 Eur. Ct. H.R. 447 (2002) (holding that a failure to grant legal recognition of a person’s gender re-assignment was a breach of Article 8). Domestically, courts have recognized similar rights. See, e.g., Ghaidan v. Ghodin-Mendoza, [2004] 2 A.C. 557 (H.L.) (holding that tenancy succession rules which treated survivors of homosexual partnerships less favorably than survivors of heterosexual partnerships breached Articles 8 and 14).
life provision.\footnote{Schalk v. Austria, 2010 Eur. Ct. H.R. (holding that the relationship between a cohabiting same-sex couple living in a stable de facto partnership, fell within the notion of “family life,” just as the relationship of a different-sex couple in the same situation would).} This clause gives wide meaning to the concept of “discrimination,” so as to address not merely formal distinctions in treatment, but also structural and institutional forms of inequality.\footnote{See, e.g., Jordan v. United Kingdom, 37 Eur. Ct. H.R. 2 (2003) (a general policy or measure that has disproportionately prejudicial effects on a particular group, may be considered discriminatory notwithstanding that it is not specifically aimed or directed at that group and breach Article 14); Thlimmenos v. Greece, 31 Eur. Ct. H.R. 411 (2000) (the right not to be discriminated against in the enjoyment of the rights under the ECHR is also violated when States without an objective and reasonable justification fail to treat differently persons whose situations are significantly different); D.H. v. Czech Republic, 47 Eur. Ct. H.R. 3 (2008) (a breach of Article 14 may occur where a general policy or measure which, though couched in neutral terms, results in disproportionately prejudicial effects against a group); Opuz v. Turkey, 50 Eur. Ct. H.R. 695 (2010) (holding that general and discriminatory judicial passivity in Turkey on the issue of domestic violence, albeit unintentional, mainly affected women and the violence could therefore be regarded as gender-based violence and a form of discrimination against women, and that in the circumstances, the overall unresponsiveness of the judicial system and impunity enjoyed by the aggressors indicated that there was insufficient commitment to take appropriate action to address domestic violence and this amounted to a breach of Article 14).} It covers discrimination on the basis of a long list of enumerated grounds, including sex and religion, and “other status,” which has long since been held to be sufficiently expansive to cover sexual orientation.\footnote{Salgueiro da Silva Mouta v. Portugal, 31 Eur. Ct. H.R. 47 (1999).} The prohibition on discriminatory treatment appears absolute, but the courts have approached the interpretation of the equality clause on the assumption that it implicitly allows for the justification of discrimination. For any discrimination to be lawful, however, it must be objectively and reasonably justified. The level of scrutiny applied to any discriminatory act or measure for the purposes of determining whether it is justified will reflect the social and legal importance placed upon the relevant distinguishing characteristic. Distinctions based on “suspect” grounds, those being, inter alia,
RELIGIOUS FREEDOM AND EQUAL TREATMENT 685

religion, sex, and sexual orientation will be subject to particularly rigorous scrutiny and will require “very weighty reasons” if they are to be justified.

The scheme of the ECHR with its broad protections promotes, or should promote, respect for diversity and for a plurality of divergent beliefs. Difficulties arise, however, because the guarantees of freedom of religion and the prohibition on discrimination may be used by members of more than one suspect class with respect to the same act. This typically occurs when the act in issue discriminates against women or sexual minorities, but the proscribing of it discriminates against those holding particular religious beliefs. While the possibility of justifying what would otherwise be prohibited discrimination seems to be the obvious route by which such conflicts might be resolved, this results in uncertainty. Where the balance will be struck in any particular case can be hard to predict and will sometimes depend upon unknowable factors, such as the personal views of the judge hearing the case. Further, little guidance can be found in the ECtHR’s jurisprudence because of its reliance on the “margin of appreciation.”

In essence, the doctrine of the “margin of appreciation” affords Member States a degree of latitude in their compliance with the Convention rights, reflecting the principle of subsidiarity and the ECtHR’s cognizance of the special

---

48 For a full discussion, see generally R. CLAYTON & H. TOMLINSON, THE LAW OF HUMAN RIGHTS § 6.42 (2d ed. 2009).
49 This principle requires that States are to be held primarily responsible for securing compliance with the ECHR, supra note 25. See generally H. Petzold, THE CONVENTION AND THE PRINCIPLE OF SUBSIDIARITY, in THE EUROPEAN SYSTEM FOR THE PROTECTION OF HUMAN RIGHTS (R. St. J. Macdonald, eds., 1993). This principle is reflected in both case law under the ECHR and in European Union law. See Treaty on European Union, Mar. 2, 2010, 2010 O.J. (C83/13), art. 5(3) [hereinafter TEU] (“Under the principle of subsidiarity, in areas which do not fall within its exclusive competence, the Union shall act only if and insofar as the objectives of the proposed action
conditions that might exist at local level. Its effect is that, in recognition of the status of the ECtHR as a supranational court, the ECtHR will refrain from laying down strict, universally applicable principles. Instead the ECHR confers on States a degree of discretion as to the means by which, or indeed the extent to which, they implement the Convention in domestic law. This doctrine is generally applied “when it comes to striking a balance between competing Convention rights.” This can create particular problems in controversial areas where there is a lack of consensus across States, such as in the interface between religious rights and the rights of sexual minorities.

2. The Court’s Approach to Competing Claims

The ECHR does not contain any explicit protection for sexual minorities. Instead, the Court’s developing jurisprudence in relation to sexual minorities has largely fashioned what are in essence nondiscrimination rights, from the prohibition on interferences with private life. As the scope for asserting the right to nondiscrimination in relation to sexual orientation has expanded, so the opportunity for a conflict between the right to freedom of religion and the rights to nondiscrimination has increased. This has resulted in the court resorting to the “margin of appreciation.”

Two cases illustrate this. Both concern the right to freedom of religion, and the right to nondiscrimination enjoyed by sexual minorities. The two cases are Ladele v. United Kingdom and McFarlane v. United Kingdom. In these cases the applicants,

---

52 See supra notes 38–40 and accompanying text.
53 Ladele and McFarlane were heard together with Eweida v. United Kingdom. See Eweida, 2013 Eur. Ct. H.R.
both holding orthodox Christian beliefs, declined to provide certain services to gay and lesbian people, namely the registration of civil partnerships\(^{54}\) in *Ladele*, and psycho-sexual counseling in *McFarlane*. In the first case, *Ladele*, the applicant, Lillian Ladele, was a civil registrar and as such was obliged to register civil partnerships (the status which affords legal recognition to same-sex couples in the United Kingdom) as part of her job.\(^{55}\) She complained of indirect discrimination when, having refused to engage in the registration of civil partnerships, she was directed to do so by her employer.\(^{56}\) Ladele’s refusal was contrary to her public authority employer’s equal opportunities policy, and the requirements of her job. However, she held the orthodox Christian belief that marriage is the union of one man and one woman for life and that same-sex unions are contrary to God’s will. She believed, therefore, that it would be wrong for her to participate in the creation of an institution equivalent to marriage (as she saw it) between a same-sex couple.\(^{57}\) Her claims failed in the domestic courts on the ground that requiring her to conduct civil partnerships was justified under the circumstances.\(^{58}\)

Ladele brought a complaint against the United Kingdom before the ECtHR, in reliance upon the right to freedom of religion and the nondiscrimination guarantee under the ECHR. The ECtHR accepted that her employer’s policy of requiring, without exception, that all registrars of births, marriages, and deaths be designated civil partnership registrars had a particularly detrimental impact on her because of her religious beliefs and that it was therefore, prima facie, indirectly discriminatory. However, that was not sufficient to make the requirement unlawful.

As with the right to manifest religious belief, the ECtHR held that in order to determine whether her employer’s policy violated the nondiscrimination guarantee, it was necessary to decide whether the policy was justified as pursuing a legitimate aim, and

\(^{54}\) This is available to same sex partners only. Civil Partnership Act, 2004, c. 33, § 1 (U.K.).


\(^{56}\) *Id.*

\(^{57}\) *Id.*

\(^{58}\) *Id.* at [3].
was proportionate.\textsuperscript{59} The ECtHR noted that the aim of the policy was not limited to providing a service which was effective in terms of practicality and efficiency. It was also to ensure compliance with the employer’s overarching policy of being “wholly committed to the promotion of equal opportunities and to requiring all its employees to act in a way which does not discriminate against others.”\textsuperscript{60} The ECtHR held that same-sex couples are in a similar situation to different-sex couples with regard to their need for legal recognition and protection of their relationships, and were thus in an analogous situation to couples seeking to marry. In those circumstances, the ECtHR considered that the aim pursued by the policy was legitimate.\textsuperscript{61} Further, following its own case law, the court held that differences in treatment based on sexual orientation require particularly serious reasons if they are to be justified. The ECtHR concluded that notwithstanding the serious impact on Ladele (the loss of her job), her employer and the domestic courts, which had rejected her discrimination claim, had not exceeded the margin of appreciation afforded to them, and accordingly there was no breach of the Convention rights.\textsuperscript{62}

In \textit{McFarlane}, the applicant, Gary McFarlane, was employed by a private enterprise providing relationship counseling and psycho-sexual counseling. Contrary to the policies of his employer, McFarlane refused to provide sexual counseling to same-sex couples because of his Christian beliefs. Consequently, he was dismissed from his employment. According to the ECtHR, as in \textit{Ladele}, the most important factor to be taken into account was the fact that the employer’s action was intended to secure the implementation of its policy of providing services without discrimination. The court held that State authorities enjoyed a wide margin of appreciation in deciding where to strike the balance between the right to manifest religious belief and the employer’s interest in securing the rights of others. This meant that the United Kingdom was not in breach of the Convention in

\textsuperscript{60} \textit{Id.} at para. 105.
\textsuperscript{61} \textit{Id.}
\textsuperscript{62} \textit{Id.} at para. 106.
failing to ensure a remedy for McFarlane’s dismissal. As with Ladele, the domestic courts had not exceeded the margin of appreciation available to them.\(^{63}\)

The question left unanswered by these cases is whether the margin of appreciation would have been exceeded if the domestic courts had decided otherwise: that is, that the refusal to provide services to same-sex couples was protected by the right to freedom of religion and that any decision requiring the applicants to deliver such services was unlawful. Whether the right to manifest religious belief would trump the nondiscrimination rights of women, and gay men and lesbians is of particular concern in the United Kingdom, given the role of religion, and specifically Christianity, in public life. Unfortunately, the jurisprudence of the ECtHR does not provide any clear framework for addressing this issue in controversial cases.

**B. Human Rights Act 1998**

The ECHR has been to a large extent transposed into United Kingdom domestic law by the Human Rights Act 1998 (“HRA”), a purely domestic statute. The HRA now allows persons who allege that their “Convention rights”\(^{64}\) have been violated to bring claims in the domestic courts, where previously a complainant could only vindicate those rights by an application to the ECtHR.\(^{65}\) The coming into force of the HRA has given greater prominence to the ECHR within the domestic legal order and has raised society’s consciousness with regard to the rights it protects. This has resulted in a great deal of case law on matters touching upon the issue of religious freedom.

The HRA does not give the Convention rights the same constitutional status seen in the Bills of Rights and other

---

\(^{63}\) Id. at para. 109.

\(^{64}\) See Human Rights Act, 1998, c. 42, § 7 (U.K.). These Convention rights are articles 2–12, 14, 1–18; First Protocol, articles 1–3; and Thirteenth Protocol, article 1.

\(^{65}\) The Human Rights Act 1998 c. 42 does not preclude an applicant from pursuing an application to the ECtHR once all domestic remedies have been exhausted. See ECHR, supra note 25, art. 35.
constitutional instruments in jurisdictions elsewhere.\textsuperscript{66} However, it does provide remedies to victims of a violation of the Convention rights.\textsuperscript{67} Further, although it does not permit the striking down of primary legislation,\textsuperscript{68} the HRA requires that legislation “be read and given effect in a way which is compatible with the Convention rights” so far as it is possible to do so.\textsuperscript{69} Further, while the Convention rights, both as a matter of international law and domestic law, bind only public authorities, a court is treated as a public authority for these purposes.\textsuperscript{70} This gives the Convention rights significant horizontal impact.\textsuperscript{71}

As the HRA is a domestic legal measure, the doctrine of a “margin of appreciation” does not apply to a ruling made under it by the United Kingdom courts. This is for the obvious reason that it is the State to which any “margin” is accorded, and the principle operates only at supranational level. However, the courts have shown a willingness to afford a degree of respect to acts of the legislature (particularly recent ones).\textsuperscript{72} As one member of the Supreme Court\textsuperscript{73} put it:

\begin{quote}
[\textit{W}hen we can reasonably predict that [the ECtHR] would regard the matter as within the
\end{quote}

\textsuperscript{66} It does not allow, therefore, for the striking down of legislation incompatible with the Convention rights as would be common in Bills of Rights and constitutional instruments.

\textsuperscript{67} Human Rights Act, 1998, c. 42, § 7 (U.K.).

\textsuperscript{68} This ensures that the operation of the Human Rights Act 1998 does not result in any challenge to the primacy of Parliament. It allows, however, for the making of a declaration of incompatibility and provides an expeditious route to amending the law in view of that incompatibility, but it does not invalidate the law in the meantime or compel Parliament to amend it. Id. § 4.

\textsuperscript{69} Id. § 3.

\textsuperscript{70} Id. §§ 6(1), 6(3).

\textsuperscript{71} By which it is meant that it impacts on private parties (through the decisions of the courts), as well as State actors. See generally CLAYTON & TOMLINSON, supra note 48.


\textsuperscript{73} Then sitting as a committee of the House of Lords (before the enactment of the Constitutional Reform Act of 2005).
margin of appreciation left to the member states . . . [the Court] should not attempt to second guess the conclusion which Parliament has reached. I do not think that this has to do with the subject matter of the issue, whether it be moral, social, economic or libertarian; it has to do with keeping pace with the [ECtHR] jurisprudence as it develops over time, neither more nor less. 74

This approach has proved to be just as significant domestically as the doctrine of the “margin of appreciation” in the ECtHR, especially where strongly contested matters are in issue. When the legislature has chosen a particular course in an area of controversy, the courts will tend to avoid second-guessing the legislature’s decision, particularly where it is presumed that the ECtHR would not interfere if it were the subject of challenge there. This has broad ramifications, but is specifically relevant to domestic equality laws protecting sexual minorities and the legality of exemptions directed at accommodating religious belief. In short, the courts will be reluctant to interfere with any legislative expression by Parliament as to the balance between competing rights. The impact of this is considered in Section III below.

C. European Union Law

Protection for religious belief and practice in the United Kingdom today is in large part derived from European Union (“EU”) law. There are both similarities and differences as between the protections afforded by the ECHR and EU law. EU law, like the ECHR, guarantees the rights to freedom of religion and to nondiscrimination but these rights are only operative in situations covered by EU law, unlike the ECHR which is of broader impact. 75

74 Countryside Alliance, [2007] UKHL 52 at [126].

75 In addition, the general impact of EU law even within those parameters may be narrower than the ECHR in some circumstances, see, for example, Association de médiation sociale v. Union locale des syndicats CGT (Union départementale CGT des Bouches-du-Rhône and another intervening), [2014] WLR (D) 2.
The EU Treaties regulating the European Union\textsuperscript{76} contain equality guarantees,\textsuperscript{77} as does some secondary legislation under the Treaties.\textsuperscript{78} Importantly too, the European Union Charter of Fundamental Rights\textsuperscript{79} guarantees the right to freedom of religion\textsuperscript{80} and to nondiscrimination.\textsuperscript{81} The nondiscrimination provisions in the various EU legal instruments address discrimination across a number of grounds, including religion, gender, and sexual orientation. Some address the concept of discrimination in an open-textured way, while some adopt more formalistic meanings of discrimination.\textsuperscript{82}

The system of legal rules that flows from the United Kingdom’s accession to the European Economic Community ("EEC"), now the EU, means that EU law is of very great importance to both the interpretation and application of domestic equality law and confers, in some cases, directly effective\textsuperscript{83}.

\textsuperscript{76} TEU, supra note 49; Treaty on the Functioning of the European Union, Mar. 2, 2010, 2010 O.J. (C83/47) [hereinafter TFEU].

\textsuperscript{77} See, e.g., TEU, supra note 49, art. 2; TFEU, supra note 76, arts. 10, 18, 19, 45, 153(1)(i)–(j), 157.


\textsuperscript{79} Charter of Fundamental Rights of the European Union, art. 10, 2000 O.J. (C 364/01).

\textsuperscript{80} Id. at art. 10.

\textsuperscript{81} Id. at art. 21.


\textsuperscript{83} See KARON MONAGHAN, MONAGHAN ON EQUALITY LAW 68 (2d ed.
nondiscrimination rights upon individuals.

D. Equality Act 2010

Finally, the United Kingdom’s Equality Act 2010 (“Act”) outlaws discrimination in certain spheres, on the grounds of religion or belief, gender, gender reassignment, and sexual orientation. The Act is more prescriptive than the ECHR and some parts of European Union law. It enacts closely formulated concepts of discrimination, and to a significant degree the legislature has decreed within it when an interference with the right of an individual or group to be free from discrimination is justified, leaving the courts with little or no discretion. This occurs in a number of areas but in particular at the interface between gender and sexual orientation, and religion, as discussed below.

The concept of “religion” is given wide reach under the Act. The Explanatory Notes to the Act, reflecting the jurisprudence of the ECtHR, state that “[i]t is a broad definition in line with the freedom of thought, conscience and religion guaranteed by . . . the European Convention on Human Rights. The main limitation . . . is that the religion must have a clear structure and belief system.” The Explanatory Notes observe that all the main religious groups are covered, including the Baha’i faith, Buddhism, Christianity, Hinduism, Islam, Jainism, Judaism, Rastafarianism, Sikhism, and Zoroastrianism, as well as denominations or sects within a religion, such as Protestants and Catholics within Christianity. The domestic courts have held that a belief that homosexual activity is “sinful,” a belief that

2013) for a full discussion.

84 See Equality Act, 2010, c. 15, pt. 2, § 4 (UK). The Equality Act also protects against discrimination on other grounds. Id.


87 Id. cmt. 51.

88 Id. cmts. 51–53.

children should not be placed for adoption with same-sex couples, and a belief that marriage is the union of one man and one woman for life (and as such enabling same-sex unions to be formed is contrary to God’s instructions), are all protected beliefs when they form part of a broader Christian faith.

The concept of discrimination adopted by the Act, however, is narrower than under the ECHR and some EU instruments. “Discrimination” for the purposes of the Act is defined in largely formalistic terms, and the Act proscribes such discrimination only in certain closely defined, albeit wide, circumstances, such as in the provision of services, the exercising of public functions, and in employment. The Act contains a number of exemptions applying to gender, gender reassignment and sexual orientation. It is by this means that the legislature has identified how in certain circumstances conflicts between the rights of different protected groups are to be resolved. As they apply to actions motivated by religious belief, these exemptions are very controversial and are seen by some as giving special privileges to religious individuals and organizations. Addressing potential conflicts through specific and closely circumscribed exemptions does have the virtue of certainty. However, as is discussed under Sections III and IV below, that certainty comes at the expense of full equality for women and sexual minorities whose rights are sometimes subordinated to claims to religious freedom.

City Council, [2011] EWHC 375, [6] (Admin) (a belief by members of the Pentecostal Church that sexual relations other than those within marriage between one man and one woman were morally wrong and a belief, therefore, that homosexuality was “against God’s laws and morals”).


93 Id. pts. 3, 5 (UK).

94 The Court has a role in other circumstances. See id. § 19. In particular, where a complaint is made of indirect discrimination, it is left to the courts to determine whether there is justification for any prima facie discrimination. Id.

95 Id. at sch. 9, para 2; sch. 3, para. 29; sch. 22, para. 3; sch. 23, para. 2.
III. PROTECTING THE RIGHTS OF MINORITIES AGAINST DISCRIMINATORY ACTS MOTIVATED BY RELIGIOUS BELIEF

There are, then, four distinct legal schemes addressing equality and nondiscrimination in the United Kingdom. These schemes provide for fairly comprehensive, albeit not always coherent, protections. They also do not stand in isolation since each informs the other. The meaning to be afforded the various expressions under the Equality Act 2010 will so far as possible conform to EHCR law because the HRA requires as much. In addition, a considerable amount of EU law is directly effective in the United Kingdom even without transposing legislation, and in any event may require that existing domestic legislation be construed compatibly with it. The EU is about to become a member of the Council of Europe and will accede to the ECHR. Further, the EU is bound to act in accordance with the ECHR, the contents of which now comprise general principles of EU law. The Court of Justice of the European Union and the ECtHR have anyway long since taken account of each other’s jurisprudence in formulating their own case law.

Taken together these legal schemes ensure that religious freedoms are robustly protected in the United Kingdom. But these legal instruments also protect against gender and sexual orientation discrimination. Refusing to provide employment or services to a woman, a gay man, a lesbian, or a transgendered person because of their status as such, is made unlawful under the ECHR, EU law, and the Equality Act 2010. This creates

97 European Communities Act, 1972, c. 68, § 2 (UK).
98 The system of legal rules that flow from the UK’s accession to what was then the European Economic Community (“EEC”), now the EU, is complex but the impact of EU law on domestic law largely derives from Section 2 of the European Communities Act. Id.
99 TEU, supra note 76, art. 6(2).
101 ECHR, supra note 25, arts. 8, 14.
friction since the expression of religious belief may well impair the enjoyment of the equality rights of women, gay men, lesbians, or transgendered persons.

As domestic case law has made clear, “religious conviction is not a solvent of legal obligation.” This does not mean, however, that freedom of religion must always give way to competing rights; this is certainly not the case. There are numerous examples where religious beliefs are accommodated even where that impinges on the rights of women and sexual minorities. This is sometimes explicitly provided for in legislation, and at other times it occurs through the interpretation (or perhaps stretching) of legislation in such a way as to accommodate religious belief. This allows those motivated by religious beliefs “to be true to their beliefs while remaining respectful of the law.”

Examples abound: The Abortion Act 1967 makes abortion lawful in certain circumstances and at the same time excuses a person from “participat[ing] in any treatment” authorized by the 1967 Act to which they have a conscientious objection. The recent case of \textit{Doogan v. NHS Greater Glasgow & Clyde Health Board} broadly interprets the conscientious objection clause. According to the court in \textit{Doogan}, the clause extends not only to the actual medical or surgical termination but to the “whole process of treatment” given for that purpose. In \textit{Doogan}, two Catholic midwives succeeded in their claim that in addition to refusing to participate in the conducting of an abortion, they were

\begin{footnotesize}
\footnotesize
  \begin{itemize}
    \item 103 Equality Act, 2010, c. 15, §§ 4, 29, 39 (UK).
    \item 104 R (Williamson) v. Secretary of State for Educ. & Emp’t, [2005] 2 A.C. 246 [58] (citing Church of the New Faith v. Comr. of Pay-Roll Tax, 154 CLR 120 [136] (1983)).
    \item 105 Doogan v. NHS Greater Glasgow & Clyde Health Bd., [2013] CSIH 36 [37].
    \item 106 Abortion Act, 1967, c. 87, § 4 (U.K.).
    \item 107 \textit{Doogan}, [2013] CSIH 36.
    \item 108 \textit{Id.} at [37].
  \end{itemize}
\end{footnotesize}
entitled to refuse to carry out supervisory and management responsibilities in relation to staff assisting in abortions, and could refuse to participate in the provision of care to patients undergoing abortions at any stage in the process.\textsuperscript{109} There are serious practical consequences in permitting senior midwives to refuse to engage in any activity connected, however remotely, with the carrying out of abortions.\textsuperscript{110} Nevertheless, the court chose to interpret the 1967 Act in a way that allowed the Catholic midwives to “to be true to their beliefs while remaining respectful of the law.”\textsuperscript{111} It did this by giving a meaning to the conscientious objection clause that extended well beyond what the 1967 Act appeared to intend, and beyond what the professional nursing bodies had until then understood the limits of the clause to be. Unless the decision is overturned by the United Kingdom Supreme Court,\textsuperscript{112} there is a real possibility that the right to access a legal abortion will be impeded, with the foreseeable attendant risk to the health and well-being of women seeking to terminate a pregnancy.

Another such example is the Marriage (Same Sex Couples) Act 2013 (“Same Sex Marriage Act”) which made marriage between same-sex partners lawful, but also introduced provisions that prohibit same-sex marriages from being contracted on religious premises unless very tight conditions are met.\textsuperscript{113} The

\textsuperscript{109} Id. at [6].

\textsuperscript{110} Elizabeth Prochaska, Abortion and Conscientious Objection: What about Human Rights?, U.K. HUMAN RIGHTS BLOG (May 22, 2013), http://ukhumanrightsblog.com/2013/05/22/comment-abortion-and-contentious-objection-what-about-human-rights-elizabeth-prochaska/ (last visited Feb. 22, 2014). While section 4(2) of the Abortion Act does not enable a conscientious objection to be raised where the mother is in danger of grave permanent injury or death, as Elizabeth Prochaska points out: “[T]hat is a difficult assessment to make and a woman’s condition can deteriorate rapidly. Savita Halappanavar’s death shows how a system which ostensibly permitted abortion to save the mother’s life failed to protect her from the conscientious objection of her caregivers. The practical consequences of the judgment may put women at risk.” Id.

\textsuperscript{111} Doogan, [2013] CSIH 36 at [37].

\textsuperscript{112} An appeal is currently pending before the Supreme Court under case number UKSC 2013/0124.

\textsuperscript{113} Marriage (Same Sex Couples) Act, 2013, c. 30, § 1 (UK).
Same Sex Marriage Act allows for same-sex marriages on religious premises only where a religious organization has “opted-in,” in accordance with a prescribed procedure.\textsuperscript{114} Further, the Same Sex Marriage Act strictly prohibits a person being compelled by any means (including by the enforcement of a contract or a statutory or other legal requirement) to conduct a same-sex marriage according to religious rites. This is so even where those persons have the authority to conduct a marriage and are members of a religious organization that has “opted in.”\textsuperscript{115}

Also, as mentioned, the Equality Act 2010 contains exemptions privileging religion in certain cases. For example, it allows an employer to require that in order to be hired for a particular job, a person must be of a particular sex; must not be a transsexual person; must not be married or a civil partner, or must not be of a specified sexual orientation. This is permitted where (i) the employer can show that the employment is for the “purposes of an organised religion;” (ii) the application of the requirement engages “the compliance or non-conflict principle,” and (iii) the person to whom the requirement is applied does not meet it, or (save in relation to sex) the person applying the requirement has reasonable grounds for not being satisfied that the person meets it.\textsuperscript{116}

The “compliance principle” is engaged where a requirement

\textsuperscript{114} The Church of England and the Catholic Church have made clear that they will not apply to “opt in” as per the Marriage (Same Sex Couples) Act of 2013, § 3 (UK). \textit{Same-sex Marriage and the Church of England}, CHURCH OF ENG., http://www.churchofengland.org/our-views/marriage,-family-and-sexuality-issues/same-sex-marriage/same-sex-marriage-and-the-church-of-england-an-explanatory-note.aspx (last visited Feb. 22, 2014). The bar on civil partnerships being conducted on religious premises has been lifted for those religious groups who seek permission to have their premises approved for the registration of civil partnerships. \textit{See} Marriages and Civil Partnerships (Approved Premises) (Amend.) Regulations 2011, SI 2011/2661 (amending SI 2005/3168) (UK). So far, the Society of Friends (Quakers), Spiritualists, Unitarians, and the United Reformed Church have applied. The mainstream Christian churches (the Church of England and Catholic Church, specifically) have indicated that they will not apply, and protections are built into the legislation to ensure that they will not be required to do so.

\textsuperscript{115} Marriage (Same Sex Couples) Act of 2013, § 2 (UK).

\textsuperscript{116} Equality Act, 2010, c. 15, § 2(1)(a)--(b), sch. 9 (UK).
is applied so as to comply with the doctrines of the religion concerned. The “non-conflict principle” is engaged where, because of the nature or context of the employment, the requirement is applied so as to avoid conflicting with the strongly held religious convictions of a significant number of the religion’s followers.\(^\text{117}\) This exemption does not contain any threshold of proportionality,\(^\text{118}\) nor does it define what is meant by the expression “for the purposes of an organised religion.”\(^\text{119}\) The Government had intended introducing a proportionality requirement, and defining employment “for the purposes of an organised religion” as being employment wholly or mainly involving (a) leading or assisting in the observation of liturgical or ritualistic practices of a religion, or (b) promoting or explaining the doctrine of the religion (whether to followers of the religion or otherwise). The Catholic Bishops’ Conference of England and Wales expressed strong concern that defining employment for the purposes of an organized religion in that way “would unduly narrow the scope of the exception and limit the ‘essential’ ability of the Church in filling posts with a pastoral role ‘to prefer a candidate whose life is in accordance with its ethos.’”\(^\text{120}\) Due to the expression of such concerns, the definition of relevant employment for these purposes and the proportionality condition were removed by amendments made in the House of Lords during the passage of the Equality Bill (prior to its enactment as the Equality Act 2010).\(^\text{121}\) The Equality Act 2010

\(^{117}\) Id.

\(^{118}\) Unlike other exemptions, it is not necessary that any requirement be a proportionate means of achieving a legitimate aim. See, e.g., id. § 1(1), sch. 9 (ordinary occupational requirements).

\(^{119}\) See CLAYTON & TOMLINSON, supra note 48 (discussing the impact of Article 9 in this respect).

\(^{120}\) HOUSE OF LORDS & HOUSE OF COMMONS JOINT COMMITTEE ON HUMAN RIGHTS, LEGISLATIVE SCRUTINY: EQUALITY BILL, 26th Report, Sess. 2008–09 at para. 166 (citing Memorandum from the Catholic Bishops’ Conference of England and Wales (E14) to the Public Bills Committee).

\(^{121}\) Though this exemption is controversial, interference with the rights of sexual minorities in consequence of a similar exemption under earlier Regulations was found to be lawful. See, e.g., R (on the application of Amicus-MSF Section) v. Secretary of State for Trade & Indus., [2004] IRLR 430.
also exempts, in prescribed circumstances, sexual orientation discrimination in the provision of services by religious organizations, and it allows discrimination by religious ministers against women through the provision of services only to persons of one sex or separately for men and women. Wide exemptions also apply to faith-based schools (which are lawful in the United Kingdom as part of the State education system).

Where domestic statutory law prohibits discrimination against women or sexual minorities, then unless there is an exemption for those whose acts are motivated by religious belief, or there is a need to accommodate those beliefs under the ECHR, such discrimination will be unlawful whatever the motivation for it. This is illustrated by the case of Bull v. Hall. In Bull, the United Kingdom Supreme Court found that a couple operating a small hotel had acted unlawfully in refusing a double-bedded room to a same-sex couple in a civil partnership. This was because the law prohibits sexual orientation discrimination, as it does religious discrimination, in the provision of hotel accommodation and related services. This is without exception for private profit-making businesses run along religious lines. The fact that the discrimination was motivated by the hotel owners’ orthodox Christian belief “that the only divinely ordained sexual relationship is that between a man and a woman within bonds of matrimony,” did not rescue the hoteliers. As the Supreme Court held, finding that the hoteliers had acted unlawfully was simply to treat them equally to all other hoteliers. If the claimants in Bull, the same-sex couple seeking a room, ran a hotel and denied a double room to the defendants on the ground

122 Equality Act, 2010, c. 15, § 2, sch. 9 (UK).
123 Id. § 29, sch. 3.
125 Equality Act 2010, c. 15. § 5, sch. 11; see also id. § 11, sch. 3.
126 Bull v. Hall, [2013] UKSC 73 (appeal taken from Eng.).
127 The Justices gave different reasons for so holding. Compare id. at [24]–[30] (discussing the impact of the couple being in a civil partnership on the conclusion reached), with id. at [74]–[76] (Neuberger, L., concurring) (disagreeing as to the significance of a civil partnership).
128 Id. at [9].
of their Christian beliefs, they too would have been acting unlawfully. Neither group was privileged in law in this context. However, this does little to abate concerns about those cases where the law does prioritize religious belief over equality for women and sexual minorities.

IV. RELIGIOUS FREEDOM: A MINORITY RIGHT OR A MAJORITARIAN CLAIM?

Religion may not be the solvent of all legal obligation, but as the discussion above demonstrates, it can serve to avoid it, at least where women and sexual minorities are concerned. This is either because Parliament has enacted statutory exceptions applicable in cases where discriminatory acts are motivated by religious belief, or because the HRA (or the ECHR) will require that a particular religious belief be accommodated. Given the privileged space occupied by religion in the United Kingdom and, in the case of the Church of England, its legislature, this is perhaps of little surprise.

It is a peculiarity, however, that though the vast majority of the population of the United Kingdom self-identify as “Christian” (73.8%), the official law reports and mainstream media reports indicate that claimants in religious discrimination cases are overwhelmingly likely to be Christian where a “clash” with another’s nondiscrimination rights is engaged. This is most

---

129 See id. at [4], [54] (comments by Lady Hale).

130 This would not be tolerated if the discrimination were race-based. R (E) v. Governing Body of JFS, [2009] UKSC 15, [33]-[46] (appeal taken from Eng.). See Timeshev v. Russia, 2005-XII Eur. Ct. H.R. 169, 187 (holding that “no difference in treatment which is based exclusively or to a decisive extent on a person’s ethnic origin is capable of being objectively justified”). In the context of religion, this prospect is not fanciful. A number of predominantly U.S.-based “Christian” churches including the “Christian Knights of the Ku Klux Klan,” “Aryan Nations,” and a variety of other churches within the “Christian Identity” movement advocate white supremacism and anti-Semitism.

notably the case where the conflict concerns religion on the one hand, and gender or sexual orientation on the other.\textsuperscript{132} Christians can barely be said to comprise a minority group by any measurement, whether in actual numbers, or by distribution of power or privilege. This distinguishes this group of claimants from the usual claimant profile in discrimination claims. Women make up the vast majority of claimants in sex discrimination claims; gay men and lesbians in sexual orientation claims; and so on, for obvious historical and structural reasons. Women, gay men, and lesbians generally call on the law to remedy disadvantage experienced by them. However, certainly some Christian claimants pursuing religious discrimination claims might instead be said to be calling on the law to ensure that their privileged place in the public life of the United Kingdom is not displaced.

The dominance of Christianity in the United Kingdom's constitutional settlement, perhaps inevitably, obstructs the achievement of full equality for women and sexual minorities. Until the United Kingdom's formal Head of State and legislature discard their anachronistic ties to the Church of England, it is likely that those whose gender or sexual orientation are not accorded equal respect\textsuperscript{133} in the Church's theology and institutions of power will find their equality rights subordinated to the demands of Christianity. This is harmful to women and sexual minorities. However, the continuance of the status quo, though conflicting with all modern concepts of liberal, pluralistic democracy seems likely for some time. There is a political feebleness about tackling the Church of England and the place of Christianity in public life, and none of the mainstream parties have indicated any intention to do so. There are no present proposals to disestablish the Church of England or to otherwise


\textsuperscript{133} Equal respect must ultimately mean equal treatment.
reduce the influence of religion, in particular Christianity, in public life. This is a matter of considerable regret to many.
INCONVENIENCE OR INDIGNITY?
RELIGIOUS EXEMPTIONS TO PUBLIC
ACCOMMODATIONS LAWS

Marvin Lim* & Louise Melling†

Discrimination is not simply dollars and cents, hamburgers and movies; it is the humiliation, frustration, and embarrassment that a person must surely feel when he is told that he is unacceptable as a member of the public . . . .¹

—Senate Commerce Committee Report on the Civil Rights Act of 1964

To be sure, gays and lesbians also suffer serious economic injustices . . . . But far from being rooted directly in the economic structure, these injustices derive instead from the status order, as the institutionalization of heterosexist cultural norms produces a class of devalued persons who suffer economic liabilities as a byproduct. The remedy for the injustice, consequently, is recognition, not redistribution.²

—Nancy Fraser, The Tanner Lectures

* Marvin Lim is a Peter and Patricia Gruber Fellow at the American Civil Liberties Union.
† Louise Melling is the Deputy Legal Director of the American Civil Liberties Union and Director of its Center for Liberty.
INTRODUCTION

States are increasingly recognizing the right of LGBT individuals to live free from discrimination. Across the country, more and more state laws are prohibiting discrimination in public accommodations based on sexual orientation or gender identity. Confronted with these laws, some businesses have refused to comply, invoking the owners’ religious objections. As a result, inns, cake shops, and florists are closing their doors to customers because of these customers’ sexual orientation.

Invariably, these refusals to provide service come at the expense of the dignity of LGBT individuals. This harm is clear from the words of those refused service because of their sexual orientation. For example, “[i]t is hurtful to see that we are less welcome than the family dog,” stated a lesbian couple refused a room at a Vermont inn. Another gay couple emphasized the “shock and hurt” they experienced after being turned away by a florist in Washington State. “I was devastated . . . . I was crying,” explained a lesbian in New Jersey as she described the aftermath of being sent out of a bridal shop. “I can’t tell you

---


4 E.g., Katie McDonough, Oregon Baker Denies Lesbian Couple a Wedding Cake, SALON (Feb. 4, 2013), http://www.salon.com/2013/02/04/oregon_baker_denies_lesbian_couple_a_wedding_cake.


6 See also Fraser, supra note 2, at 14 (discussing the “serious economic injustices” suffered by gays and lesbians).

7 Dorning, supra note 3.

8 Valdes, supra note 5.

how much it hurt to be essentially told, ‘we don’t do business with your kind of people,’” said a woman who, along with her long-term girlfriend, was denied accommodations at a hotel in Hawaii.10 “We don’t want anyone else to experience that and [be] made to feel like they have no place in society,” she continued.11 “It still stings to this day.”12

As suggested in these statements and in the Senate Commerce Committee Report on the Civil Rights Act of 1964 quoted earlier, discrimination harms a person’s dignity. Yet this harm has been given little voice in the debates over religious exemptions to laws prohibiting discrimination in public accommodations. Proponents of exemptions have typically framed religious objectors’ compliance with LGBT antidiscrimination laws as pitting one person’s religious conscience against another person’s mere inconvenience and mild sense of offense.

This Article does not question the harm a person experiences when required to comply with a law that conflicts with his or her religious beliefs—that harm, whether or not it is legally cognizable, is real. Rather, this Article aims to shed light on what has been less articulated and appreciated: the dignitary harm that results when businesses turn away LGBT individuals based on the owners’ religious beliefs. Part I discusses how some proponents of religious exemptions, understating or overlooking the deeper harm at stake, frame the debate as one of religious conscience versus customer inconvenience. Part II shows how the U.S. Supreme Court has long recognized the dignitary harm inflicted by discrimination and the critical role antidiscrimination laws play in preventing that harm. Part III illustrates that, in the transnational debate, the courts of other countries have repeatedly recognized the dignitary harm of discrimination against LGBT people, even in the face of competing religious liberty claims. Finally, Part IV argues that, as American courts and legislatures now consider the scope of protections for LGBT people in

9900898_1_bridal-shop-dresses-gay-marriage.
11 Id.
12 Id.
antidiscrimination laws, they must give weight to the harm to dignity as they have in other contexts. Accordingly, they should reject calls for religious exemptions to public accommodations laws that protect LGBT people.

I. REFUSAL OF SERVICE: A MERE INCONVENIENCE?

Proponents of religious exemptions have argued that refusals by businesses to serve LGBT people cause little harm if the individual can obtain the services elsewhere. Accordingly, they frame the issue of compliance with laws prohibiting sexual orientation-based discrimination as pitting customers’ inconvenience against a much deeper harm to business owners: a burden to their religious beliefs. Among the most prominent and representative of these proponents are Professors Doug Laycock and Thomas Berg. In the book, Same-Sex Marriage and Religious Liberty, Laycock characterizes the harm generated by not allowing for religious exemptions as “forcing the merchant to violate a deeply held moral obligation.” 13 Similarly, Berg emphasizes that the harm to religious merchants would cut to the core of their being, since religious beliefs “affect virtually all of the defining decisions of personhood.” 14

The harm to those turned away from businesses is far less, according to Laycock and Berg. They characterize this harm as “the insult of being refused service and the inconvenience of going elsewhere.” 15 It should be unsurprising, then, that they would deny religious exemptions “only in cases of concrete, tangible hardship,” where the customer will struggle to secure similar services elsewhere. 16 But in “the large majority of cases,” Berg argues, 17 there will be no such hardship, considering particularly that “the large majority of gay couples” live in urban

13 Douglas Laycock, Afterword to Same-Sex Marriage and Religious Liberty: Emerging Conflicts 189, 197 (Douglas Laycock et al. eds., 2008).
15 Laycock, supra note 13, at 197.
16 Berg, supra note 14, at 229; see also Laycock, supra note 13, at 198.
17 Berg, supra note 14, at 229.
areas where presumptively there are more vendors. Laycock even goes so far as to say that he “would have no objection to a requirement that merchants that refuse to serve same-sex couples announce that fact on their website or, for businesses with only a local service area, on a sign outside their premises.” Such signals, Laycock argues, would avoid “unfair surprise,” which presumably exacerbates any inconvenience that an LGBT person experiences.

Laycock and Berg do concede imperfections with their approach, including the idea of harm beyond the potential transaction costs. Laycock argues:

From the gay rights perspective, discrimination gets a certain legitimacy, and in the worst case, the stream of commerce might be sprinkled with public notices of discriminatory intent. In more traditional communities, same-sex couples planning a wedding might be forced to pick their merchants carefully, like black families driving across the South half a century ago. All of this is true, and in some parts of the country it would be very real . . .

In the end, though, Laycock finds these concerns to be insufficiently alarming, arguing that “in most cities, such problems would be minimal.” Berg reaches the same conclusion:

Denials of service do affect gay couples by causing them disturbance, hurt, and offense. While acknowledging that harm, one must also acknowledge, I think, that the harm to the objector from legal sanctions is greater and more concrete. In most cases, the offended couple can go to the next entry in the phone book or the Google result. The individual or organization held liable for

---

18 Id. at 233.
19 Laycock, supra note 13, at 198.
20 Id.
21 Id. at 200.
22 Id.
discrimination, by contrast, must either violate the tenets of her (its) faith or else exit the social service, profession, or livelihood in which she (it) has invested time, effort, and money. One simply has not given the religious dissenter’s interest significant weight if one finds that offense or disturbance from messages of disapproval [is] sufficient to override it.  

As Laycock and Berg have jointly stated, denying religious exemptions threatens serious harm to a religious minority while conferring no real benefits to same-sex couples. Same-sex couples will rarely if ever actually want such personalized services from providers who fundamentally disapprove of their relationship, and they will nearly always be able to readily obtain these services from others who are happy to serve them.  

The approach advocated by Laycock, Berg, and others who adhere to similar arguments has faced no shortage of critique. Much of this critique, however, focuses on the inaccuracy of their economic arguments, most often disputing the contention that LGBT people turned away will usually have no trouble finding alternate service providers. This Article takes a different

23 Berg, supra note 14, at 229.


25 See, e.g., Michael Kent Curtis, A Unique Religious Exemption from Antidiscrimination Laws in the Case of Gays? Putting the Call for Exemptions for Those Who Discriminate Against Married or Marrying Gays in Context,
tack: even in the absence of economic transaction costs, discrimination against LGBT people motivated by sincere religious objection should not be permitted, because of the significant harm to dignity that it inflicts.

II. AT THE U.S. SUPREME COURT: THE LONGSTANDING RECOGNITION OF THE DIGNITARY HARM OF DISCRIMINATION

Dignity can be defined in various ways. But one conceptualization of dignity clearly recognized within American constitutional law—and the conception of dignity that is most relevant to the current debate over protecting LGBT people from discrimination—is the dignity of public respect and recognition. As Professor Neomi Rao states, this conception of dignity “requires more than . . . equal benefits in order to recognize belonging.” Instead, “[i]nherent in this conception of dignity is

47 WAKE FOREST L. REV. 173, 199–200 (2012) (arguing that it is hard to show, one way or the other, whether LGBT people actually have sufficient access to alternate service providers, as this would require difficult judgments about whether alternative providers are qualitatively comparable to the original provider, whether they are sufficiently close, and other such questions); Ira C. Lupu & Robert W. Tuttle, Same-Sex Family Equality and Religious Freedom, 5 NW. J.L. & SOC. POL’Y 274, 290 (2010) (criticizing Laycock and Berg’s argument insofar that it fails to address the imbalance between customer and merchant and fails to address the extra cost of “locating providers willing to serve [same-sex couples]”). One author who discusses the dignitary implications for Laycock’s approach is Shannon Gilreath, who argues that reducing the harm experienced by LGBT people to inconvenience is made possible by looking at the individual harm alone. See Shannon Gilreath, Not a Moral Issue: Same-Sex Marriage and Religious Liberty, 2010 U. ILL. L. REV. 205, 219 (2010) (book review) (arguing that a focus on “‘dignitary rights’ . . . rhetorizes the harm inherent in the proposed system as an individual harm” of embarrassment, insult, or inconvenience, which “can then easily be balanced against the individual rights of religious objectors”).

the idea that public respect and recognition are necessary to lead a full private life. An individual’s private choices gain meaning and validation in part through their recognition by the social and political community.”

The U.S. Supreme Court has long recognized this particular conception of dignity in its antidiscrimination jurisprudence. Beginning with Brown v. Board of Education, the Court has consistently understood the harm to dignity that discrimination causes, and recognized it to be distinct from the more “tangible” harm of being unable to access a particular benefit or entitlement.

Most relevant to the current debate over discrimination against LGBT people, the Court has, in the context of race, repeatedly recognized the dignitary harm of being turned away from public accommodations. The most prominent instance is in Heart of Atlanta Motel v. United States, where the Court upheld the constitutionality of the Civil Rights Act’s prohibition of discrimination in public accommodations, known as Title II. Writing the majority opinion, Justice Clark affirmed that “the fundamental object of Title II was to vindicate ‘the deprivation of personal dignity.’” Here Clark quoted from the Senate Commerce Committee’s report on the bill, which states:

The primary purpose . . . is to solve this problem, the deprivation of personal dignity that surely accompanies denials of equal access to public establishments. Discrimination is not simply dollars and cents, hamburgers and movies; it is the

27 Rao, supra note 26, at 262.
28 See Brown v. Bd. of Educ., 347 U.S. 483, 494 (1954) (“To separate [black children] from others of similar age and qualifications solely because of their race generates a feeling of inferiority as to their status in the community that may affect their hearts and minds in a way unlikely ever to be undone.”); see also Batson v. Kentucky, 476 U.S. 79 (1986) (holding that a state’s discriminatory selection of jurors violates a defendant’s Equal Protection rights, without regard to whether the absence of such discrimination would have changed the outcome of the jury’s decision).
30 Id. at 250 (citing S. REP. NO. 88–872, at 2370 (1964)).
humiliation, frustration, and embarrassment that a person must surely feel when he is told that he is unacceptable as a member of the public because of his race or color. It is equally the inability to explain to a child that, regardless of education, civility, courtesy, and morality, he will be denied the right to enjoy equal treatment even though he be a citizen of the United States and may well be called upon to lay down his life to assure this Nation continues.31

The Civil Rights Act’s legislative history makes clear that, while Title II is concerned with remedying the “adverse economic effect of discrimination,” the “fundamental purpose . . . is directed at meeting a problem of human dignity”32 This is shown as well in the report’s citation of Roy Wilkins, then-executive secretary of the NAACP: “The truth is that the affronts and denials that this section, if enacted, would correct are intensely human and personal. Very often they harm the physical body, but always they strike at the root of the human spirit, at the very core of human dignity.”33

Though perhaps most prominent, Heart of Atlanta is only one instance in which the Court has given recognition and weight to harm to dignity. Indeed, the Court has recognized this harm across many different contexts where discrimination occurs. In Roberts v. Jaycees, the Court held that discrimination—in that case, turning women away from a private organization—“deprives persons of their individual dignity and denies society the benefits of wide participation in political, economic, and cultural life.”34 In doing so, the Court found that the state’s public accommodations law served a compelling interest, one outweighing the First Amendment right to freedom of association

32 Id. at 2371 (emphasis added).
33 Id. at 2369.
34 Roberts v. U.S. Jaycees, 468 U.S. 609, 625 (1984) (rejecting the all-male club Jaycees’ freedom of association claim because the organization lacked sufficient intimacy in size and selectivity, and because the state had a compelling interest in eradicating gender discrimination).
in this particular case.\textsuperscript{35} In \textit{JEB v. Alabama ex rel. T.B.}, when striking down gender-based preemptory challenges in jury selection, the Court stated that such discrimination can be an “assertion of . . . inferiority”\textsuperscript{36} that “denigrates the dignity of the excluded” and “reinvokes a history of exclusion.”\textsuperscript{37} In \textit{Curtis v. Loether}, a case arising out of a Fair Housing Act racial discrimination claim, the Court stated that “[a]n action to redress racial discrimination may also be likened to an action for defamation or intentional infliction of mental distress . . . . [U]nder the logic of the common law development of a law of insult and indignity, racial discrimination might be treated as a dignitary tort.”\textsuperscript{38} And in the employment context, the Court explained in \textit{Price Waterhouse v. Hopkins} that “[w]hile the main concern of [Title VII] was with employment opportunity, Congress was certainly not blind to the stigmatic harm which comes from being evaluated by a process which treats one as an inferior by reason of one’s race or sex.”\textsuperscript{39} The Court continued, “whatever the final outcome of a decisional process, the inclusion of race or sex as a consideration within it harms both society and the individual.”\textsuperscript{40}

\textsuperscript{35} Id. at 625–26.

\textsuperscript{36} J.E.B. v. Alabama \textit{ex rel.} T. B., 511 U.S. 127, 142 (1994) (quoting Strauder v. West Virginia, 100 U.S. 303, 308 (1880)).

\textsuperscript{37} Id.; see also Camille Gear Rich, \textit{What Dignity Demands: The Challenges of Creating Sexual Harassment Protections for Prisons and Other Nonworkplace Settings}, 83 S. CAL. L. REV. 1, 56 (2009) (discussing both the “public” dignity concerns touched on by the Court’s “branding” argument, and the “private” dignity concerns touched on by the Court’s “subordination” argument).


\textsuperscript{40} Price Waterhouse, 490 U.S. at 265. It is worth noting that the Court has also recognized that such harm occurs beyond race- or gender-based discrimination. See, e.g., Commissioner v. Schleier, 515 U.S. 323 (1995). In \textit{Schleier}, the Court recognized that age discrimination may cause economic loss and a separate “psychological or ‘personal’ injury”—and that remedying
The Supreme Court has also recognized the dignity at stake in sexual orientation discrimination. In *United States v. Windsor*, the Court emphasized how a state’s decision to give LGBT people the right to marry “conferred upon them a dignity and status of immense import.”\(^{41}\) Consequently, the Defense of Marriage Act effectuated not just a denial of the economic benefits tied to marriage but also a “differentiation [that] demeans the couple, whose moral and sexual choices the Constitution protects and whose relationship the State has sought to dignify.”\(^{42}\) Hearkening to the language of Title II’s legislative history,\(^{43}\) the Court also recognized the problem that discrimination presents for children. According to the Court, DOMA “humiliates tens of thousands of children now being raised by same-sex couples. The law in question makes it even more difficult for the children to understand the integrity and closeness of their own family and its concord with other families in their community and in their daily lives.”\(^{44}\)

In short, American jurisprudence amply recognizes the harm to dignity resulting from discrimination. It takes more seriously than do Laycock and Berg the harm of being turned away. The harm goes to a person’s core, to her dignity. The question then remains: how does this harm to dignity factor into the issue of today—namely, whether the law should accord exemptions to businesses that object, on religious grounds, to complying with laws that prohibit discrimination against LGBT people in public accommodations? Put differently, how does this harm to dignity weigh against the harm to business owners, who today object to serving LGBT people on religious grounds?

III. DIGNITY IN THE TRANSNATIONAL CONTEXT

Courts and legislative bodies throughout the United States are one does not necessarily remedy the other. *Id.* at 330 (explaining that with respect to the remedy for loss of wages and the remedy for “personal” injury, “neither is linked to the other”).

\(^{41}\) United States v. Windsor, 133 S. Ct. 2675, 2692 (2013).

\(^{42}\) Id. at 2694 (citation omitted).


\(^{44}\) Windsor, 133 S. Ct. at 2694.
grappling with these very questions today. As they do, it is critical to remember, as this Symposium Issue shows, that debates about the intersection of religious freedom and equal treatment are not taking place in just one country. Instead, such debates are happening across multiple continents. Only recently has this issue begun to percolate in the American courts in the context of LGBT rights. However, the courts of other countries have already confronted claims for religious exemptions with respect to LGBT antidiscrimination laws—and they have repeatedly rejected such claims. In the process, they have also repeatedly recognized the central thesis of this Article: that preventing the dignitary harm of discrimination is a paramount interest.

The European Court of Human Rights,45 and courts in the United Kingdom,46 Israel,47 and South Africa have all rejected the notion that violation of antidiscrimination laws could be sanctioned in the name of religion.48 In addition, the French Constitutional Court rebuffed a claim that its national marriage equality law was constitutionally defective because it did not have a religious exemptions provision.49 It is beyond the scope of this

45 See Eweida et al. v. United Kingdom, 57 Eur. Ct. H.R. 213, 239 (2013) (upholding British court decision denying a British civil servant’s religious discrimination claim, arising out of the government’s requiring that she register same-sex civil partnerships, in spite of her religious objections to doing so); id. at 215 (upholding British court decision denying religious discrimination claim by a psychosexual therapist, who was dismissed by his employer after he refused to provide sex therapy to LGBT individuals).

46 See Bull v. Hall & Preddy, [2013] UKSC 73 (S.C.) (upholding a discrimination claim against the owners of a bed-and-breakfast who refused to serve a gay couple, on the grounds of their religious beliefs).

47 CS 5901/09 Tal Ya’akovovich and Yael Biran v. Yad Hashmona Guest House (2012) (Isr.) (Jerusalem Magistrate Court ruling that the owners of a reception hall violated Israeli antidiscrimination law by cancelling a reservation to host a wedding reception after discovering that the reception was for a lesbian couple).

48 See Strydom v. Nederduitse, 2009 (4) SA 510 (CC) at para. 6 (S. Afr.) (finding by South African court that Christian church violated antidiscrimination law when it fired a music teacher for being gay).

49 See Franck M. et al, Conseil constitutionnel [CC] [Constitutional Court] decision No. 2013-353 QPC, Oct. 18, 2013 (Fr.).
Article to engage in a comprehensive analysis of all of these countries and their jurisprudence. Instead, we briefly highlight one country whose courts have already repeatedly faced this issue: Canada.

Perhaps more frequently than the courts of any other country, Canadian courts have recognized the dignitary harm of discrimination against LGBT people, even in the face of competing religious liberty claims. In one such case, a challenge to a refusal by the Knights of Columbus to rent out a hall for a same-sex marriage reception, Smith & Chymyshyn v. Knights of Columbus et al., 2005 BCHRT 544 (Can.), the British Columbia Human Rights Tribunal awarded damages to compensate the plaintiffs “for injury to dignity, feelings and self-respect.”

Other Canadian decisions have spoken at greater length on the dignitary harm of discrimination against LGBT people. Deciding In the Matter of Marriage Commissioners Appointed Under the Marriage Act, the Saskatchewan Court of Appeal held that an amendment to Saskatchewan’s Marriage Act, which would have allowed individual marriage commissioners to refuse to perform same-sex marriages, violated the Canadian Constitution. In so holding, the court emphasized the harm of being turned away, a harm not mitigated simply by finding another commissioner to perform the marriage:

[T]his submission overlooks, or inappropriately discounts, the importance of the impact on gay or lesbian couples of being told by a marriage commissioner that he or she will not solemnize a same-sex union. As can be easily understood, such effects can be expected to be very significant and genuinely offensive. It is easy to imagine the personal hurt involved in a situation where an individual is told by a governmental officer “I won’t help you because you are black (or Asian or [Native Canadian]) but someone else will.” [B]eing

---

50 Smith & Chymyshyn v. Knights of Columbus et al., 2005 BCHRT 544 (Can.).

51 Id. at para. 151.

52 Marriage Commissioners Appointed Under the Marriage Act, 2011 SKCA 3 (Can.).

53 Id. at paras. 2–3.
told “I won’t help you because you are gay/lesbian but someone else will” is no different.\textsuperscript{54}

The court also emphasized that the dignitary harm will hardly be isolated—and gives credence in the process to the “legitimation of discrimination” that proponents of accommodation in the U.S. often de-emphasize:

\textsuperscript{55} Important \ldots is the affront to dignity, and the perpetuation of social and political prejudice and negative stereo-typing that such refusals would cause. Furthermore, even if the risk of actual refusal were minimal, knowing that legislation would legitimize such discrimination is itself an affront to the dignity and worth of homosexual individuals. History has established and jurisprudence has confirmed the extreme vulnerability of this group to discrimination and even hatred.\textsuperscript{56}

To the Saskatchewan Court of Appeal, it thus hardly matters that this regime may inflict no economic transaction costs on LGBT individuals; what matters is that their dignity will be harmed regardless. The reasoning of these courts follows that of U.S. courts when the latter addresses the harm of discrimination in the context of both race and gender. However, the question remains: will the U.S. debates concerning LGBT discrimination follow this lead?

\footnotesize{\textsuperscript{54} Id. at para. 41.  
\textsuperscript{55} See supra notes 21-22 and accompanying text.  
\textsuperscript{56} Marriage Commissioners, 2011 SKCA 3, para. 107. In another case, the Ontario Superior Court affirmed the claim of a student who wanted to bring a same-sex date to his prom at his Catholic high school. Hall v. Durham Catholic Dist. Sch. Bd. [2002] O.J. No. 1803 (Can.). The court emphasized “the impact of stigmatization on gay men in terms of denial of self, personal rejection discrimination and exposure to violence.” Id. at para. 53. Being barred from bringing a same-sex date to a culturally significant event like a prom, the court argued, is a “harm that cannot be properly compensated in damages.” Id. at para. 51.}
IV. GIVING DIGNITY ITS DUE IN THE CURRENT DEBATES OVER LAWS PROHIBITING LGBT DISCRIMINATION IN PUBLIC ACCOMMODATIONS

We are now at a critical moment in the United States in the debate over the propriety of religious exemptions to laws ending discrimination against people based on sexual orientation and gender identity. The question of exemptions arises whenever legislatures consider enacting protections against discrimination;\(^57\) where these protections already exist, but businesses claim a right of religious exemption;\(^58\) and, as in Arizona this year, where legislatures call for religious freedom protections that would make it much easier for objectors to secure exemptions.\(^59\) The


acknowledgement of dignity is thus critical because, at its core, the question in these contestations is whether there is a governmental interest in prohibiting the discrimination of sufficient strength to override any harm to the business owner.

In other contexts, we have already rejected the notion of exemptions to antidiscrimination measures predicated on religious beliefs. The Civil Rights Act of 1964, which prohibits discrimination based on race, among other predicates, has no exception for those who object to racial integration in public accommodations based on religious grounds.\(^6^0\) And the courts have also rejected claims for exemptions to integration of the races based on religious grounds.\(^6^1\) The question then becomes, is there any basis to reason differently here?

There is no question that a business owner experiences an affront when she is required to comply with the law in spite of her sincere religious objection to doing so. The harm described by business owners arises from their role as an agent in facilitating what, to them, is a moral wrong. The issue for the merchant is participation; it is irrelevant to them whether these individuals will likely find another service provider. To those seeking services, harm arises from the denial of agency, whether or not they could easily obtain the same services elsewhere. This is because a person refused help in this manner is essentially “told that [he or she] is unacceptable as a member of the public,”


eliciting “humiliation, frustration, and embarrassment,” as the Senate Commerce Committee stated in discussing Title II. 62 This harm is unlike that of a business turning away a customer merely for lack of appropriate attire, as this harm is set against a history of discrimination. 63

This brand of harm can only be addressed, as philosopher Nancy Fraser has stated, by public “recognition, not redistribution.” 64 Antidiscrimination laws provide a form of this recognition: by declaring a group to have a right to access goods and services, for example, the political community takes an affirmative step to accord respect and recognition to that group. Exemptions to these laws undermine that communal respect and recognition. And they legitimize discrimination, even if only in small pockets of society, and thus undermine the traditionally stigmatized group’s belief that the community will ever give them a fair shake. 65

This harm would only be exacerbated, not mitigated, by Laycock’s suggestion that merchants could post notice that they will not serve LGBT people or couples, so as to avoid “inconveniencing” LGBT clients. Such a sign would only reinforce LGBT people’s feeling of exclusion—regardless of whether other proprietors were more welcoming. To be fair, Laycock does recognize that public notices of discriminatory intent are problematic: he observes that “[i]n more traditional communities, same-sex couples planning a wedding might be

63 See, e.g., United States v. Windsor, 133 S. Ct. 2675, 2693 (2013) (“The avowed purpose and practical effect of the [Defense of Marriage Act] are to impose a disadvantage, a separate status, and so a stigma upon all who enter into same-sex marriages made lawful by the unquestioned authority of the States.”); Romer v. Evans, 517 U.S. 620, 632 (1996) (holding that a state constitutional amendment removing all LGBT-specific public antidiscrimination protections “seems inexplicable by anything but animus”).
64 Fraser, supra note 2.
65 See R.A. Lenhardt, Understanding the Mark: Race, Stigma, and Equality in Context, 79 N.Y.U. L. REV. 803, 840 (2004) (noting that “racial stigma deprives individuals of the confidence that they are being dealt with in good faith, leaving them (quite understandably) somewhat mistrustful of even those individuals who expressly claim and perhaps even believe that they are [nondiscriminatory]”).
forced to pick their merchants carefully, like black families
driving across the South half a century ago.”66 But he argues that
this is the “worst case” scenario, and that “in most cities, such
problems would be minimal.”67 Even accepting the assertion that
“Heterosexuals Only” signs would not crop up across many
segments of the country, the question remains: why should this be
acceptable anywhere, even where motivated by religious belief?

It is also no response to say, as Laycock and Berg do, that we
can sanction or tolerate the exemptions because those turned away
by religious merchants should not want to be served by them
anyway.68 This essentially amounts to: “If they don’t want you,
why would you want them?” Like the argument emphasizing the
availability of welcoming proprietors, it is another way of saying,
what is the harm? Ironically, this argument sounds in dignity—
that individuals turned away because of who they are deserve
better. And it is an argument that, at its core, fails to consider
that grudging respect and recognition, even if not ideal, is still
better than no recognition—or rejection.69 One need only consider
the protests at Woolworth’s,70 early efforts at school integration
in Arkansas,71 protests at all-male clubs,72 or employment
discrimination lawsuits73 to appreciate the value of inclusion, even
if forced. Moreover, this argument fails to appreciate that legally
enforced recognition can in fact be the spark that changes minds
and institutions in the long run. As then-Solicitor General
Thurgood Marshall stated in 1966, “There is very little truth to

66 Laycock, supra note 13, at 200.
67 Id.
68 See supra note 24 and accompanying text.
69 This is particularly true when one considers the negative psychological
effects of discrimination. See, e.g., Vickie M. Mays & Susan D. Cochran,
Mental Health Correlates of Perceived Discrimination Among Lesbian, Gay,
and Bisexual Adults in the United States, 91 AMER. J. OF PUB. HEALTH 1869, 1869 (2001) (finding that the social stigma of homosexuality and the higher-
than-average rate of discrimination against LGBT individuals has important
mental health consequences).
70 See Michael J. Klarman, Brown, Racial Chance, and the Civil Rights
71 See id. at 47–48.
72 See supra notes 34–35 and accompanying text.
73 See supra notes 39–40 and accompanying text.
the old refrain that one cannot legislate equality. Laws not only provide concrete benefits, they can even change the hearts of men—some men anyhow—for good or evil.”

Finally, it is no response to propose that at least small businesses should be allowed to refuse service on religious grounds. Laycock and Berg have made this argument on the basis that “very small businesses... are essentially personal extensions of the individual owner.” But the dignitary harm is no less significant merely because the business that refuses the customer happens to be small. Notably, federal law banning discrimination in public accommodations has no such broad exemption. The question thus presents itself once more: where federal law banning discrimination based on race has no such exemption, why should laws prohibiting discrimination based on sexual orientation or gender identity have one?

In short, the question remains: once you acknowledge the harm to dignity that LGBT people experience when they are turned away, why should our laws and court decisions permit exemptions in this context when we have rejected them in other contexts? Why sanction “Heterosexuals Only” signs when we reject the notion of a restaurant posting, “Christians Only” or “Citizens Only”? Why permit this, even in the limited number of

---

74 Thurgood Marshall, Solicitor General, Address at the 1966 White House Conference on Civil Rights (June 1, 1966).


76 See 42 U.S.C. § 2000a(b), (e) (2006) (providing Title II exemptions only to inns “containing not more than five rooms for rent or hire and which [are] actually occupied by the proprietor of such establishment as his residence” and to any “private club or other establishment not in fact open to the public”).
instances when the refusal is premised on religious grounds? We see no compelling reason for a difference.

CONCLUSION

Going forward, American courts and legislatures should reaffirm the dignitary harm of discrimination. Accordingly, they should greet any calls for exemptions motivated by religious beliefs with great skepticism. In doing so, they will align themselves with longstanding American tradition and with the courts of many other countries. This tradition recognizes that, while the right to religious freedom is fundamental, religion cannot be used to discriminate, and thus to harm the dignity of people who deserve basic respect and recognition in our society.

Emphasizing precisely this point about respect, we conclude with the words of Justice Bosson in his concurrence in the case of *Elane Photography*, a case in which the New Mexico Supreme Court rejected a photography studio’s call for an exemption to that state’s antidiscrimination law, predicated on religious and speech grounds.\(^77\) Justice Bosson states:

> In the smaller, more focused world of the marketplace, of commerce, of public accommodation, the [company owners] have to channel their conduct, not their beliefs, so as to leave space for other Americans who believe something different. That compromise is part of the glue that holds us together as a nation, the tolerance that lubricates the varied moving parts of us as a people. That sense of respect we owe others, whether or not we believe as they do, illuminates this country, setting it apart from the discord that afflicts much of the rest of the world. In short, I would say to the [company owners], with the utmost respect: it is the price of citizenship.\(^78\)

---


\(^78\) *Id.* at 79 (Bosson, J., concurring).
WHO DECIDES CONSCIENCE?
RFRA’S CATCH-22

Priscilla J. Smith*

INTRODUCTION

This Article examines application of the Religious Freedom Restoration Act (“RFRA”) in cases challenging the contraception coverage rules under the Affordable Care Act.¹ I will discuss a problem with the application of RFRA’s statutorily mandated strict scrutiny test in this context that has not received attention—a problem I’ll call the RFRA Catch-22. The Court first

* Director, Program for the Study of Reproductive Justice, Information Society Project, Yale Law School. Thanks to the speakers and panelists at the symposium, Religious Freedom and Equality: An International Look, held at Brooklyn Law School in 2013. Thanks to Brooklyn Law School, its Dean, faculty and staff for hosting the symposium, to those whose generous support for the conference made it possible, and to Louise Melling, who was the primary organizer of the event. For comments on previous versions of this essay that have greatly improved it, thanks are due to Jack Balkin and the ISP fellows, especially Andrew Tutt and Kara Loewentheil. For sharing their thoughts about religion and a commitment to social justice, thanks are due to Robert M. Pennoyer, Rev. John F. Smith, and James Carroll. Finally, I am also extremely grateful to David Giller and Florence Mao for their excellent editorial guidance and suggestions.

¹ Religious Freedom Restoration Act (RFRA), 42 U.S.C. §§ 2000bb to 2000bb-4 (2012). In City of Boerne v. Flores, 521 U.S. 507 (1997), the Supreme Court held that the RFRA was unconstitutional as applied to state laws. It remains applicable to federal laws, like the Affordable Care Act. See also, e.g., O’Bryan v. Bureau of Prisons, 349 F.3d 399, 401 (7th Cir. 2003) (holding that RFRA is a valid exercise of congressional authority under the necessary and proper clause); Guam v. Guerrero, 290 F.3d 1210, 1221 (9th Cir. 2002); Kikumura v. Hurley, 242 F.3d 950, 958 (10th Cir. 2001); Henderson v. Kennedy, 265 F.3d 1072 (D.C. Cir. 2001); Christians v. Crystal Evangelical Free Church, 141 F.3d 854, 856 (8th Cir. 1998).
confronted the identical Catch-22 in Employment Division v. Smith, a case I discuss in detail below, when it attempted to apply strict scrutiny to the constitutional free exercise claims of Native Americans whose ceremonial peyote use was proscribed by state law. On the one hand, the Court recognized that the First Amendment prohibits judicial review of the “centrality” of conduct to an individual’s religion, the “relative merits of differing religious claims,” or “the determin[ation] of the place of a particular belief in a religion or the plausibility of a religious claim.” On the other hand, “[d]ispensing with a ‘centrality’ inquiry is utterly unworkable,” said the Court. It would require courts to grant all claims, and to equate burdens on throwing rice at church weddings to burdens on getting married in church, or, more relevant to today’s cases, the “burden” of having one’s employees covered by insurance that includes coverage for contraception with the “burden” of being forced to use contraception oneself. Faced with this Catch-22, this choice between an all or nothing approach to free exercise claims seeking accommodation from generally applicable nondiscriminatory laws, in Smith the Court chose nothing, and rejected application of the strict scrutiny test to claims under the Free Exercise Clause. The Court wrote, “[t]he government’s ability to enforce generally applicable prohibitions of socially harmful conduct, like its ability to carry out other aspects of public policy, ‘cannot depend on measuring the effects of a governmental action on a religious objector’s spiritual

---

3 Id. The plaintiffs in Smith were denied unemployment compensation benefits for “misconduct” when they violated the state’s drug laws by using peyote. Id.
4 Smith, 494 U.S. at 886–87 (“Nor is it possible to limit the impact of respondents’ proposal by requiring a ‘compelling state interest’ only when the conduct prohibited is ‘central’ to the individual’s religion. [A]s we reaffirmed only last Term, ‘[i]t is not within the judicial ken to question the centrality of particular beliefs or practices to a faith, or the validity of particular litigants’ interpretations of those creeds.’” (internal quotation marks and citations omitted)).
5 Id. at 887 n.4.
6 Id.
development.”

Under RFRA, however, courts are again required to apply the strict scrutiny test in challenges to federal government conduct claimed to burden “religious exercise,” requiring precisely the sort of judicial measurement of religious tenets and impact on spiritual matters that the Smith Court recognized are precluded by the Establishment Clause. Therefore, the U.S. Supreme Court will have to grapple with this familiar Catch-22 as it considers the expansive interpretations of “religious exercise” and “substantial burden” under RFRA promoted by the plaintiffs in Conestoga Wood Specialties Corp. v. Sebelius and Sebelius v. Hobby Lobby Stores, challenges to the contraceptive coverage rules being heard by the Court this term. As Georgetown Law Professor Marty Lederman’s detailed writings revealing the minimal burden on Hobby Lobby Executives’ religious exercise establish, in order to find for Hobby Lobby the Court would have to adopt a broad hands-off view of RFRA’s protections.

---

7 Id. at 885 (internal quotation marks omitted).
9 See infra Part II; Smith, 494 U.S. at 887 (equating evaluation of centrality with, inter alia, substantiality, discussing “unacceptable ‘business of evaluating the relative merits of differing religious claims,’” and citing Justice Stevens’ warning in United States v. Lee that this type of judicial evaluation of religious tenets would create “the risk that governmental approval of some and disapproval of others will be perceived as favoring one religion over another” (citing United States v. Lee, 455 U.S. 252, 263 n.2 (1982) (Stevens, J., concurring))); id. at 889 n.5, 887 n.4. See also Samuel J. Levine, Rethinking the Supreme Court’s Hands-Off Approach to Questions of Religious Practice and Belief, 25 FORDHAM URB. L.J. 85, 122–23 (1997) (noting that RFRA’s substantial burden test “appear[s] to require courts to engage in the kind of investigation into religious beliefs that Supreme Court Justices have increasingly and nearly uniformly rejected”).
12 See infra note 22.
Under this view, it is the RFRA claimant, not the court, who decides if something is a “substantial burden” on “religious exercise” under RFRA. This broad interpretation was articulated clearly by counsel for the University of Notre Dame in a recent oral argument in a related case in the Seventh Circuit in which counsel stated that it is enough if Notre Dame believes something is a “substantial burden” under RFRA. As counsel argued, “[i]t is up to the believer to draw the line.”

I won’t hide my views of these broad claims. Better to confess them now. If the Court upholds the plaintiffs’ RFRA claims and the broad hands-off interpretation of “religious exercise” and “substantial burden” they necessitate, rather than finding a way to limit RFRA’s scope constitutionally to deny accommodations in these cases, RFRA will have no boundaries. A broad RFRA, read as the Court must read it—and to read it fairly and in accordance with the Establishment Clause—will mean a vastly different society, but that’s not necessarily a bad thing. If I were confident that the courts would in fact review RFRA claims equitably, showing equal respect to all claims of religious exercise, I could allow myself to see a silver lining in a dark cloud—one that could be brought to bear in challenges to numerous federal laws under the aegis of federal RFRA. The broad interpretation of “religious exercise” and “substantial burden” being promoted in these cases could even be

---

13 See, e.g., Oral Argument at 27:23, Univ. of Notre Dame v. Sebelius, (7th Cir. argued Feb. 12, 2014) (No. 13-3853), available at http://media.ca7.uscourts.gov/sound/2014/rs.13-3853.13-3853_02_12_2014.mp3 [hereinafter Notre Dame Oral Argument]. This case involves the even more fantastical claim that even invoking a statutorily-granted accommodation from the contraceptive coverage requirements for non-profit religious institutions, who self-certify, was a “substantial burden.” See infra notes 109–10 and accompanying text.

14 I discuss possible narrowing techniques below. See infra Part III.

15 Smith, 494 U.S. at 888–89 (“The rule respondents favor would open the prospect of constitutionally required religious exemptions from civic obligations of almost every conceivable kind-ranging from compulsory military service, . . . to the payment of taxes, . . . [to] manslaughter and child neglect laws, . . . compulsory vaccination laws . . . drug laws, [to] environmental protection laws, . . . and laws providing for equality of opportunity for the races.”).
persuasive in challenges brought under state versions of RFRA that prohibit state restrictions that “substantially burden” “religious exercise.” Defined as broadly as the plaintiffs in these contraceptive coverage challenges advocate, requiring a hands-off judicial approach to evaluating burdens, RFRA’s protections could mean a new birth of freedom—freedom from draconian limits on reproductive choice, limits on sexual expression, limits on drug possession and drug use, requirements of service on juries, requirements that certain taxes be paid and census questions answered, and limitations on who and how many one may marry.

Unfortunately, though, I am not confident of the courts’ ability to apply a broad RFRA fairly. In rejecting strict scrutiny in *Smith*, the Court admitted that it cannot apply the unbounded strict scrutiny test equitably or in a manner in accordance with the Establishment Clause. Dueling opinions of two panels of the Seventh Circuit—one insisting on judicial evaluation of the “substantiality” of burden and the other limiting review of “substantiality” drastically—confirm this view.16

By reimposing the strict scrutiny test rejected in *Smith*, Congress has put the Court into the same untenable position it faced in *Smith*. The Court can choose “nothing” again, insisting that conducting these determinations is beyond the “judicial ken.” It could choose “all,” deferring to the plaintiffs’ characterization of religious “exercise” and the “substantiality” of burden, as Notre Dame’s counsel urged. Or it could, as I expect it to, claim to be evaluating the substantiality of the burden in this case but in practice conduct no real evaluation at all, ignoring its earlier warnings about the discriminatory results that have occurred under this standard and are likely to occur again in the future. Thus, the most likely result is a broad and protective RFRA for

16 The *Notre Dame* panel rejected the broad hands-off view, but another panel of the Seventh Circuit appears to endorse it. Compare *Notre Dame v. Sebelius*, No. 13-3853, slip op. at 21 (7th Cir. Feb. 21, 2014) (noting that “substantiality . . . is for the court to decide”), *with Korte v. Sebelius*, 735 F.3d 654, 683 (7th Cir. 2013) (noting that “substantial burden inquiry” must be limited to the evaluation of “the coercive effect of the government pressure” to act against beliefs). See also infra notes 109–19 and accompanying text.
some, those with religious exercise claims with which judges are most familiar, and a weak RFRA for the rest of us.

In Part I below, I will outline the relevant RFRA standards and ACA requirements, and briefly discuss arguments made by others that RFRA violates the Establishment Clause on its face, or alternatively, that RFRA would be unconstitutional “as applied” if applied to grant accommodations in *Hobby Lobby* and *Conestoga Wood*. In Part II, I will discuss judicial review of free exercise claims, explain the Catch-22 the Court faced in *Employment Division v. Smith*, how RFRA creates the same Catch-22, and how the breadth of the claims in *Hobby Lobby* and *Conestoga Wood* traps the Court in the Catch-22. Finally, in Part III, I will suggest two ways for the Court to limit RFRA and avoid the Catch-22 at least in these cases, and then close with a discussion of the ramifications of granting accommodations in these cases, either by granting all accommodations requested under RFRA or by conducting only a perfunctory examination of “substantial burden.”

I. THE RELIGIOUS FREEDOM RESTORATION ACT AND THE AFFORDABLE CARE ACT

The question of how to balance competing claims of religious conscience and equality mandates imposed by secular authority is currently being played out most prominently in the ongoing battle over the Affordable Care Act (“ACA”). The ACA requires employers with fifty or more employees, who are not otherwise exempt from the Act’s requirements, to provide their employees

---


18 The ACA provides broad exemptions for religious institutions and nonprofit organizations who self-certify that they oppose providing contraception. In such circumstances, health plans will provide the coverage without the involvement of the employer. See, e.g., 45 C.F.R. § 147.131 (2012).
WHO DECIDES CONSCIENCE? RFRA’S CATCH-22

with a minimum level of health insurance or pay an assessment to the Internal Revenue Service.\(^\text{19}\) Nonexempt group plans must provide coverage without cost-sharing for preventive care and screening for women\(^\text{20}\) that includes “[a]ll Food and Drug Administration approved contraceptive methods, sterilization procedures, and patient education and counseling for women with reproductive capacity.”\(^\text{21}\) Notably, employers do not have to pay for contraception themselves because contraception is a cost-saving preventive service and is therefore routinely offered at no additional cost. Moreover, as Professor Lederman has explained in detail, although this provision of the law has been widely described as a contraception “mandate,” this term is a misnomer, both because of the numerous accommodations and exemptions from the requirements granted by statute and regulation to religious institutions and nonprofit organizations, and because the statute also provides objectors with a way to avoid the contraception coverage requirements altogether.\(^\text{22}\) If these entities

\(^\text{19}\) Conestoga Wood, 724 F.3d at 381.

\(^\text{20}\) Id. (citing 42 U.S.C. § 300gg–13(a)(4) (2012)).

\(^\text{21}\) Id. (quoting 77 C.F.R. 8725 (2012)).

object to providing the health insurance package outlined in the Affordable Care Act, they can choose not to provide health insurance and instead pay an assessment to the IRS.\footnote{For a full description of the alternative to providing a plan with the required services and its implications, see Lederman, Hobby Lobby Part III-A, supra note 22.}

Despite this alternative, numerous cases have been filed throughout the country challenging the requirement in different postures.\footnote{For updates on these cases, see Challenges to the Federal Contraceptive Coverage Rule, ACLU, https://www.aclu.org/reproductivefreedom/challenges-federal-contraceptive-coverage-rule (last updated Mar. 13, 2014).} In March 2014, the Court heard arguments in two of those cases, \textit{Sebelius v. Hobby Lobby Stores,}\footnote{723 F.3d 1114 (10th Cir. 2013); see also Sebelius v. Hobby Lobby Stores, Inc., SCOTUSBLOG, http://www.scotusblog.com/case-files/cases/sebelius-v-hobby-lobby-stores-inc/ (last visited Mar. 25, 2014) (list of briefs filed for this case).} and \textit{Conestoga Wood Specialties Corp. v. Sebelius,}\footnote{724 F.3d 377 (3d Cir. 2013); see also Conestoga Wood Specialties Corp. v. Sebelius, SCOTUSBLOG, http://www.scotusblog.com/case-files/cases/conestoga-wood-specialties-corp-v-sebelius/ (last visited Mar. 25, 2014) (list of briefs filed for this case).} both of which involve objections of for-profit businesses to the contraceptive coverage requirement of the ACA. In \textit{Hobby Lobby}, the plaintiff is a for-profit corporation that claims the contraceptive coverage requirements violate the corporation’s right to religious exercise under RFRA because of the religious objections of the corporation’s owners.\footnote{Hobby Lobby Stores, Inc. v. Sebelius, 723 F.3d 1114 (10th Cir. 2013).} In \textit{Conestoga Wood}, the Plaintiffs include the individual “religious owners” of the “family business” as well as the for-profit corporation itself. They assert the claims of the individuals and the for-profit corporation under both RFRA,\footnote{42 U.S. C. §§ 2000bb to 2000bb-4 (2012).} as well as the Constitution’s Free Exercise Clause.\footnote{Conestoga Wood Specialties Corp. v. Sec’y of Health & Human Servs., 724 F.3d 377 (3d Cir. 2013).}

It is extremely unlikely that the Court will consider the free exercise claim in \textit{Conestoga Wood} because the standard applicable to constitutional free exercise claims is lower than the
RFRA standard.\textsuperscript{30} This means that if the plaintiffs prevail on their RFRA claims, there will be no need to look to the constitutional claim; and if the plaintiffs lose their RFRA claim under an application of RFRA’s strict scrutiny standard, it is “virtually inconceivable” that they’d win under the less stringent Free Exercise Clause claim, or even under a Free Exercise Clause reinterpreted to require application of strict scrutiny.\textsuperscript{31} Therefore, assuming the Court disregards, as have the lower courts,\textsuperscript{32} the plaintiffs’ option to avoid the requirement to provide health insurance that includes contraception by declining to offer any health insurance at all, the Court will be faced with difficult questions about how to evaluate the claims of conscience in these cases. While much has been written about RFRA generally\textsuperscript{33} and

\textsuperscript{30} In Employment Division v. Smith, 494 U.S. 872 (1990), the Court rejected the strict scrutiny standard of review for reasons discussed more fully below. See infra notes 37–40 and accompanying text.

\textsuperscript{31} See, e.g., Lederman, Hobby Lobby Part I, supra note 22 (predicting that “the constitutional question, as such, will consume only a tiny fraction of the total briefing, and virtually none of the Court’s attention.”).

\textsuperscript{32} See Lederman, Hobby Lobby Part III, supra note 22.

the contraceptive mandate challenges in particular, one issue that has not received attention is the impact that the Court’s analysis in *Smith* should have on its method of review of the contraceptive coverage challenges brought under RFRA and the potential Catch-22 that the RFRA revives.

A. *The Religious Freedom Restoration Act: A Response to Employment Division v. Smith*

RFRA was enacted in 1993 with broad bipartisan support in response to the U.S. Supreme Court’s decision in *Employment

---


34 I will not discuss here, but will refer the interested reader to, excellent literature analyzing claims that the contraceptive coverage requirements violate the Free Exercise Clause, the Free Speech Clause, or RFRA under current doctrine. For example, Caroline Corbin has written a series of articles outlining many flaws in the arguments of those who claim their free exercise rights are violated by the contraceptive coverage requirements. *See, e.g.*, Caroline M. Corbin, *The Contraception Mandate*, 107 NW. U. L. REV. 1469 (2013) (contraception coverage requirement does not violate Free Exercise Clause, the Free Speech Clause, or the Religious Freedom Restoration Act); *id.* at 1477 (citing Zelman v. Simmon-Harris, 536 U.S. 639 (2002) (availability of federal funds to religious schools through voucher programs was too indirect to create Establishment Clause problem)) (pointing out that any “burden” of providing health insurance that includes coverage for contraception is not “substantial” because it is so “indirect”); Corbin & Smith, Debate, *supra* note 33 (debating status of contraceptive coverage requirement under RFRA and arguing that corporations are not eligible “persons” under RFRA). For the argument that for-profit corporations are not “persons” under RFRA, see generally Corbin, *Corporate Religious Liberty, supra* note 33 (citing articles on the subject).

35 494 U.S. 872 (1990). *But see* Levine, *supra* note 9, at 122–23 (noting that RFRA’s substantial burden test “appear[s] to require courts to engage in the kind of investigation into religious beliefs that Supreme Court Justices have increasingly and nearly uniformly rejected”).
Division v. Smith. In Smith, the Court held that the Free Exercise Clause did not prohibit application of Oregon drug laws to the use of peyote during the religious ceremony of Native Americans and, therefore, the state could deny claimants unemployment compensation for work-related “misconduct” based on their use of the drug. In rejecting the free exercise claims of the Native Americans, the Court also rejected the strict scrutiny standard it had previously claimed was applicable to free exercise claims, writing:

To make an individual’s obligation to obey such a law contingent upon the law’s coincidence with his religious beliefs, except where the State’s interest is “compelling”—permitting him, by virtue of his beliefs, “to become a law unto himself”—contradicts both constitutional tradition and common sense.

As a result, the Court held that “if prohibiting the exercise of religion (or burdening the activity of printing) is not the object of the tax but merely the incidental effect of a generally applicable and otherwise valid provision, the First Amendment has not been offended.”

Reflecting a concern that the decision in Smith put religious exercise at risk, Congress enacted RFRA and reinstituted the strict scrutiny standard. RFRA requires that “[g]overnment shall

---

36 494 U.S. 872 (1990). For a description of the bipartisan movement to enact RFRA as a response to Smith, see, for example, Eisgruber & Sager, Why RFRA is Unconstitutional, supra note 33, at 438–41.

37 In Eisgruber & Sager, Why RFRA is Unconstitutional, supra note 33, at 446–47, the authors argue that in rejecting the strict scrutiny standard, Smith was actually just bringing doctrine in line with past results. While in other constitutional areas the compelling state interest test has been “‘strict’ in theory and fatal in fact,” in the pre-Smith religious exemption cases, they point out that the test was “strict in theory but feeble in fact.” Id. (noting the Court had only applied the test to mandate accommodations from generally applicable laws in the unemployment compensation cases and Yoder).

38 Smith, 494 U.S. at 885 (quoting Reynolds v. United States, 98 U.S. 145, 167 (1878)).

39 Id. at 878.

40 Eisgruber & Sager, Why RFRA is Unconstitutional, supra note 33, at 438 (describing “self-congratulatory hoopla” from both sides of the aisle that...
not substantially burden a person’s exercise of religion even if the burden results from a rule of general applicability,” unless “it demonstrates that application of the burden to the person—(1) is in furtherance of a compelling governmental interest; and (2) is the least restrictive means of furthering that compelling governmental interest.” But while some who supported RFRA may have been motivated by a reaction to the Court’s inconsistency in and seemingly discriminatory pattern with which the Court had applied the Free Exercise doctrine generally—and by the rejection of a “minority” religious claim in Smith in particular—enactment of RFRA simply reimposed the standard the courts had applied inconsistently in the past. Moreover, RFRA did nothing to solve the Catch-22 at the heart of the jurisprudence that led the Court to walk away from the strict scrutiny standard in Smith.

B. RFRA—Unconstitutional on its Face?

Law Professors Christopher Eisgruber and Lawrence Sager

accompanied enactment of RFRA).


42 Compare Wisconsin v. Yoder, 406 U.S. 205 (1972) (granting an accommodation from compulsory school-attendance laws to Amish parents who refused on religious grounds to send their children to school and discussing the Amish’s generally civilized behavior); Sherbert v. Verner, 374 U.S. 398 (1963) (granting Seventh Day Adventist an exemption from laws requiring her to make herself available to work on a Saturday), with Dep’t of Human Res. of Or. v. Smith, 494 U.S. 872 (1990) (rejecting application of the Sherbert test to peyote ban that prevented Native Americans from performing a religious ritual that was widely acknowledged to be central to their religious practice); O’Lone v. Estate of Shabazz, 482 U.S. 342 (1987) (sustaining, without mentioning the Sherbert test, a prison’s refusal to excuse Muslim inmates from work requirements to attend worship services); Goldman v. Weinberger, 475 U.S. 503 (1986) (rejecting application of the Sherbert test to military dress regulations that forbade the wearing of yarmulkes).

43 Professor Eisgruber was a Professor of Law at NYU Law School, and is currently the President of Princeton University. News at Princeton, PRINCETON UNIV. (Apr. 21, 2013), http://www.princeton.edu/main/news/archive/S36/65/54C75/index.xml?section=featured. Lawrence Sager is Professor of Law at University of Texas School of Law. See UT Law Faculty – Lawrence Sager, UT LAW, http://www.utexas.edu/law/faculty/sager/ (last
have argued that RFRA is unconstitutional on its face because it improperly privileges religion in violation of the Establishment Clause.\textsuperscript{44} This argument has been presented to the Court in an amicus brief filed on behalf of the Freedom from Religion Foundation and others.\textsuperscript{45} Justice Stevens adopted this view in his concurrence in \textit{City of Boerne v. Flores},\textsuperscript{46} where he argued that RFRA required granting an exemption from a generally applicable neutral civil law to religious practice, something that no atheist or agnostic could obtain, thus establishing a governmental preference for religion that is forbidden by the First Amendment.\textsuperscript{47} Other than Justice Stevens though, no other Justice was persuaded by the argument, or even commented on it, in \textit{Boerne}. Moreover, in \textit{Cutter v. Wilkinson},\textsuperscript{48} the Court explicitly rejected a facial Establishment Clause challenge to a law quite similar to RFRA, section 3 of the Religious Land Use and Institutionalized Persons Act of 2000 (“RLUIPA”). That statute mandates application of the RFRA strict scrutiny standard to patients or inmates confined to a federal institution.\textsuperscript{49} In \textit{Cutter},

\textsuperscript{44} See Eisgruber & Sager, \textit{Protecting Without Favoring}, supra note 33; Eisgruber & Sager, \textit{Why RFRA is Unconstitutional}, supra note 33.


\textsuperscript{46} 521 U.S. 507 (1997).

\textsuperscript{47} \textit{City of Boerne}, 521 U.S. at 537 (Stevens, J., concurring) (citing \textit{Wallace v. Jaffree}, 472 U.S. 38, 52–55 (1985)).

\textsuperscript{48} 544 U.S. 709 (2005).

\textsuperscript{49} \textit{Compare} 42 U.S.C. § 2000cc-1(a)(1)-(2) (2012) (providing in part: “[n]o government shall impose a substantial burden on the religious exercise of a person residing in or confined to an institution,” unless the burden furthers “a compelling governmental interest” and does so by “the least restrictive means”) \textit{with} 42 U.S.C. §§ 2000bb-1 (providing “(a) [g]overnment shall not substantially burden a person’s exercise of religion even if the burden results from a rule of general applicability [unless] . . . (b) . . . it demonstrates that application of the burden to the person—(1) is in furtherance of a compelling governmental interest; and (2) is the least restrictive means of furthering that compelling governmental interest”). \textit{See also} 42 U.S.C. § 2000cc (preventing implementation of land use regulation in a manner that “imposes a substantial burden on . . . religious exercise,” unless the government demonstrates that
state officials mounted a facial challenge to RLUIPA under the Establishment Clause after prisoners who were members of nontraditional religions claimed that their rights were violated under the Act. The Court held that while “[a]t some point, accommodation may devolve into ‘an unlawful fostering of religion,’” section 3 of RLUIPA did not cross this line.

There are certainly ways to distinguish Cutter and the RLUIPA from RFRA that could support a holding that RFRA violates the Establishment Clause in Hobby Lobby and Conestoga Wood, despite Cutter. First, and most obviously, RLUIPA is much “less sweeping” than RFRA. It targets two specific areas, land-use regulation and religious exercise by institutionalized persons, while RFRA is a seemingly unlimited mandate to privilege religious exercise over nonreligious conduct, and in Cutter, the Court emphasizes RLUIPA’s targeted nature. The Court made much of Congress’s extensive documentation in hearings spanning three years of the specific problems of institutionalized persons and the “frivolous or arbitrary,” “egregious and unnecessary” barriers to their religious exercise that they faced that could limit its holding to the narrow situation of institutionalized persons.

Second, the Cutter decision is quite narrow in other ways.

the burden “is in furtherance of a compelling governmental interest” and “is the least restrictive means” of furthering that interest).

Cutter, 544 U.S. at 714 (citing Corp. of Presiding Bishop of Church of Jesus Christ of Latter-day Saints v. Amos, 483 U.S. 327, 334–35 (1987)).

Id. The Sixth Circuit had agreed with the state officials, holding that RLUIPA violated the Establishment Clause on its face because it “impermissibly advances religion by giving greater protection to religious rights than to other constitutionally protected rights . . . .” Id. at 709 (discussing Cutter v. Wilkinson, 349 F.3d 257 (6th Cir. 2003)).

Id. at 715, 720–21 (“Section 3 covers state-run institutions—mental hospitals, prisons, and the like . . . .”); see also id. at 722 (citing appropriate accommodation of religion in military context where it did not interfere with “military duties”).

Id. at 716 (noting that “[b]efore enacting [section] 3, Congress documented, in hearings spanning three years, that . . . ‘some institutions restrict religious liberty in egregious and unnecessary ways’”) (quoting 146 Cong. Rec. 16698, 16699 (2000) (joint statement of Sen. Orrin Hatch and Sen. Edward M. Kennedy on RLUIPA)).
The Court has recognized that the Free Exercise Clause “requires governmental respect for, and noninterference with, the religious beliefs and practices of our Nation’s people.” In upholding RLUIPA, though, the Court treads carefully, navigating a narrow path between the “conflicting pressures” of the Free Exercise and the Establishment Clauses. As the Court explained in Hosanna

54 Id. at 719. See also Hosanna-Tabor Evangelical Lutheran Church & Sch. v. EEOC, 132 S.Ct. 694, 706 (2012) (noting that the First Amendment “gives special solicitude to the rights of religious organizations”). While some saw the Court’s recent decision in Hosanna—agreeing with Lutheran Church

55 Id. at 719–20 (citing Smith, 494 U.S. at 890; Corp. of Presiding Bishop of Church of Jesus Christ of Latter-day Saints v. Amos, 483 U.S. 327, 329–30 (1987) (approving federal exemption for religious organizations from Title
Tabor, the Religion Clauses must be interpreted in concert to both protect against government action that promotes the majority’s favored brand of religion (Establishment Clause) and government action that impedes religious practices not favored by the majority (Free Exercise Clause). On the one hand, if legislatures and judges were precluded by the Establishment Clause from adopting or granting, respectively, exemptions from generally applicable laws at least in some circumstances, then much of the protection the Free Exercise Clause is designed to provide—the “special solicitude” to religious practice it endorses—could be nullified by generally applicable laws that proscribe religious practices. The danger to religious exercise rights would be particularly acute in the institutions whose inhabitants RLUIPA was designed to protect. In those contexts, “government exerts a degree of control unparalleled in civilian society and severely disabling to private religious exercise.”

On the other hand, this ability to grant accommodations in honor of the Free Exercise guarantees has never been, nor should it be, unlimited, lest the accommodations tilt too far into Establishment Clause territory.

VII’s prohibition on religious discrimination) (noting that the Religion Clauses are “cast in absolute terms,” and “if expanded to a logical extreme, would tend to clash with each other”); Walz v. Tax Comm’n of N.Y.C., 397 U.S. 664, 668–69 (1970)).

Hosanna-Tabor, 132 S.Ct. at 730.

See id. at 706 (noting that the First Amendment “gives special solicitude to the rights of religious organizations”). Surely, though privileging of individual religions or the privileging of religion over non-religion is not allowed, some solicitude to religious exercise is required. Smith, 494 U.S. at 890 (“A society that believes in the negative protection accorded to religious belief can be expected to be solicitous of that value in its legislation . . . .”).


Even proaccommodationist Michael McConnell argues for “rigorous limitations” on accommodations. Michael W. McConnell, Accommodation of Religion: An Update and a Response to the Critics, 60 GEO. WASH. L. REV. 685, 687–88 (1992) (“[A]ccommodations are . . . sometimes required and, within rigorous limitations . . . are always permitted. That does not mean, of course, that every benefit to religion masquerading as an accommodation is constitutional, but it does mean that the principle of accommodation, when properly applied, is consistent with the requirements of the Religion Clauses.” (emphasis added)).
In *Cutter*, the Court stressed three important aspects of RLUIPA that established it, at least in a facial challenge, as a permissive legislative accommodation that fits between the Scylla and Charybdis\(^{60}\) of the Religion Clauses, demonstrating solicitude to religious exercise, an appropriate protection of religious freedom, rather than an inappropriate privileging of religion. First, to be permissive, an accommodation must “alleviate[] exceptional government-created burdens on private religious exercise.”\(^{61}\) A legislative act that protects religious expression by *removing government-imposed* burdens rather than creating privilege where no burden existed is more likely to be perceived as “an accommodation of the exercise of religion rather than as a Government endorsement of religion.”\(^{62}\) As a response to burdens imposed in government-run institutions, RLUIPA met this part of the test.

Second, courts must be satisfied that the permissive accommodation will be “administered neutrally among different faiths.”\(^{63}\) Noting there was no reason on the face of RLUIPA to believe that it would not be applied neutrally, the Court found this second aspect of the test satisfied.\(^{64}\) Third, in evaluating accommodations, the Court noted that courts “must take adequate account of the burdens a requested accommodation may impose on non-beneficiaries,” and, in granting an accommodation, courts must not “override other significant interests.”\(^{65}\) The Court held

---

\(^{60}\) In Greek mythology, Scylla and Charybdis were two monsters living on either side of the waterway between Italy and Sicily. Scylla was a six-headed beast who was said to eat ships and their sailors, while Charybdis was a whirlpool who would suck the boats and men down into her watery abyss. *Scylla and Charybdis, Myth Encyclopedia*, http://www.mythencyclopedia.com/Sa-Sp/Scylla-and-Charybdis.html. Ships rarely made it between the two unscathed. *Id.*

\(^{61}\) *Cutter*, 544 U.S. at 720.

\(^{62}\) *Id.* (quoting Corp. of Presiding Bishop of Church of Jesus Christ of Latter-day Saints v. Amos, 483 U.S. 327, 349 (1987)).

\(^{63}\) *Id.* (citing Bd. of Ed. of Kiryas Joel Village School Dist. v. Grumet, 512 U.S. 687, 705 (1994)) (statute that created special school district for one religious enclave and excluded all others violated the Establishment Clause).

\(^{64}\) *Cf. Kiryas Joel*, 512 U.S at 705.

\(^{65}\) *Cutter*, 544 U.S. at 720. For further discussion of “significant interests” that must be considered, see Gedicks & Van Tassel, *supra* note 33.
that RLUIPA must be applied so as not to elevate accommodation of religious observances over an institution’s need to maintain order and safety, and so found this third aspect of the permissive accommodation test satisfied as well.

RFRA is expected to survive on its face, not least because the government has not pressed a facial challenge. It is, after all, a legislative accommodation that allows plaintiffs of all religions to seek relief from a government imposed burden, and thus seems to meet the first, and perhaps the second, part of the test that the Court lays out in Cutter. However, the third criterion should create problems when analyzing whether RFRA is constitutional as applied to the case of contraceptive coverage requirements as discussed next.

C. RFRA—Unconstitutional As Applied to For-Profit Employers Seeking an Accommodation from the Contraceptive Coverage Requirements

Applying RFRA to grant accommodations from the contraceptive coverage requirements to for-profit employers, as requested in the two cases before the Court, is constitutionally problematic because it would not “take adequate account of the burdens a requested accommodation may impose on non-beneficiaries,” and thus would “override other significant interests” in violation of the Court’s limitations on permissive accommodations set out in Cutter. Professors Gedicks and Van Tassell argue that Establishment Clause doctrine prohibits “accommodations that shift the costs of an accommodated religion from those who practice it to those who don’t.”

66 See, e.g., id. (arguing that “it is likely that RFRA facially complies with the Establishment Clause”). But see Hamilton, Religious Freedom Restoration Act, supra note 33.

67 Cutter, 544 U.S. at 720.

68 See Gedicks & Van Tassel, supra note 33. Professors Gedicks and Van Tassell argue, “Neither courts nor commentators seem aware that a line of permissive-accommodation prohibits shifting of material costs of accommodating anticontraception beliefs from the employers who hold them to the employees who do not. The impermissibility of cost-shifting under the Establishment Clause is a threshold doctrine whose application is logically
employees who do not share their employer’s anticontraception beliefs would be denied their statutory and regulatory entitlement to contraception coverage without cost sharing, and thus would be directly saddled with material costs they would not incur in the absence of the exemption. Employees and their families would be deprived of the benefits of the Mandate to which they are otherwise legally entitled. The RFRA exemption would require that they pay the out-of-pocket expense of contraceptives and related services that they ought to receive at no expense beyond their monthly health care insurance premium. This is a direct burden that would not exist without the permissive accommodation of RFRA exemption. 69

Applying RFRA to grant exemptions from the mandate in the cases before the Supreme Court, the authors argue, would exceed the Establishment Clause’s “limits on permissive accommodation.” 70 Drawing from current doctrine, Gedicks and Van Tassell suggest a limitation on the right to free exercise that ends where the rights of nonadherents begin.

prior to all of the RFRA issues on which the courts are now focused.” Id. This argument is before the Court in the form of an amicus brief. See Brief for Amici Curiae Church-State Scholars Frederick Mark Gedicks et al., Sebelius v. Hobby Lobby Stores, Inc., Nos. 13-354 and 13-356, 2014 WL 333891 (U.S. Jan. 27, 2014).

69 Gedicks & Van Tassel, supra note 33, at 47-48 (noting that “[t]he externalized cost will be material for most employees. Effective oral contraceptive drugs cost between $180 and $960 per year, depending on the drug prescribed and the area of the country where the prescription is filled”).

70 Id. Yet, another “as-applied” Establishment Clause problem would arise if the RFRA was not “administered neutrally among different faiths.” See Cutter, 544 U.S. at 720. For example, if courts granted accommodations under the RFRA for one religion, like Christian practices for example, but for few or no others, one could argue that this mosaic of accommodations added up to a preference for Christianity over other religions in violation of the Establishment Clause. See id.
Kara Loewentheil takes this argument one step further.\textsuperscript{71} Like Gedicks and Van Tassell, she focuses on Cutter’s third limitation, the requirement that the courts consider the impact of accommodations on third parties. She argues that the current doctrine applicable to religious accommodation claims, under both the First Amendment and the Religious Freedom Restoration Act, is ill-suited to cases—like the contraceptive coverage requirement cases—in which the “primary conflict lies between different sets of non-state rights-holders: specifically religious objectors and existing rights-holders whose interests or rights would be negatively impacted (or completely blocked) by a grant of religious accommodation to an objector, particularly when such [existing rights-holders] have equality-implicating rights at stake.”\textsuperscript{72} For these cases, Loewentheil argues, a framework is needed that would “vindicate[e] the purpose of religious accommodation rights [while also] protecting [existing rights-holders] from the negative impact of accommodations.”\textsuperscript{73} While Loewentheil argues that “current doctrine can be argued to obliquely support an emphasis on the[] interests” of existing rights holders, she also proposes “a framework that places a positive obligation on the state to respect all the substantial rights involved when possible—and that prioritizes equality-implicating rights when not possible.”\textsuperscript{74} If the Supreme Court is brave enough in the Conestoga Wood and Hobby Lobby cases to deny the exemptions, an unlikely but vaguely possible outcome,\textsuperscript{75} the

\textsuperscript{71} See Loewentheil, supra note 33.
\textsuperscript{72} Id. at 501.
\textsuperscript{73} Id.
\textsuperscript{74} Id.
\textsuperscript{75} If the Court’s recent action in Little Sisters of the Poor is any indication, the Court is intent on avoiding any serious examination of these issues. Little Sisters of the Poor Home for the Aged v. Sebelius, No. 13A691, 2014 WL 272207 (U.S. Jan. 24, 2014). In Little Sisters of the Poor, the plaintiffs were eligible for a statutory exemption from the contraceptive coverage requirement. Nonetheless, they filed suit against the requirement and refused to comply with the administrative procedures created to notify the government and the insurance companies that they were eligible for the exemption, complaining that doing so would enable the insurance companies to provide the coverage on its own, thus making Little Sisters complicit in someone else’s sin. The Court simply created a different administrative
Court is most likely to deny them based on the harm to third party interests under either the Gedicks/Van Tassell or Loewentheil framework.

II. JUDICIAL REVIEW OF FREE EXERCISE CLAIMS

There remains a larger question at issue about the nature of “religious exercise” for which legislatures and courts are granting exemptions, a question that Loewentheil has pointed out is “a neglected and under-theorized area in accommodation law, with no satisfactory framework yet advanced.” Although this general question has always been at the heart of the Court’s difficulty and inconsistency in evaluating free exercise claims, as well as claims under RFRA, no one has adequately grappled with the specific question of when something constitutes an “exercise” of religion. This inquiry—including an inquiry into the “substantiality” of a burden on religious exercise required under RFRA—if it is to be meaningful, requires that questions be asked about the nature of the religious practice at issue, about what has been termed the “centrality” of a practice to religious belief. The problem for the courts is that these inquiries are precluded by Establishment Clause principles. Without examining the nature of religious “exercise,” however, there can be no meaningful limitation on mechanism that required the plaintiffs to inform the government in writing that they qualified for the exemption, relieved them of the responsibility of filling out the government form, and required the government to inform the insurance company on behalf of the objectors. Filling out the government form and sending that form to the insurance companies was apparently a “substantial burden” under the RFRA, but informing the government in writing of the same information and having the government inform the insurance companies was not. Of course, all this was done at the preliminary injunction stage and so “should not be construed as an expression of the Court’s views on the merits.”

---

Loewentheil, supra note 33, at 451 n.89. See also, e.g., Eisgruber & Sager, Why RFRA is Unconstitutional, supra note 33. Cf. Edward Whelan, The HHS Contraception Mandate v. The Religious Freedom Restoration Act, 87 NOTRE DAME L. REV. 2179, 2182 (2012) (claiming that “there can be no serious dispute that a person engages in an ‘exercise of religions . . . when, for religious reasons, he performs, or abstains from performing, certain actions’”).
what can be claimed as religious exercise. Unless some other limitation on RFRA claims is adopted, such as the requirement that third party interests not be harmed, there will be no limitation on what must be accommodated using the strict scrutiny test required by RFRA. This is exactly the Catch-22 the Court found itself in in Smith.

A. Smith’s Catch-22

The First Amendment’s free exercise doctrine has always struggled with questions about the nature of religious exercise that should be protected, with miserable results that tend to undermine the mandate for religious equality embodied in the Religion Clauses. The ultimate problem with application of strict scrutiny to free exercise claims, such as the ones in Smith, Conestoga Wood, and Hobby Lobby, is the difficulty of governing a society where a claim that a law interfered with one’s religious exercise mandates, almost automatically, that the person be granted an exemption from that law. By demanding the highest standard of review—requiring that governmental actions that substantially burden a religious practice must be justified by a compelling governmental interest—the strict scrutiny standard creates a presumption of invalidity for any regulation of conduct that is claimed to interfere with religious exercise. The Smith Court recognized that in the past it had applied the strict scrutiny standard—and thus the presumption of invalidity—inconsistently, striking down regulations in the face of complaints in very limited circumstances, without limiting the standard’s reach in the context of religion in any principled way. The reason it should be impossible to turn down a claim under strict scrutiny? Because of the First Amendment’s separate prohibition on judicial review of the “centrality” of conduct to an individual’s religion, the “relative merits of differing religious claims,” or “the

77 See Eisgruber & Sager, Why RFRA is Unconstitutional, supra note 33.
78 Smith, 494 U.S. at 888.
79 See id. at 883 (citing Sherbert v. Verner, 374 U.S. 398 (1963)).
80 Id. at 888 (“We cannot afford the luxury of deeming presumptively invalid, as applied to the religious objector, every regulation of conduct that does not protect an interest of the highest order.” (emphasis in original)).
determin[ation] of the place of a particular belief in a religion or the plausibility of a religious claim.”81 As the Court notes in Smith:

It is no more appropriate for judges to determine the “centrality” of religious beliefs before applying a “compelling interest” test in the free exercise field, than it would be for them to determine the “importance” of ideas before applying the “compelling interest” test in the free speech field.82

In support of its position that courts must not examine the centrality of religious belief, the Court in Smith cites a string of cases grounding the prohibition against judicial scrutiny of centrality of religious belief in the First Amendment and, in one case, specifically in the Establishment Clause.83 The Court cites

81 Id. at 887, 886 (“Nor is it possible to limit the impact of respondents’ proposal by requiring a ‘compelling state interest’ only when the conduct prohibited is ‘central’ to the individual’s religion.”).

82 Id. at 886–88.

83 Smith, 494 U.S. at 886–88 (citing Hernandez v. C.I.R., 490 U.S. 680, 699 (1989)) (“It is not within the judicial ken to question the centrality of particular beliefs or practices to a faith, or the validity of particular litigants’ interpretations of those creeds. We do, however, have doubts whether the alleged burden imposed by the deduction disallowance on the Scientologists’ practices is a substantial one.”); United States v. Lee, 455 U.S. 252, 263 n.2 (1982) (Stevens, J., concurring); Thomas v. Review Bd. of Ind. Emp’t Sec. Div., 450 U.S. 707, 715–16 (1981) (“One can, of course, imagine an asserted claim so bizarre, so clearly nonreligious in motivation, as not to be entitled to protection under the Free Exercise Clause; but that is not the case here, and the guarantee of free exercise is not limited to beliefs which are shared by all of the members of a religious sect. Particularly in this sensitive area, it is not within the judicial function and judicial competence to inquire whether the petitioners or his fellow worker more correctly perceived the commands of their common faith. Courts are not arbiters of scriptural interpretation.”); Presbyterian Church in U.S. v. Mary Elizabeth Blue Hull Mem’l Presbyterian Church, 393 U.S. 440, 450 (1969) (refusing to conduct inquiry “to determine matters at the very core of a religion—the interpretation of particular church doctrines and the importance of those doctrines to the religion,” because “[p]lainly, the First Amendment forbids civil courts from playing such a role”); Jones v. Wolf, 443 U.S. 595, 602 (1979) (noting that “the First Amendment severely circumscribes the role that civil courts may play in resolving church property disputes . . . . Most importantly, the First Amendment prohibits civil courts from resolving church property disputes on
footnote number two of Justice Stevens’s concurrence in *United States v. Lee* in which Justice Stevens explained that the principal reason for avoiding this inquiry is the overriding interest in keeping the government—whether it be the legislature or the courts—out of the business of evaluating the relative merits of differing religious claims. The risk that governmental approval of some and disapproval of others will be perceived as favoring one religion over another is an important risk the Establishment Clause was designed to preclude.84

Because the Establishment Clause precludes the Court from reviewing the validity of a claim of religious interference, if a compelling interest test was to be applied, it must be applied across the board, to all actions claimed to be religiously commanded. This is different, as the Court seemed to be admitting, from how the Court had applied the compelling interest test piecemeal, read discriminatorily, in the past. The Court then listed a number of important statutes, such as manslaughter and child-abuse statutes, from which exemptions could be sought. Importantly, the purpose of this list was not to suggest that the courts would necessarily grant exemptions from these eminently reasonable statutes, but to point out that denial of any of these exemptions would require denial of all exemptions. Any grant of one exemption but not another put the Court in violation of the Establishment Clause because it risked that the Court was itself comparing the weight of the burden on one person’s religious belief as against the weight of the burden on another person’s religious belief, a comparison that required prohibited inquiry into the centrality of a practice under religious

---

84 *Lee*, 455 U.S. at 263 n.2 (Stevens, J., concurring). *See also Smith*, 494 U.S. at 887 (discussing the “unacceptable ‘business of evaluating the relative merits of differing religious claims’”).
Thus, the real problem that the Court faced in *Smith* is the impossibility of evaluating the extent of a burden on religious exercise without inquiring into the “centrality” of the religious belief being claimed. In *Smith*, Justice Scalia writing for the Court recognized the Catch-22 that this created. On the one hand, courts cannot inquire into centrality because of the risk of creating a widespread perception of favoritism that will lead to internecine conflicts between individuals of different faiths and faith traditions. On the other hand, “[d]ispens with a ‘centrality’ inquiry is utterly unworkable.” As Justice Scalia wrote for the Court:

> [i]t would require, . . . the same degree of “compelling state interest” to impede the practice of throwing rice at church weddings as to impede the practice of getting married in church. There is no way out of the difficulty that, if general laws are to be subjected to a “religious practice” exception, both the importance of the law at issue and the centrality of the practice at issue must reasonably be considered.”

As the Court in *Smith* recognized, if strict scrutiny is indeed to be applied, “it must be applied across the board, to all actions thought to be religiously commanded.” This is to say it must be applied consistently and fairly. But “if ‘compelling interest’ really means what it says, . . . many laws will not meet the test.” In *Smith*, the Court was distressed at the prospect of applying a real strict scrutiny test that would require courts to grant

---

85 *Smith*, 494 U.S. at 888–89 (listing compulsory military service, payment of taxes, manslaughter and child neglect laws, compulsory vaccination laws, traffic laws, minimum wage legislation, child-labor and animal-cruelty laws, laws protecting the environment, and equality of opportunity for the races).

86 *Id.* at 886–87.

87 *Id.* at 887 n.4.

88 *Id.* (emphasis in original).

89 *Id.* at 888.

90 *Id.*

91 *Id.* (noting that because “we are a cosmopolitan nation made up of
accommodations from all sorts of “reasonable” laws, such as laws prohibiting drug use to laws prohibiting murder and mayhem, all in the service of religious belief. In the face of this crisis, the Court throws up its hands and passes the buck to the legislature. Categorically unable to adjudicate such claims in a manner that does not create the appearance of establishing some religions as favored and some as disfavored, the Court in Smith concluded it was proscribed from deciding whose claims to individual religious exemptions were valid and whose were invalid under the Free Exercise Clause by independent counterforce embodied in the Establishment Clause.92

The Justices struggle with their inability to review religious practice to determine the scope of Free Exercise claims throughout the opinions in Smith. However, while swearing off such inquiries in writing, in practice the Courts have been unable to resist making these determinations in Free Exercise cases. For example, Justice O’Connor in her concurrence repeats the Court’s constant refrain that “it is not within the judicial ken to question the centrality of particular beliefs or practices to a faith,”93 and states that the determination of the constitutionality of the ban on peyote use “cannot, and should not, turn on the centrality of the particular religious practice at issue.”94 On the other hand, Justice O’Connor proposes,

the sounder approach—the approach more consistent with our role as judges to decide each

people of almost every conceivable religious preference,” the compelling interest standard would require exemptions “from civic obligations of every conceivable kind” (quoting Braunfeld v. Brown, 366 U.S. 599, 606 (1961))).

92 This is not a wholly unusual situation in Constitutional law. The Court has held that the Constitution’s commitment to equality means that race-conscious measures to remedy racially-disparate impacts must also be narrowly circumscribed to avoid embroiling courts in a process of enacting the very inequality they seek to remedy. See Ricci v. DeStefano, 557 U.S. 557 (2009). Courts are limited in their powers to enjoin and limit speech to protect the privacy of individuals in legal proceedings lest they themselves violate the First Amendment, even though privacy is generally regarded as itself protected by the First Amendment at least to some extent.


94 Id. at 906–07.
case on its individual merits—is to apply [the compelling state interest test] in each case to determine whether the burden on the specific plaintiffs before us is constitutionally significant and whether the particular criminal interest asserted by the State before us is compelling.  

Justice O’Connor would allow the courts to make factual findings “as to whether a claimant holds a sincerely held religious belief that conflicts with, and thus is burdened by, the challenged law.” Admitting that “[t]he distinction between the question of centrality and questions of sincerity and burden” is “fine,” Justice O’Connor nonetheless insists that “it is one that is an established part of our Free Exercise doctrine, and one that courts are capable of making.” She then determines that the prohibition of peyote places a severe burden on the respondents’ religious practice, and that the state has a compelling interest in controlling use of illegal drugs. Finally, while claiming she is not questioning the centrality of the peyote use to the church, Justice O’Connor appears to do just that, questioning whether the claimant holds a “sincerely held religious belief that conflicts with, and thus is burdened by, the challenged law.”

Justice Blackmun writing in dissent similarly demonstrates the problem with review of the sincerity of religious belief or the centrality of a belief to religious practice. Justice Blackmun, like Justice O’Connor, applied strict scrutiny, but weighed the importance of the peyote ritual differently than Justice O’Connor and so would have granted an exemption from the peyote ban. Like the other Justices, Blackmun begins by agreeing that “courts should refrain from delving into questions whether, as a matter of religious doctrine, a particular practice is ‘central’ to the

---

95 Id. at 899.
96 Id. at 907.
97 Id. (citing Tony & Susan Alamo Found. v. Sec’y of Labor, 471 U.S. 290, 303–05 (1985); United States v. Ballard, 322 U.S. 78, 85–88 (1944)).
98 Id. at 903–04 (O’Connor, J., concurring) (citing scholarly work on Peyote Religion).
99 Id. at 907.
100 Id. at 919 (Blackmun, J., dissenting).
religion.”¹⁰¹ He goes on, though, to advocate that courts do just that, citing to Yoder’s determination that “education is inseparable from and a part of the basic tenets of the [Amish] religion . . . [, just as] baptism, the confessional, or a Sabbath may be for others,” noting “I do not think this means that the courts must turn a blind eye to the severe impact of a State’s restrictions on the adherents of a minority religion.”¹⁰² He then finds that “[w]ithout peyote, [the Respondents] could not enact the essential ritual of their religion.”¹⁰³

Justice Scalia writing for the Court explicitly rejects the approaches of both Justice O’Connor and Justice Blackmun. In response to Justice O’Connor, he dismisses the “fine” distinction between judicial review of the centrality of religious belief and review of the “significance” of a burden, noting that “‘[c]onstitutionally significant burden’ would seem to be ‘centrality’ under another name.”¹⁰⁴ To Justice Blackmun he similarly replies that there is no difference between inquiry into “severe impact” and inquiry into “centrality” of a religious belief and notes that Blackmun’s evaluation of the impact of the peyote ban demonstrates this fact. Justice Scalia declares that Justice Blackmun “has merely substituted for the question ‘How important is X to the religious adherent?’ the question ‘How great will be the harm to the religious adherent if X is taken away?’ There is no material difference.”¹⁰⁵

It is into Smith’s breach that Congress—emboldened by support from left, right, and middle—inserted RFRA, leaving the Court in the untenable position of reviewing the “substantiality” of the burden on “religious exercise,” the inquiry the Court rejected in Smith as precluded by the Establishment Clause.¹⁰⁶

¹⁰¹ Id.
¹⁰² Id.
¹⁰³ Id. at 919–20 (Blackmun, J., dissenting) (citing a brief filed by the Association on American Indian Affairs et al., a scholarly history of the peyote religion, and a popular mystery novel by Tony Hillerman, describing ritual in which the “sacrament Peyote is the means for communicating with the Great Spirit”).
¹⁰⁴ Id. at 887 n.4 (majority opinion).
¹⁰⁵ Id.
¹⁰⁶ It could be argued that “substantiality” of a burden is a different
Under RFRA, Congress demands the “horrible” result the Court decried in *Smith*, mandating that “federal judges will regularly balance against the importance of general laws the significance of religious practice.”

**B. Breadth of the Hobby Lobby and Conestoga Wood Claims**

One could argue that there are cases that would not require the Court to evaluate the significance of certain behaviors and their importance to religious practice in a way that implicates the concerns articulated in *Smith*. For example, challenges to laws that so obviously burden clear and well-established rules of a given religion, such as a law that prevents a person from becoming a minister to her chosen congregation, a law that prevents a person from using the sacramental wine, a law that prevents a person from using peyote in a religious ceremony, these might not test the court or require it to evaluate religious doctrine in a threatening way. These all seem obviously burdensome in a substantial way, even though the last was not obvious to the Court in *Smith*. One could argue that because these laws directly contradict central aspects of the actual ceremonial “celebration” of religion, the courts can avoid the Catch-22 because they need not enter into a prohibited area of review to hold that these laws violate RFRA.

But in *Hobby Lobby* and *Conestoga Wood* the claims are more complicated. These plaintiffs are not claiming that how they celebrate their religion is burdened. Rather, they claim that by offering plans that include the means to obtain contraception, they are somehow complicit in what they see as the sin of the person who chooses to use contraception, even if the person who

---

107 *Smith*, 494 U.S. at 889 n.5. See Levine, supra note 9, at 122–23 (noting that RFRA’s substantial burden test “appear[s] to require courts to engage in the kind of investigation into religious beliefs that Supreme Court Justices have increasingly and nearly uniformly rejected”).
uses the contraception does not see it as a sin, and indeed, even if the person who uses contraception sees it as religiously mandated. They argue that their “religious exercise” will be “substantially burden[ed]” under RFRA if the insurance plans provided to their employees include contraceptives as part of a package of preventive services, even if the employer does not pay anything for the contraception, and even though the employee’s choice to use or not use contraception stands between the employer and the employee’s alleged “sin.” Furthermore, they claim this chain of causation between sin and employer is not broken in this circumstance, but that it is broken where the employee uses other financial benefits they receive from their employers, i.e., salary, to pay for contraception. Moreover, the plaintiffs in *Hobby Lobby* make this claim even though the health insurance plans they offered before the ACA required contraceptives to be covered, covered contraception. *Hobby Lobby* only dropped coverage for contraceptives when federal law required coverage, a fact that should lead the Court to question the substantiality of the burden imposed.\(^\text{108}\) The breadth of these claims puts the Court squarely in forbidden waters.

This Catch-22 came to a head recently in the oral argument and subsequent decision in *Notre Dame v. Sebelius*, a case which involves the more “fantastic” claim that even requiring Notre Dame to *invoke* the statutory exemption granted under ACA from the contraception coverage requirements violates RFRA.\(^\text{109}\) As Notre Dame’s counsel stated, if Notre Dame believes it is a substantial burden on its religious exercise to even apply for the accommodation, the court must grant an accommodation *from applying for the accommodation*, no questions asked: “It is up to the believer to draw the line.”\(^\text{110}\)

\(^{108}\) *See* Eden Foods, Inc. v. Sebelius, 733 F.3d 626 (8th Cir. 2013) (questioning sincerity of the plaintiff’s claimed religious belief); *see also infra* note 124.

\(^{109}\) Univ. of Notre Dame v. Sebelius, No. 13-3853, slip op. at 20 (7th Cir. Feb. 21, 2014) (“What makes this case and others like it involving the contraception exemption paradoxical and virtually unprecedented is that the beneficiaries of the religious exemption are claiming that the exemption process itself imposes a substantial burden on their religious faiths.”).

The *Notre Dame* panel rejected this position, holding that the burden of invoking a statutory exemption was minimal, and that though “Notre Dame may consider the process a substantial burden, but substantiality—like compelling governmental interest—is for the court to decide.”\(^{111}\) Notably, though, the court escaped the Catch-22 because the minimal nature of the
colloquy went as follows:

Judge Hamilton: We have a long history in this country of accommodating religious faith in various ways that are not required by free exercise or prohibited by the Establishment Clause, there’s some play in the joints, we know . . . . Can you point me to any other example in our legal history where the accommodation *itself* has been challenged as a burden on free exercise?

Mr. Kairys: You mean other accommodations? This mandate is new.

Judge Hamilton: This mandate is new, yes, I’m trying to understand though . . . .

Mr. Kairys: I cannot your honor . . . .

Judge Hamilton: [h]ow complying with minimal . . . I mean, to provide an accommodation at all requires at least some minimal invocation, say “yes I want to take advantage of the accommodation.”

Mr. Kairys: Right, but it is up to the believer to draw that line. And Notre Dame has made that religious determination and Korte says it is not for you to engage in this issue of minimal or not minimal.

Judge Hamilton: It sounds like what you are telling us is that the entire U.S. Code then is subject to strict scrutiny any time somebody raises a sincere religious objection.

*Id.*; see also *id.* at 18:50 (where counsel for Notre Dame argues that under RFRA, “It is not for the Court to determine what is ‘meaningful’ or what’s ‘insignificant.’ [Notre Dame has] made their own religious determination that the role required of it from signing the form to maintaining a contractual relationship [violates its religious tenets].)

\(^{111}\) *Notre Dame*, No. 13-3853, slip op. at 21. The Court held that the paperwork burden is “the opposite of cumbersome,” *id.* at 13, and while expressing skepticism of the second burden, the “triggering” burden, by writing “[t]hat seems a fantastic suggestion,” *id.* at 18, the court ultimately finds it unconvincing as a factual matter, *id.* at 15 (“Notre Dame’s signing the form no more ‘triggers’ Meritain’s obligation to provide contraceptive services than a tortfeasor’s declaring bankruptcy ‘triggers’ his co-tortfeasors’ joint and several liability for damages.”).
burden claimed made the claim practically ridiculous in its eyes. The court distinguished a different Seventh Circuit case, *Korte v. Sebelius*,\(^{112}\) in which the panel had an opposite reaction to the RFRA Catch-22,\(^{113}\) taking the hands-off approach in the more difficult case of two individual owners of a for-profit business that was not entitled to the statutory exemption, and who, like the plaintiffs in *Hobby Lobby* and *Conestoga*, sought an accommodation under RFRA. In *Korte*,\(^{114}\) the panel upheld the plaintiffs’ RFRA claims, and contradicting the *Smith* Court’s determination, wrote:

>[i]mportantly, the substantial-burden inquiry does *not* invite the court to determine the centrality of the religious practice to the adherent’s faith; RFRA is explicit about that. And free-exercise doctrine makes it clear that the test for substantial burden does not ask whether the claimant has correctly interpreted his religious obligations. Indeed, that inquiry is prohibited . . . . It is enough that the claimant has an “honest conviction” that what the government is requiring, prohibiting, or pressuring him to do conflicts with his religion.\(^{115}\)

The *Korte* panel attempted to draw a distinction between “sincerity” and “religiosity,” both factual inquiries it claimed are within the court’s authority and competence, and the “substantial-burden” inquiry which the panel argues, along with the Tenth Circuit, is “primarily” an evaluation of the “intensity of the coercion applied by the government to act contrary to [religious] beliefs.”\(^{116}\) By defining the substantial burden inquiry as an evaluation of the “coercive effect of the governmental pressure on the adherent’s religious practice,” as the adherent defines the religious practice, *Korte* intends to steer the substantial burden inquiry “well clear of deciding religious questions.”\(^{117}\)

\(^{112}\) 735 F.3d 654 (7th Cir. 2013).

\(^{113}\) *Notre Dame*, No. 13-3853, slip op. at 21 (distinguishing *Korte*, 735 F.3d at 654).

\(^{114}\) *Korte*, 735 F.3d at 654.

\(^{115}\) *Id.* at 683 (citations omitted).

\(^{116}\) *Id.*

\(^{117}\) *Id.*
Notre Dame panel declined to state its disagreement with the Korte panel on this view of the substantial burden inquiry explicitly, if the “coercive effect” of the government pressure were the only issue, the result would have been the same in both cases, because the penalty for failure to comply is the same. Instead, in Notre Dame, the panel refuses to abdicate its role of reviewing substantial burden, but it still appears to leave open the possibility that abdication of judicial review of the burden must occur where the question is harder. If the hands-off view of judicial review of burden prevails under Hobby Lobby and Conestoga Wood, as it did in Korte, the courts’ role under RFRA will simply be to decide whether the plaintiffs are forced to do something they claim violates their religious beliefs, something they oppose other people doing because it contrasts with their moral beliefs as religious people. The courts will not be conducting a meaningful review of whether or not the action burdens, in any significant way, their exercise of religion. As Judge Hamilton warns during oral argument in Notre Dame, “It sounds like what you are telling us is that the entire U.S. Code then is subject to strict scrutiny any time somebody raises a sincere religious objection.”

III. AVOIDING THE CATCH-22

The Court recently heard argument in Hobby Lobby and Conestoga Wood and may issue decisions before this essay goes to print. I conclude here by discussing three possible outcomes of these cases.

A. Limiting RFRA

There are two ways the Court in Hobby Lobby and Conestoga Wood could limit RFRA to avoid the Catch-22 altogether. First, the Court could recognize that accommodations may not be granted in these cases, because the harm to third parties is too

---

118 Notre Dame, No. 13-3853, slip op. at 21 (distinguishing Korte, 735 F.3d at 654).
great, under either the Gedicks/Van Tassell or the Loewentheil theory.\textsuperscript{120} If the Court adopts this approach, it will not be called upon to address the substantiality of the burden on religious exercise under RFRA. Second, the Court could interpret “religious exercise” under RFRA strictly to mean only those actions that constitute religious “practice.” Limiting “religious exercise” to actions such as, for example, celebrating religion or wearing religious garb identifying one’s religious affiliation, would allow the Court to avoid the issue in this case because the attenuated claims of harm here would not qualify as this type of religious exercise.\textsuperscript{121}

\textbf{B. Choosing “All” or Choosing “Some”}

If the Court does consider the merits of the claims here, it strays into the area of considering claims for exemptions from secular mandates because the mandates prevent someone from living one’s life “in accordance with one’s religious belief.” Consideration of such claims leads the Court into the dangerous waters between Scylla and Charybdis. Either the Court will choose “all,” accepting at face value any claim that a secular rule conflicts with a religious belief, or it will choose “some,” and evaluate whether being tangentially involved in giving a third party the freedom to act or not act in a way that would conflict with one’s own religious belief is itself in violation of the plaintiff’s religious belief.

The Court could agree with the \textit{Korte} panel and the position of the University of Notre Dame and hold that it is the believer who draws the line. It may be that the Court will distinguish RFRA’s requirement that the Court conduct a review of the substantiality of a burden and the same inquiry conducted under the Free Exercise Clause, based on the former being a

\textsuperscript{120} See supra notes 66–75 and accompanying text.

\textsuperscript{121} See \textit{Hosanna-Tabor Evangelical Lutheran Church & Sch. v. EEOC}, 132 S.Ct. 694, 711–12 (2012) (“The First Amendment protects the freedom of religious groups to engage in certain key religious activities, including the conducting of worship services and other religious ceremonies and rituals, as well as the critical process of communicating the faith.”) (Alito and Kagan, JJ., concurring).
legislatively mandated review while the latter was a judicially created standard. The Court could then conduct a perfunctory review of “substantial burden,” applying a very limited inquiry into the burden the attenuated claims place on the plaintiffs in these cases. It is also possible that the Court will hold that the connection here between government action and religious exercise is simply too attenuated to amount to a substantial burden on religious exercise.

In any case, application of RFRA’s standard to allow for-profit businesses an exemption from the ACA contraceptive coverage requirement, like application of the strict scrutiny standard to free exercise claims under the Constitution, would, as the Court warned in *Smith*:

> open the prospect of constitutionally required religious exemptions from civic obligations of almost every conceivable kind—ranging from compulsory military service, . . . to the payment of taxes . . . to health and safety regulation such as manslaughter and child neglect laws, . . . compulsory vaccination laws, . . . , drug laws, . . ., and traffic laws . . .; to social welfare legislation such as minimum wage laws, . . ., child labor laws, . . ., animal cruelty laws, . . . environmental protection laws, . . . and laws providing for equality of opportunity for the races, . . .

122

Given that many think the Court will grant the exemptions in this case, we should prepare ourselves to use RFRA to enforce equal protection for our own religious freedom, and to challenge restrictions on actions mandated by our consciences in our relationships with our own “divinities.” Courts have rejected the claims of the Church of Marijuana, 123 but the claims of many religious people, for example, religious people who were integrally involved in the movement for reproductive freedoms in

---

123 United States v. Meyers, 95 F.3d 1475, 1484 (10th Cir. 1996) (denying RFRA claim of adherent of the Church of Marijuana as espousing a philosophy and/or way of life rather than a “religion”).
the 1960s and 1970s will not be so easily sloughed off. Moreover, if the Court grants the exemptions in these cases, any attempt to deny claims brought under RFRA by those whose religious beliefs lead them to choose abortions or contraceptives, or those whose religious beliefs mandate they make abortions or contraceptives available to others, would be an unconstitutional application of RFRA under the Establishment Clause. We await the Court’s move.

CONCLUSION

I am not optimistic that the Court will follow these suggestions. In defending the contraceptive coverage requirements, the federal government has been extremely reticent to criticize RFRA from what I imagine is a political desire to avoid seeming hostile to any religious claims. They have studiously avoided questioning the sincerity of religious beliefs and tried to avoid anything that would entangle the Court in questioning the claimed burdens on religious exercise. The government has done this despite ample evidence in this and other challenges to the contraceptive mandates that these claims are being made only as part of a broader objection to federal power, and/or as an effort to prevent women from accessing contraception. Nor has anyone questioned Catholic plaintiffs’

124 In its recent decision in Eden Foods, Inc. v. Sebelius, 733 F.3d 626 (8th Cir. 2013), the Eighth Circuit questioned the sincerity of the plaintiff’s claimed religious belief. The court notes that the plaintiff Michael Potter, a Roman Catholic, claims that he

follows the teachings of the Catholic Church, and has

“deeply held religious beliefs” “that prevent him from participating in, paying for, training others to engage in, or otherwise supporting contraception, abortion, and abortifacients.” In fact, Potter claims that “these procedures almost always involve immoral and unnatural practices.”

Id. at 629. The court then notes in a footnote:

Interestingly, in a conversation with salon.com’s Irin Carmon, Potter’s “deeply held religious beliefs,” more resembled a laissez-faire, anti-government screed. Potter stated to Carmon, “I’ve got more interest in good quality long underwear than I have in birth control pills.” Carmon
objections to providing access to all contraceptions despite the approval by the Catholic Church of these medications when used therapeutically, not as a form of contraceptive.\footnote{125} In fact, unless the Court follows the path recommended by Gedicks and Van Tassell or Loewentheil in implementing an alternative limitation on the grant of accommodations under RFRA, I suspect that the Court will grant an accommodation in the contraception cases, while refusing accommodations in future cases to those whose religions are less palatable to them,\footnote{126} like mine.\footnote{127} This will

then asked the Eden Foods chairman why he didn’t seem to care about birth control when he had taken the step to file a lawsuit over the contraceptive mandate. Potter responded, “Because I’m a man, number one[,] and it’s really none of my business what women do.” The article continued: So, then, why bother suing? “Because I don’t care if the federal government is telling me to buy my employees Jack Daniel’s or birth control. What gives them the right to tell me that I have to do that? That’s my issue, that’s what I object to, and that’s the beginning and end of the story.” He added, “I’m not trying to get birth control out of Rite Aid or Wal-Mart, but don’t tell me I gotta pay for it.”

\textit{Id.} at 629 n.3 (citation omitted). Similarly, in the \textit{Hobby Lobby} case, the plaintiff’s “religious exercise” only became “substantially burdened” when the federal government adopted the contraceptive coverage requirement. \textit{See, e.g.}, Jaime Fuller, \textit{Here’s What You Need to Know About the Hobby Lobby Case}, WASH. POST (Mar. 24, 2014), \url{http://www.washingtonpost.com/blogs/the-fix/wp/2014/03/24/heres-what-you-need-to-know-about-the-hobby-lobby-case/}.

Until that point, the health plans offered contraceptive coverage without objection from the plaintiffs.

\footnote{125} The Humanae Vitae, the document setting out the Roman Catholic Church’s position on contraception, permits therapeutic uses of contraceptives to treat organic diseases, even though they have a contraceptive effect. \textit{PAUL VI, HUMANAE VITAE} (1968), \url{http://www.vatican.va/holy_father/paul_vi/encyclicals/documents/hf_p-vi_enc_25071968_humanae-vitae_en.html} (“On the other hand, the Church does not consider at all illicit the use of those therapeutic means necessary to cure bodily diseases, even if a foreseeable impediment to procreation should result there from—provided such impediment is not directly intended for any motive whatsoever.”).

\footnote{126} \textit{See, e.g.}, Meyers, 95 F.3d at 1484 (denying RFRA claim of adherent of the Church of Marijuana as espousing a philosophy and/or way of life rather than a “religion.”).

\footnote{127} \textit{See} Planned Parenthood v. Casey, 505 U.S. 833, 850 (“Men and women of good conscience can disagree, and we suppose some always shall
produce exactly the Establishment Clause violation the Court rejected in *Smith* and Justice Stevens warned against in *Lee*. If this is the route the Court takes, it should expect that its religious neutrality will be tested, and the religious underpinnings of civil rights movements will rise again.

INTRODUCTION

A generation ago, only 1% of the United States population lived in a privately governed common-interest community ("CIC"). Today, approximately 64 million people (20% of the...
country’s population) reside in one of the more than 300,000 CICs in the United States.\(^2\) Residents in CICs are bound to a private governance scheme that includes written obligations that have been recorded in the local land records and run with the land as well as rules and regulations enacted from time to time by the board of directors of the community association.\(^3\) These covenants and rules form the private law of the community, and generally courts will grant injunctions or specific performance to enforce such regulations. State law also permits a CIC association to assess lien-backed fines for non-compliance.\(^4\) Buyers of homes in a CIC are deemed to have voluntarily elected to be legally bound to all the private community rules, to have such rules specifically enforced, and to subject their property to a security interest securing their obligations to the community.

For the most part, courts do not undertake a substantive

\(^{1}\) CAI tracks data regarding the number of CICs and their residents. \(\text{CMTY. ASS'NS INST., INDUSTRY DATA, http://www.caionline.org/info/research/Pages/default.aspx (last visited Feb. 12, 2014). CAI’s data indicates that the number of residents of common interest communities has increased to 63.4 million today. This figure represents 20.2% of the population of the U.S.A., estimated by the U.S. Census Bureau in 2012 to be approximately 313.9 million. \textit{U.S. & World Population Clock}, U.S. CENSUS BUREAU, http://www.census.gov/popclock/ (last visited Feb. 12, 2014). The percentage of the population residing in a CIC continues to grow. WAYNE S. HYATT & SUSAN F. FRENCH, \textit{COMMUNITY ASSOCIATION LAW: CASES AND MATERIALS ON COMMON INTEREST COMMUNITIES} 3 (2d ed. 2008); Andrea J. Boyack, \textit{Community Collateral Damage: A Question of Priorities}, 43 LOY. U. CHI. L.J. 53, 58 (2011) [hereinafter Boyack, \textit{Community Collateral Damage}]. The proliferation of the CIC form is not uniformly heralded as a positive development. See David E. Grassmick, \textit{Minding the Neighbor’s Business: Just How Far Can Condominium Owners’ Associations Go in Deciding Who Can Move into the Building?}, 2002 U. ILL. L. REV. 185, 189 (asserting that in a sort of “Gresham’s Law” (bad money drives out good) a “condominium or owners’ association-governed community is crowding out other types of housing from the market”).

analysis of the desirability of individual community covenants.\textsuperscript{5} Courts reason that all members of a community have agreed to be contractually bound to this private governance scheme,\textsuperscript{6} and therefore judicial deference to community choices is mandated by freedom of contract policies.\textsuperscript{7} The proper judicial role, under this conception of the CIC, is to ensure that any changes to the private legislative content (covenant amendments or rule enactments) occur according to the privately enumerated process.\textsuperscript{8} Focusing

\textsuperscript{5} For example, the court in \textit{Powell v. Washburn}, 125 P.3d 373, 376 (Ariz. 2006), enjoined a homeowner from keeping a recreational vehicle on his property by holding that CIC covenants should be enforceable according to the intent of the drafting party, specifically departing from and rejecting the rule of strict construction of covenants that run with the land. \textit{See also} Jeffrey A. Goldberg, Note, \textit{Community Association Use Restrictions: Applying the Business Judgment Doctrine}, 64 CHI.-KENT L. REV. 653, 673; (1998); Robert G. Natelson, \textit{Consent, Coercion, and “Reasonableness” in Private Law: The Special Case of the Property Owners Association}, 51 OHIO ST. L.J. 41, 45–47 (1990).

\textsuperscript{6} \textit{Hidden Harbour Estates, Inc. v. Basso}, 393 So. 2d 637, 639–40 (Fla. Dist. Ct. App. 1981) (asserting that CIC restrictions “are clothed with a very strong presumption of validity which arises from the fact that each individual unit owner purchases his unit knowing of and accepting the restrictions to be imposed”); \textit{see also}, e.g., \textit{Lookout Mountain Paradise Hills Homeowners’ Ass’n v. Viewpoint Assocs.}, 867 P.2d 70, 74 (Colo. App. 1993); Joslin v. Pine River Dev. Corp., 367 A.2d 599, 601 (N.H. 1976).

\textsuperscript{7} Courts reason that while a community’s group preferences may not coincide with individual owner preferences, those owners have agreed to subordinate their individual wishes to the choices of a group. This concept, that the interrelationship among owners in a CIC justifies some curtailment of individual rights, is a fundamentally accepted aspect of CIC covenant enforcement. Hidden Harbour Estates, Inc. v. Norman, 309 So. 2d 180, 181–82 (Fla. Dist. Ct. App. 1975) (“[T]o promote the health, happiness, and peace of mind of the majority of the unit owners since they are living in such close proximity and using facilities in common, each unit owner must give up a certain degree of freedom of choice which he might otherwise enjoy in separate, privately owned property.”).

\textsuperscript{8} If regulations and amendments apply equally to all members and are promulgated according to the procedures set forth in the governing documents, courts will generally uphold them. \textit{Hyatt}, \textit{supra} note 4, at 56, 173. \textit{See}, e.g., Kroop v. Caravelle Condo, Inc., 323 So. 2d 307 (Fla. Dist. Ct. App. 1975) (holding that amendments severely limiting an owner’s right to lease his unit were valid because the amendment was passed according to the procedure set forth in the CIC declaration).
solely on how covenants are amended and how rules are enacted does ensure that community members enjoy some level of procedural due process with respect to changes to CIC governing provisions. However, there is little actual substantive limit on the covenants and regulations that CICs can impose, either through amendment or as part of the original recorded covenants and community rules.

Courts unrealistically presume that purchasing property within a CIC is in itself an adequate manifestation of assent to be bound to CIC governing provisions. General deference to parties’ substantive choices in contracting is proper. But freedom of contract is an inadequate justification for covenant enforcement in the context of privately governed communities. Such covenants do not necessarily represent voluntary owner assent to obligation and do not necessarily reflect neighborhood preferences. The covenants are perpetual, non-negotiable contracts of adhesion, bundled with one of the most personal, expensive, and complicated purchases an individual will ever make—the purchase of a home.\(^9\) As servitudes, CIC covenants enjoy duration and specific enforceability that go beyond typical contract rights.\(^10\) In addition, the terms of a community’s laws are not self-imposed; instead, they are crafted by developers and driven by the requirements of lenders and governments.\(^11\) The only escape from a given CIC governance scheme is sale of one’s home,\(^12\) and in some markets even this will be ineffective due to lack of real choice among residential neighborhood options.\(^13\)

Contract analogy should not create presumptive validity for all CIC covenants and properly enacted rules. The reality of CIC governance is more complicated and implicates property and

---

\(^9\) See infra Part II.A.


\(^11\) See infra Part II.D. The content of CIC covenants is motivated in part by mortgage market constraints imposed by federal agencies or Fannie Mae and Freddie Mac.

\(^12\) See infra Part II.C.

\(^13\) See infra notes 94–95 and accompanying text.
constitutional concerns as well as contract law. The proper approach to CIC governance review must draw from all three of these areas of the law. The subject matter scope of CIC governance should be limited based on servitude law principles. Constitutional protections should be legislated for members of CICs. And bona fide, deliberate assent should be prerequisite to holding owners bound to CIC obligations.

Part I of this Article explores the origins and judicial treatment of the private laws of self-governed communities. CIC covenants are legal hybrids—enforced as contracts but specifically enforceable against successive landowners because they are servitudes. Part II explains how CIC covenants and rules diverge from the typical contractual model. CIC covenants are contracts of adhesion, made up of completely non-negotiable, recorded terms bundled into home acquisition. Developers and lenders generally prescribe the content of such covenants, and they may not reflect community desires or values. Part III explains how a refocused freedom of contract rationale, an updated variant of traditional servitude requirements, and new legislation regarding important personal freedoms can bring clarity and fairness to common interest community law.

I. THE CURIOUS CASE OF COMMON INTEREST COMMUNITIES

A. Legal Hybrids: Contracts Enforced as Servitudes and Functioning as Constitutions

CICs are creatures both of property law and of contracts. In terms of function, they are akin to “mini governments.”

foundational structure of CICs, however, is servitude law.\textsuperscript{15} In a CIC, all property owners are bound together under a system of real covenants and share certain financial obligations and property rights.\textsuperscript{16} Every property owner within a CIC is also a mandatory member of a contractually defined association that provides private governance for the community.\textsuperscript{17} The power of an association to govern, to assess owners for upkeep, and to enforce rules regarding use and appearance of individual properties is established through a recorded declaration of covenants (sometimes called CC&Rs). These covenants bind all successive owners of the property by virtue of their ownership, a concept called “running with the land.” Although framed much like a multilateral contract, CIC covenants transcend typical contractual obligation and become obligations of the property itself, binding its successive owners and specifically enforceable in perpetuity.\textsuperscript{18} The covenant obligations in CICs are not static because the association can amend the CC&Rs or pass rules to further clarify or carry out the purposes of the community.\textsuperscript{19}

\textsuperscript{15} A servitude is a legal device that creates a right or obligation that runs with the land. \textsc{Restatement (Third) of Prop.: Servitudes} § 1.1(1). A servitude can be an easement, profit, or covenant. \textit{Id.} § 1.1(2). The Restatement calls covenants that are servitudes “covenants running with the land.” \textit{Id.} § 1.3. Modern courts do not distinguish between equitable and real covenants. Lake Limerick Country Club v. Hunt Mfg. Homes, Inc., 84 P.3d 295, 298–99 (Wash. App. 2004). In this article, I use both “covenant” and “real covenant” to refer to covenants running with the land.

\textsuperscript{16} See Boyack, \textit{Community Collateral Damage}, supra note 2, at 60 (“All types of CICs . . . share the same essential service and payment structure: homeowner-elected directors manage common upkeep, and all homeowners contribute their pro rata portion of the common costs.”); see also \textsc{Hyatt \& French}, supra note 2, at 11 (discussing the power of an elected board of directors); \textsc{Hyatt}, supra note 4, at 84–88, 105, 121 (discussing powers of a board, community assessments, and collection devices).

\textsuperscript{17} \textsc{Hyatt \& French}, supra note 2, at 6, 13–14.

\textsuperscript{18} Any associated financial obligations are secured by a lien on the subject property. \textsc{Hyatt}, supra note 4, at 120–21.

\textsuperscript{19} CIC purposes are almost always defined as preserving and promoting property values and owner “lifestyle.” Apple II Condo. Ass’n v. Worth Bank & Trust Co., 659 N.E.2d 93, 95–97 (Ill. App. Ct. 1995); see also Ngai Pindell, \textit{Home Sweet Home? The Efficacy of Rental Restrictions to Promote Neighborhood Stability}, 29 St. Louis U. Pub. L. Rev. 41, 43–46 (2009);
In addition to recorded covenants contained in a community’s CC&Rs, the board of the community association can pass specific regulations authorized by the recorded declaration. These regulations can be changed as the board sees fit. CIC obligations can therefore arise either from the terms of the original recorded declaration, from amendments to the declaration, or from the rules promulgated by the board of directors to carry out the general purposes of the association. Courts generally are more deferential to recorded covenants than to rules enacted by the board, reasoning that owners had more notice of recorded covenants and that such covenants are not as easily changed. In addition, state statutes sometimes limit the ability of a board to promulgate rules governing individually owned property (as opposed to common elements) and individuals’ behavior.

Occasionally, public policy provides a substantive outer limit on restrictive covenants. For example, in a handful of cases, a non-compete covenant or a restriction on alienation has been declared unenforceable as contravening public policy. Aside from such outlier cases, however, courts today will generally enforce covenant obligations that have something to do with the


HYATT, supra note 4, at 82–88 (discussing the powers of a board of directors of a CIC association); see also Todd Brower, Communities Within the Community: Consent, Constitutionalism, and Other Failures of Legal Theory in Residential Associations, 7 J. LAND USE & ENVT’L. L. 203, 242 (1992) (noting that CIC enforcement is justified based on the unanimous assent of its members to covenant terms and explaining that later amendments “pose special problems”).


property as long as the obligations have been created by an intentional, recorded writing. This is different than in the past. Traditionally, in order for landowners to create a real covenant, the covenant must be in writing, specifically intended to run with the land, touch and concern the real property, be adequately publicized (usually by recordation in the applicable local land records in order to create third party notice), and be authored by parties who were linked in “horizontal privity.” Modernly, courts have moved away from strictly requiring these elements exist in order for a covenant to have been created. The newer approach relies on an intentional, recorded writing alone, focusing on upholding as a servitude any provision specifically intended to be a servitude. This approach dispenses with the formalistic requirement of privity and, to some extent, the touch and concern test.

24 E.g., Powell v. Washburn, 125 P.3d 373, 376 (Ariz. 2006); Vulcan Materials Co. v. Miller, 691 So. 2d 908, 913 (Miss. 1997); Runyon v. Paley, 416 S.E.2d 177 (N.C. 1992). The requirement that a covenant “touch and concern” the land requires that the substance of the covenant relate to the real property itself. By requiring that a covenant touch and concern the land in order to run with the land, the common law sought to ensure that personal obligations unrelated to the ownership of the property would only bind the original parties—in contract—and would not be deemed servitudes that would continue as specifically enforceable obligations for all landowners.

25 RESTATEMENT OF PROP. § 537 cmt. h (1944) (justifying the touch and concern requirement as a means to reduce the number of permissible real covenants). “Horizontal privity” requires both parties to simultaneously hold an interest in the same property, such as a landlord and tenant or buyer and seller. Neighbors, for example, would not be in horizontal privity.

Most of the requirements for covenant creation deal with required formalities, but the touch and concern requirement—to the extent it still exists—has to do with substance and limits the scope of perpetually restraining covenants.\textsuperscript{27} For example, traditionally, a promise to pay money could not be a covenant obligation as it was considered not to touch and concern the land. But courts eventually accepted that the assessment of property to pay for joint amenities was a proper subject matter for real covenants,\textsuperscript{28} and it was this expansion of the notion of touch and concern that spurred growth of suburban planned communities across the country.\textsuperscript{29} In the past several decades, the touch and concern requirement has faded in importance. The new Restatement calls it unnecessary. But without the touch and concern requirement, covenants have no substantive limits beyond the public policy restraints placed on all contracts.

The legality of CIC governance crystallized during the last century.\textsuperscript{30} But the outer boundary of permissible subject matter


\textsuperscript{28} See, e.g., Regency Homes Ass’n v. Egermayer, 498 N.W.2d 783 (Neb. 1993); Neponsit Property Owners Association, Inc. v. Emigrant Industrial Savings Bank, 278 N.Y. 248 (1930).


\textsuperscript{30} Although the modern CIC did not appear until the 1970s, the underlying legal forms that make CICs possible can be traced back to the sixteenth century’s breakdown of the English common field system. The Industrial Revolution heralded changes in land use that increased potential negative externalities on neighbors. Property law expanded the law of servitudes as an adaptation to these new developments. *Id.*; see also JESSE DUKEMINIER & JAMES E. KRIER, *PROPERTY* 668–70 (5th ed. 2002). Initially courts were worried that this ownership structure would negatively impact alienability. See, e.g., Hutchinson v. Ulrich, 34 N.E. 556 (Ill. 1893) (holding
for CIC regulation remains the subject of heated debate.\textsuperscript{31} As courts over the past century began to take a more permissive view toward CICs and associated covenant requirements, developers increasingly structured communities with common amenities and assessment obligation servitudes, confident that courts would uphold the governance scheme. In the twentieth century, community real estate development became a big part of the real estate industry. Developers pioneered using servitude law to achieve their visions of community planning and design. At first, developers relied on restrictive covenants to limit land uses as a way to preserve values, particularly for affluent suburban communities.\textsuperscript{32} Many early generation covenant communities were created by obtaining the unanimous consent of all neighborhood residents, and these covenants focused on restricting undesirable uses\textsuperscript{33} and users.\textsuperscript{34}

\begin{itemize}
\item that since limitations on free alienability are disfavored at law, ambiguities are to be resolved against the restrictive covenants; Carol M. Rose, Property Law and the Rise, Life, and Demise of Racially Restrictive Covenants (Ariz. Legal Studies Discussion Paper No.13-21, 2013), reprinted in POWELL ON REAL PROPERTY (Michael Allan Wolf & Richard R. Powell, eds., 2013), available at http://ssrn.com/abstract=2243028. Initially, courts were concerned that enforcing this new brand of servitude would adversely affect alienability of land. M\textsc{ckenzie}, supra note 29, at 32.
\item \textsuperscript{31} See, e.g., Uriel Reichman, \textit{Residential Private Governments: An Introductory Survey}, 43 U. Ch. L. Rev. 253, 293–94 (1976) (advocating a robust “touch and concern” test as a way of limiting the scope of permissible CIC regulations); Brower, \textit{supra} note 20, at 272–73 (advancing a theory that presumptive enforceability of CIC covenants should turn on the extent of the particular liberty right curtailed).
\item Neponsit, 15 N.E.2d at 793.
\item Community covenants are very useful in addressing negative external impacts that the use of one parcel imposes upon other proximate parcels, and are preferable to reliance on nuisance law to protect property from such negative externalities. See Andrea J. Boyack, \textit{Community Covenant Alienation Restraints and the Hazard of Unbounded Servitudes}, 42 Real Estate L.J. 450 (2013) [hereinafter Boyack, \textit{Community Covenant Alienation Restraints}].
\item “Occupancy restrictions perhaps were the \textit{raison d’être} of early-generation covenant-based communities.” \textit{Id.;} see also Grassmick, \textit{supra} note 2; \textsc{Lee Anne Fennell}, \textit{The Unbounded Home: Property Values Beyond Property Lines} 123 (2009). For a thorough discussion and analysis of historic racial occupancy restrictions in CICs, see \textsc{Richard R.W. Brooks} & \textsc{Carol M. Rose}, \textit{Saving the Neighborhood: Racially Restrictive}
In the 1960s and the 1970s, there was a further revolution in CIC ownership form through the increased use of condominiums. During this time, new developments increasingly were structured as privately governed communities prior to sale of the first unit, and in these communities owner assent was presumed through purchase of property already burdened with CC&Rs. Although the cooperative form had previously been used to approximate real property ownership of a unit in a multi-family building, in the 1960s, actual fee simple property ownership of apartment units was made possible by the enactment of condominium-enabling statutes. By the 1970s, every state had adopted a

---

35 The cooperative ownership structure allowed shareholders of an owning entity to obtain exclusive, perpetual possessory rights with respect to a single apartment unit. Cooperatives are generally included in the definition of CICs even though their ownership form is based on lease and corporate law. Cooperatives, often known as co-ops, are more commonly found in earlier urbanized areas, such as New York City. Susan Stellin, Co-op vs. Condo: The Differences Are Narrowing, N.Y. TIMES, Oct. 5, 2012, at RE9. Cooperative buildings do not permit fee simple ownership of a given unit, instead, the entire building is owned by an entity, and each “owner” holds a share of membership interest in the entity. The shareholders have, as an appurtenance to their ownership interests, a perpetual lease on “their” unit.

36 The Condominium is a creature of statute that permits fee simple ownership defined along three-dimensional planes, rather than common law two-dimensionally defined land ownership boundaries. In the common law, the third dimension is ad coelorum: a column of space “from the center of the earth to the heavens.” See William Schwartz, Condominium: A Hybrid Castle in the Sky, 44 B.U. L. Rev. 137, 141 (1964) (noting the traditional view that “whatever is attached to the land belongs to the land” and, consequently, to the person who owns the land itself); Charles W. Pittman, Note, Land Without Earth—The Condominium, 15 U. Fla. L. Rev. 203, 205–06 (1962) (noting the general hostility expressed in European civil codes to the concept of horizontal property). Condominium ownership is the only way to own an apartment in fee simple. The earliest state condominium statutes tracked the FHA Model Act and in some key aspects were insufficient, ambiguous and ineffective. See Robert Kratovil, The Declaration of Restrictions, Easements, Liens, and Covenants: An Overview of an Important Document, 22 J. Marshall L. Rev. 69 (1988). Once condominium-enabling statutes were passed in the early 1970s, condominium ownership of apartments rapidly replaced the cooperative form as the most common way to obtain “ownership” of an apartment unit. The condominium ownership structure made ownership
of urban apartment dwelling units possible and has proved so flexible that today fee simple ownership can exist with respect to “postage stamp” buildings (the outlines of the building alone without any surrounding land), parking spaces, interior store spaces, and even air space for telecommunications equipment.

37 Every state adopted a condominium statute in the 1960s, and this paved the way for a huge condominium “boom” during the next few decades. HYATT, supra note 4, at 11.

38 UCIOA, supra note 22. The UCIOA was created by combining the Uniform Condominium Act, the Uniform Planned Community Act, and the Model Real Estate Cooperative Act.

39 Boyack, Community Collateral Damage, supra note 2, at 100.

40 Id.

41 In condominium ownership, every member owns her unit in fee simple and all members collectively hold the remainder of the condominium (the roof, lobby, elevators, amenities, parking garage, electrical system, etc.) as tenants in common. PUD development is similar to condominiums, but typically the
lot owners do not own common areas as tenants in common; instead, the association owns the common areas. In all three forms of CICs, property ownership is synonymous with membership in the governing association, and in all three ownership forms, members must abide by recorded covenants and rules established by the association’s board. The association is responsible for maintenance of the CIC and is funded in full by assessments levied on the members. The obligation to pay assessments is secured by a lien on the real property owned by the member. See generally HYATT & FRENCH, supra note 2; Boyack, Community Collateral Damage, supra note 2.

42 See CLIFFORD TREÈSE ET AL., RESEARCH INST. FOR HOUS. AM., CHANGING PERSPECTIVES ON COMMUNITY ASSOCIATION MORTGAGE UNDERWRITING AND CREDIT ANALYSIS 6 (2001), available at http://www.housingamerica.org/RIHA/RIHA/Publications/48502_ChangingPerspectivesonCommunityAssociationMortgageUnderwriting.pdf (stating that government privatizes its functions, requiring community associations to fulfill an otherwise municipal obligation); see also HYATT & FRENCH, supra note 2, at 13–14 (explaining how CICs function like local governments); Boyack, Community Collateral Damage, supra note 2, at 121 (comparing the function of associations to that of local governments and comparing association assessments to property taxes).


44 TREÈSE ET AL., supra note 42, at 3 (discussing methods that communities utilize to minimize taxes); Boyack, Community Collateral Damage, supra note 2, at 60 (“The CIC structure enables more community
private CIC structuring reflects demands for lower property taxes. On the other hand, as municipalities push for CIC structuring, buyers who specifically would like to live outside a CIC may be unable to find non-CIC housing. In addition, owners in CICs effectively are taxed twice—once through municipal property taxes and once through CIC assessments.

On balance, the innovation of the CIC is a positive development. CIC structures have led to increased home ownership in the United States. CICs also address the problem of neighborhood nuisances and increase available neighborhood amenities.

Still, CIC jurisprudence shows troubling claims of amenities and upkeep, permitting neighborhoods to self-fund and allowing local governments to avoid raising taxes in response to more housing developments.

45 In California, Proposition 13 limited municipal ability to increase property taxes to meet demand for community services, and CIC governance was a way to provide community amenities without draining tax revenue. The trend away from property tax funded amenities is self-perpetuating because residents in CICs, who have to pay community assessments in addition to property taxes, are strong and local voting blocks against property tax increases. Callies & Suarez, supra note 14, at 493.

46 In one state, New Jersey, taxpayers have successfully claimed the right to offset a portion of their community assessments from property taxes, claiming that they were penalized by double taxation without this offset. HYATT, supra note 4, at 133 (citing Borough of Englewood Cliffs v. Estate of Allison, 174 A.2d 631, 640 (N.J. Super Ct. 1961) (reasoning that a property’s true value does not include the value of rights transferred to a community)). Other than in New Jersey, however, assessments are not deductible from tax impositions. Id. at 106.

47 In situations where neighbors do not have community covenants, or where covenants do not explicitly prohibit an objectionable activity, neighbors can claim that the objectionable activity should be proscribed as a nuisance. Relying on the tort of nuisance to prohibit uses of neighboring property, however, is unpredictable, inconsistent, and often ineffective. For example, In Turudic v. Stephens, an Oregon court found that keeping two “pet” cougars in a residential neighborhood did not constitute a nuisance. 31 P.3d 465 (Or. Ct. App. 2001). On the other hand, courts routinely limit uses of property based on restrictions in a community’s CC&Rs without requiring that the use be proven to be a nuisance. See, e.g., Laumbauch v. Westgate, C.A. No. 2442-VCS, 2008 WL 3846419 (Del. Ch. Aug. 19, 2008), aff’d, 966 A.2d 349 (Del. 2009).

48 See Boyack, Community Covenant Alienation Restraints, supra note
overreaching by association governments and the enforcement of abusive covenants. Some scholars bemoan the erosion of personal freedom, property rights, and neighborhood diversity that has resulted from the proliferation of the CIC ownership model.

B. Judicial Oversight of CIC Governance

The CIC phenomenon is impacted by an array of legal disciplines, including association governance, constitutional rights, and property law. But more and more, courts have conceived of CIC governing provisions under the rubric of

33; Lee Anne Fennell, Contracting Communities, 2004 U. Ill. L. Rev. 829; French, supra note 43.


50 E.g., McKENZIE, supra note 29; Paula A. Franzese & Steven Siegel, Trust and Community: The Common Interest Community as Metaphor and Paradox, 72 Mo. L. Rev. 1111 (2007).
contract jurisprudence. The rhetoric of freedom of contract is often used as the primary justification for upholding CIC regulations and restrictions.\textsuperscript{51} The reality of how parties become obligated to CIC covenants and board-enacted rules, however, calls into question just how appropriate and far-reaching freedom of contract rationale is in the CIC context.

Courts have struggled with the best way to characterize CIC covenants and rules, but for all courts, the analysis of CIC regulation validity starts with the foundational assumption that owners voluntarily obligate themselves to CIC governance when they buy into the community. Based on this presumption, courts explain that owners voluntarily agreed to relinquish “a certain degree of freedom of choice” when they became members of the CIC.\textsuperscript{52} Therefore, the covenants and properly enacted rules are presumptively binding as contract terms. In particular, provisions of the recorded declaration as of the date of an owner’s purchase are presumptively binding unless the provisions violate public policy.\textsuperscript{53}

Theoretically, a court may strike down CIC covenants based on finding that they infringe upon members’ “constitutional rights.”\textsuperscript{54} But constitutional violations must involve state action, and this is a difficult hurdle to overcome in the context of CIC associations.\textsuperscript{55} Sometimes disgruntled owners claim that an


\textsuperscript{52} HYATT, supra note 4, at 50–51 (explaining how widely cited this foundational assumption is); see also Basso, 393 So. 2d at 637.

\textsuperscript{53} Public policy limits the substance of covenants in the same way that public policy limits the substance of contracts. For example, some covenants not to compete have been held unenforceable as a matter of public policy. Davidson Bros., Inc. v. Katz & Sons, Inc., 579 A.2d 288 (N.J. 1990). Theoretically, public policy should also restrain covenants that unduly limit alienation of real property. See, e.g., Riste v. E. Wash. Bible Camp, Inc., 605 P.2d 1294 (Wash Ct. App. 1980).

\textsuperscript{54} See HYATT, supra note 4, at 62–63. The standard for review is whether any category one restriction is wholly arbitrary, in violation of public policy or an individual’s constitutional rights. Pines of Boca Barwood Condo. Ass’n v. Cavouti, 605 So. 2d 984, 985 (Fla. Dist. Ct. App. 1992).

\textsuperscript{55} There must be “state action” to enforce constitutional rights. Comm.
association’s power is restrained by state or federal constitutions based on an expansive conception of state action. For example, one theory—made in reference to the 1944 Supreme Court case of *Shelley v. Kraemer*—is that even if acts by an association are not themselves state action, state action exists when a court enforces such governance acts, and it is this judicial state action that renders the covenant’s substance vulnerable to constitutional scrutiny.\(^{56}\) Most courts, however, decline to apply *Shelley* outside the private racial zoning context.\(^{57}\) Another theory, made in reference to the 1946 Supreme Court case of *Marsh v. Alabama*, posits that CICs are the functional equivalent of local governments and should therefore be bound to the same constitutional constraints.\(^{58}\) However, this theory has not gained widespread support, perhaps because today’s CICs do not completely replace local public governments in the same way that a company town did in the time of *Marsh*.\(^{59}\)

Both of these theories

---

For A Better Twin Rivers v. Twin Rivers Homeowners’ Ass’n, 929 A.2d 1060, 1067 (N.J. 2007); see also HYATT, supra note 4, at 62–63. For example, one court specifically explained that a covenant limiting occupancy that would violate constitutional rights if created by the local government through a zoning ordinance did not create a constitutional problem because it was privately enacted. See White Egret Condo., Inc. v. Franklin, 379 So. 2d 346, 349 (Fla. 1979).


\(^{59}\) HYATT, supra note 4, at 64–65; see also, e.g., *Goldberg v. 400 E. Ohio Condo. Ass’n*, 12 F. Supp. 2d 820 (N.D. Ill. 1998) (“Demonstrating that condominiums do certain things that state governments also do doesn’t show that condominiums are acting as the state or in the state’s place.”). The holding in *Marsh* has been applied to cases having to do with public accommodation and access. *Id.; see also Amalgamated Food Emp. Union, Local 590 v. Logan Valley Plaza, Inc.*, 391 U.S. 308 (1968), *abrogated by Hudgens v. NLRB*, 424 U.S. 507 (1976). For an interesting discussion regarding the extent to which CICs function as municipal governments with
have generally been rejected by courts. Today, aside from Fair Housing Act prohibitions of sale transfer restrictions that are based on a constitutionally protected classification (race, religion, etc.), and the odd outlier decision, the U.S. Constitution apparently does not provide any substantive oversight of common interest community covenants.

A few courts have been willing to invalidate CIC governing acts on the basis of state constitutional violations. Cases where state constitutional guaranties have been applied to CIC governance mostly deal with freedom of speech and rights of access. But other constitutional challenges abound. For respect to non-members, see David J. Kennedy, Residential Associations as State Actors: Regulating the Impact of Gated Communities on Nonmembers, 105 YALE L.J. 761 (1995).

60 HYATT, supra note 4, at 67. See, e.g., Pines of Boca Barwood Condo. Ass’n v. Cavouti, 605 So. 2d 984, 985 (Fla. Dist. Ct. App. 1992); White Egret, 379 So.3d at 349.

61 Fair Housing Act, 42 U.S.C. §§ 3601–19 (2012). The Act, as amended, prohibits discrimination in the sale, rental, and financing of dwellings, and in other housing-related transactions, based on race, color, national origin, religion, sex, familial status (including children under the age of 18 living with parents or legal custodians, pregnant women, and people securing custody of children under the age of 18), and disability. Id. § 3604.

62 For example, in Gerber v. Longboat Condominium, a veteran’s right to fly the American flag in violation of CIC covenants was upheld by the court striking down the covenant prohibition as a violation of the Constitution. 724 F. Supp. 884 (M.D. Fla. 1989), aff’d in part on reh’g, Gerber g. Longboat Condominum, 757 F. Supp. 1339, 1341 (M.D. Fla. 1991). A different court criticized the Gerber decision as being based on emotion, not on law. Goldberg v. 400 East Ohio Condo., 12 F. Supp. 820 (N.D. Ill. 1998).


example, recent cases dealing with both state and federal constitutional claims have raised the issue of whether freedom of religion guaranties can prohibit CIC regulation of placement of a mezuzah on a doorframe or painting a kolam on a driveway. State action problems also plague state constitutional claims in the CIC context, and the law in this area is muddled and inconsistent.

---

65 A mezuzah is a small container holding handwritten parchment with a scriptorial passage that is affixed to the entranceway to a home by devout Jews. The Seventh Circuit in *Bloch v. Frischholz*, 587 F.3d 771 (7th Cir. 2009), held that a CIC rule prohibiting “objects of any sort” outside a resident’s door was neutral as to religion and therefore reasonable and enforceable. It is common for CICs to restrict changes to the exterior of homes without association permission. See Angela C. Carmella, *Religion-Free Environments in Common Interest Communities*, 38 Pepp. L. Rev. 57, 68 (2010) (discussing “aesthetic controls on signs, symbols, decorations, statuary, or items of any kind”).


67 Wayne S. Hyatt, *Common Interest Communities: Evolution and Reinvention*, 31 J. Marshall L. Rev. 303 (1998) [hereinafter Hyatt, *Common Interest Communities*] (explaining that the property application of constitutional principles to CIC governance is an unsettled area of the law). There are occasional cases that test the application of constitutional protections to CIC governance actions, and the most emotionally charged cases do much to muddy the jurisprudence in this area. An example is *Gerber v. Longboat Condominium*, in which a CIC denied a veteran’s right to fly an American flag. The court found that this act violated the owner’s constitutional rights. 724 F. Supp. 884 (M.D. Fla. 1989), aff’d in part on reh’g, Gerber v. Longboat Condo., 757 F. Supp. 1339, 1341 (M.D. Fla. 1991). During the post-9/11 patriotic fervor, Congress felt compelled to pass a law guaranteeing the right of homeowners to fly Old Glory. The Freedom to Display the American Flag Act, codified at 4 U.S.C.A. § 5 (2012), prohibits a CIC from adopting or enforcing any policy that would unreasonably restrict or prevent a member of the association from displaying the flag of the United States. See Robin Miller, Annotation, *Restrictive Covenants or Homeowners’ Association Regulations Restricting or Prohibiting Flags, Signage, or the Like on...*
Unless proven to be “arbitrary, against public policy or violat[e]ive of] some fundamental constitutional right of unit owners,” covenants contained in a CIC’s original declaration are presumptively valid.\(^68\) Public policy and constitutional constraints on the substance of CIC covenants are quite limited, and in the vast majority of cases, covenants are upheld. Courts and scholars reason that “[t]he initial members of a homeowners association, by their voluntary acts of joining, unanimously consent to the provisions in the association’s original governing documents.”\(^69\) Covenant amendments or rules enacted by the board of directors, however, are subject to \(slightly\) more judicial oversight, although the proper standard of review for such association or board actions is subject to some debate.\(^70\) Some courts use the Business Judgment Rule, borrowed from corporate law,\(^71\) in order to assess the validity of CIC governing acts.\(^72\) Other courts claim that CIC amendments and rules must be “reasonable” in order to be valid.\(^73\) And some jurisdictions use both tests: the more permissive Business Judgment Rule when associations are performing “business responsibilities” and the slightly less deferential rule of reasonableness when associations are engaging

---


\(^{69}\) Ellickson, \(supra\) note 49, at 1526–27.

\(^{70}\) Hyatt, \(supra\) note 4, at 89–97.


\(^{73}\) See, e.g., Villa De Las Palmas Homeowners Ass’n v. Terifaj, 90 P.3d 1223, 1234 (Cal. 2004); Noble v. Murphy, 612 N.E.2d 266, 270 (Mass. App. Ct. 1993).
in community “governance.” The problem with this approach is that it is difficult to determine when an association is acting as a business and when it is acting as a government, since “there is no bright line between the two” roles.

The Business Judgment Rule is a deliberately deferential standard of review. Under this standard, “absent a showing of fraud, dishonesty, or incompetence, it is not the court’s job to second-guess the actions of directors.” According to the Business Judgment Rule, if a business decision is made in good faith based on an honestly held rational belief that the decision is in the best interest of the entity, courts will not critique the decision. When applying the rule of reason, on the other hand, courts purport to balance the benefit of a particular governing act against its cost. In reality, however, courts do not engage in any precise cost-benefit analysis, and simply consider generally whether the particular governing act pertains to “the health, happiness and peace of mind of the unit owners.” The burden is on a complaining homeowner to prove a lack of nexus, and that a CIC governing act is therefore unreasonable.

Many scholars and judges conclude that this hands-off approach is appropriate because of freedom of contract. These commentators opine that there should be no real substantive judicial oversight of CIC governing acts and provisions. The proper role for a court, under this formulation, is to ensure the good faith of the decision-makers and the integrity of the process.

---

74 Hyatt, supra note 4, at 89.
75 Id.
76 Schwarzmann, 655 P.2d at 1181.
79 Hyatt, supra note 4, at 88–97.
80 See e.g., Richard A. Epstein, Covenants and Constitutions, 73 Cornell L. Rev. 906, 920 (1988) [hereinafter Epstein, Covenants and Constitutions].
Other scholars justify a more robust judicial review of covenant amendments and rule-making, asserting that courts have the power to make a substantive inquiry as to whether an association is acting within its scope of authority and whether the action bears a rational relationship to legitimate purposes of the CIC.81

In the corporate context, the Business Judgment Rule is justified based on judicial policy of leaving business decisions to the business experts. In the context of CIC governance, the decision-makers are volunteer laypeople, not corporate executives.82 Nevertheless, the several courts that have embraced the Business Judgment Rule standard to review CIC governance have failed to note this difference in context. For example, New York’s Superior Court, in Levandusky v. One Fifth Ave. Apartment Corp., acknowledged that “[e]ven when the governing board acts within the scope of its authority, some check on its potential powers to regulate residents’ conduct, life-style and property rights is necessary,” but then it concluded that the Business Judgment Rule is the most appropriate standard of review to achieve that “check” on association power.83 According to the Levandusky court, adopting the Business Judgment Rule means that judges should not inquire into actions taken in good faith “in the lawful and legitimate furtherance of corporate purposes.”84 California agreed with New York’s Levandusky opinion and adopted the Business Judgment Rule approach to CIC governance in Lamden v. LaJolla Shores Clubdominium Homeowners Ass’n.85 Thus, in at least the two of the most populous states, CIC governance decisions are unconstrained by any substantive judicial oversight.

82 CICs are really not corporations in the traditional sense. For example, they are not staffed by professional corporate directors and there are no disinterested directors. HYATT, supra note 4, at 90.
84 Id. The court specifically rejected the reasonableness standard adopted by the appellate court.
Other jurisdictions purport to apply the rule of reason in assessing the validity of CIC governance. The courts’ use of reasonable review theoretically includes an element of subjective review, but in practice, reasonableness review of CIC actions focuses almost exclusively on whether the association followed the enumerated procedures in amending the CC&Rs or passing community rules.\(^8\) Although most courts assert that only “reasonable” governing acts will be upheld, courts rarely explain what this standard means or engage in any methodical balancing of equities.\(^7\) In many cases, courts have essentially defined reasonable to include anything that could possibly promote community purposes, typically defined as preserving and improving property values and owner “lifestyle.”\(^8\) In circular logic, some courts give the board of the CIC association the discretion to determine which of its governing acts are “reasonable.”\(^9\) Meanwhile, other courts claim to require reasonableness but instead actually apply the Business Judgment Rule standard of review.\(^9\)

Considering the actual approach that most courts use in

\(^8\) See Hyatt, Common Interest Communities, supra note 67, at 354.

\(^7\) Id. Robert C. Ellickson opined that “reasonableness” in CIC jurisprudence means different things to different courts. Ellickson opposed “reasonableness” review in the name of freedom of contract. He stated:

“Reasonable,” the most ubiquitous legal adjective, is not self-defining. In reviewing an association’s legislative or administrative decisions, many judges have viewed the “reasonableness” standard as entitling them to undertake an independent cost-benefit analysis of the decision under review and to invalidate association decisions that are not cost-justified by general societal standards. This variant of reasonableness review ignores the contractarian underpinnings of the private association.

Ellickson, supra note 49, at 1530.


analyzing the validity of CIC amendments and rulemaking, the
distinction between the various purported standards of review
blurs. Although different jurisdictions purport to adopt distinct
oversight standards, in effect, most courts approach this issue in
essentially the same way: original covenants are presumptively
valid, and covenant amendments and rules adopted in accordance
with the procedures enumerated in the declaration are also valid
unless they are arbitrary or promulgated in bad faith.\textsuperscript{91} In \textit{Lieber v. Point Loma Tennis Club}, for example, the court held that a
regulation is deemed “reasonable” if it is not arbitrary and there
are valid reasons that an association might choose to enact the
rule.\textsuperscript{92} This standard is not a cost-benefit balancing test, but
rather mirrors oversight in administrative law, upholding rules
duly enacted as long as they are not arbitrary and capricious.\textsuperscript{93}
Regardless of standard used, courts almost universally uphold and
enforce CIC covenants and regulations.

II. THE COVENANT—CONTRACT MISMATCH

A. Adhesion and “Assent”

If contracts are not voluntary, the liberty and efficiency
justifications for their enforcement evaporate. In the context of
standard form and adhesion contracting, the voluntariness
associated with freedom of contract is diminished.\textsuperscript{94} Nevertheless,

\textsuperscript{91} HYATT, \textit{supra} note 4, at 56–57.

\textsuperscript{92} Lieber v. Point Loma Tennis Club, 47 Cal. Rptr. 2d 783, 788–89 (Cal. Ct. App. 1995) (finding that if it is not arbitrary, meaning there are valid reasons that an association might choose a regulation, it is “reasonable”).

review of CIC covenants has inspired calls for a return to a robust “touch and concern” test as a way of reigning in CICs. \textit{See, e.g.}, Reichman, \textit{supra} note 31, at 293–94.

\textsuperscript{94} In adhesion contracts, “[a]ctual assent is not just a fiction because of voluntary choices by consumers; it is effectively impossible.” Alan M. White & Cathy Lesser Mansfield, \textit{Literacy and Contract}, 13 STAN. L. & POL’Y REV. 233, 242 (2002); \textit{see also} Batya Goodman, Note, \textit{Honey, I Shrink-Wrapped the Consumer: The Shrink-Wrap Agreement as an Adhesion Contract}, 21
contracting pursuant to a non-negotiable standard form, particularly in agreements between parties of disparate bargaining power, is an increasingly common facet of modern reality, and courts have uniformly upheld the enforceability of adhesion contracts absent some special circumstance. Nevertheless, it is important to recognize that contractual theory imperfectly fits with the reality of non-negotiable forms. Standard, boilerplate terms are rarely read or negotiated. The resulting contractual


Modern contractual theory is based on objective manifestation of assent rather than subjective “meeting of the minds.” An indication of assent such as clicking “I accept” to posted terms or by initialing a form contract is clearly sufficient for legally binding obligation. Russell A. Hakes, *Focusing on the Realities of the Contracting Process—an Essential Step to Achieve Justice in Contract Enforcement*, 12 Del. L. Rev. 95, 99–100 (2011).

Several scholars have articulated the problematic disconnect between freedom of contract rhetoric and theory and the realities of the contracting process in the context of standard, non-negotiable forms. *E.g.*, Hakes, supra note 96, at 96.

substance therefore represents one party’s demands and the other’s acquiescence rather than jointly determined content. Although enforceable, the terms of such a contract do not necessarily reflect mutual intent. And when a contract’s terms are not actually elected by both parties, the contract does not necessarily promote efficient outcomes or create wealth.

Adhesion contracts are enforceable, but legal theory has


evolved to take into account the lack of voluntariness and content input inherent in adhesion contexts through modern doctrines such as unconscionability\(^{101}\) and distinct approaches to interpretation for adhesion contracts.\(^{102}\) Courts recognize that traditional deference to contractual terms may be inappropriate for contracts of adhesion, and they therefore sometimes monitor the substantive fairness of a contract in an adhesion contract context.\(^{103}\) This paternalistic approach diverges markedly from traditional hands-off contract enforcement and has led some observers to opine that contract law is now evolving along two tracks: a traditional assessment of process-based oversight for agreements between equally situated parties and a protective, regulatory approach with respect to “unsophisticated parties” in contracts of adhesion.\(^{104}\)

\(^{101}\) See Williams v. Walker-Thomas Furniture Co., 350 F.2d 445, 449 (D.C. Cir. 1965) (“Unconscionability has generally been recognized to include an absence of meaningful choice on the part of one of the parties together with contract terms which are unreasonably favorable to the other party.”).

\(^{102}\) When a written contract has been drafted solely by one party, courts invoke the doctrine of contra proferentum (“against the offeror”) that “requires that ambiguity in non-negotiated or adhesion contracts to be construed against the profferer.” Karnette v. Wolpoff & Abramson, L.L.P., 444 F. Supp. 2d 640, 647 (E.D. Va. 2006). In the context of adhesion contracts, courts sometimes construe a contract “to effectuate the reasonable expectations of the average member of the public who accepts it.” RESTATEMENT (SECOND) OF CONTRACTS § 237 cmt. E (1981); see also C & J Fertilizer, Inc. v. Allied Mut. Ins. Co., 227 N.W.2d 169, 172 (Iowa 1975).

\(^{103}\) See Stelluti v. Casapenn Enters., LLC, 975 A.2d 494, 502 (N.J. Super. Ct. App. Div. 2009), aff’d, 1 A.3d 678 (N.J. 2010) (explaining that in adhesion contracts, a court should consider the “substantive contents of the agreement” as well as the process that led to its execution); C & J Fertilizer, 227 N.W.2d at 174–75 (explaining that the court is responsible for exercising oversight with respect to the fairness and content of terms in a contract of adhesion). Professor Rakoff advocates that adhesion contracts be considered presumptively unenforceable. See Rakoff, Contracts of Adhesion, supra note 95, at 1176.

\(^{104}\) Meredith R. Miller, Contract Law, Party Sophistication and the New Formalism, 75 Mo. L. Rev. 493 (2010); Alan Schwartz & Robert E. Scott, Contract Theory and the Limits of Contract Law, 113 Yale L.J. 541, 545 (2003); see also L & L Wings, Inc. v. Marco-Destin, Inc., 756 F. Supp. 2d 359 (S.D.N.Y. 2010) (holding that party sophistication and bargaining power should be a factor to consider in determining whether a liquidated damages
Contract theorists justify the enforceability of contracts of adhesion with reference to market forces that will act to monitor and constrain the content of such contracts. But market checks only work when the market provides choices. It is increasingly true that in many areas of the country, most home purchase options are in CICs. Shopping around among various CICs offers no real choice either: most CIC declarations are virtually identical. In this context, market forces cannot justify the provision is enforceable. This latter approach has more in common with the European policy of prospectively approving the substance of form contracts prior to enforcement. See Leone Niglia, THE TRANSFORMATION OF CONTRACT IN EUROPE (2003) (explaining how contract law in Europe has evolved to deal with standard form contracts).


106 Franzese & Siegel, supra note 50, at 1113–14; Steven Siegel, THE CONSTITUTION AND PRIVATE GOVERNMENT: TOWARD THE RECOGNITION OF CONSTITUTIONAL RIGHTS IN PRIVATE RESIDENTIAL COMMUNITIES FIFTY YEARS AFTER MARSH v. ALABAMA, 6 WM. & MARY BILL RTS. J. 461, 469 (1998) [hereinafter Siegel, THE CONSTITUTION AND PRIVATE GOVERNMENT]; see also Robert Jay Dilger, NEIGHBORHOOD POLITICS: RESIDENTIAL COMMUNITY ASSOCIATIONS IN AMERICAN GOVERNANCE 38 (1992) (“Although CICs do provide more consumer options in the abstract, in many areas of the country [association-related housing] now dominate[s] the local housing market and [is] increasingly offering fairly uniform levels and types of services.”); Joel Garreau, EDGE CITY: LIFE ON THE NEW FRONTIER 189 (1991) (“If you want a new home, it is increasingly difficult to get one that doesn’t come with a homeowners’ association.”); Boyack, COMMUNITY COLLATERAL DAMAGE, supra note 2, at 59 (“The states with recent growth booms . . . have the highest percentage of citizens residing in privately governed CICs.”).

107 See Franzese & Siegel, supra note 50, at 1113–14 (“There exists no meaningful consumer choice amongst CIC organizational structures. In general, developer-imposed CIC templates are remarkably uniform.”). Even if buyers could shop around based on the particular provisions of a given CIC regime, this would be unlikely. Buyers often do not see the CIC declaration and associated documents until at or close to closing, and at closing, disclosure requirements mandate that a tremendous amount of paperwork is given to buyers. The sheer volume provided minimizes the likelihood that the buyer will review or understand the disclosures. Note, JUDICIAL REVIEW OF CONDOMINIUM RULEMAKING, 94 HARV. L. REV. 647, 650 (1981).
content of a contract of adhesion.

CIC declarations clearly fit the definition of an adhesion contract. Terms of a declaration are completely non-negotiable; in fact, prior to contracting they are prescribed and recorded in the land records. In addition, because one form binds multiple parties, no party has the ability to diverge from the recorded provisions. It is a perfect example of “take-it-or-leave-it” contracting.

Furthermore, CIC covenants are bundled with a real estate purchase. If a would-be buyer does not agree to the terms, she must relinquish the right to buy that property. Since each parcel of real property is presumed unique in our legal system, a buyer who forgoes a particular purchase has no true substitute. Homebuyers consider numerous factors in choosing which parcel of real property to buy, including school districts, lot size and configuration, tax assessment and appraisal, quality of construction, and even such things as the smell of the home and the orientation and exposure to natural light. The content of

---

108 C & J Fertilizer, Inc. v. Allied Mut. Ins. Co., 227 N.W.2d 169, 174-75 (Iowa 1975). Professor Rakoff has enumerated seven criteria that indicate a contract of adhesion: standard form drafted by one party who engages in repeated transactions of the sort presented as non-negotiable to the adhering party who enters into relatively few transactions of the sort, signed by the adhering party, and principally obligates the adhering party to pay money. Rakoff, Contracts of Adhesion, supra note 95, at 1177.

109 See Winokur, supra note 105, at 33 (concluding that such “built-in, substantive limitations on modification of uniform servitude forms present obstacles to market discipline by marginal consumers”).


111 The Department of Housing and Urban Development has even promulgated a homebuyer checklist to help purchasers track important aspects of properties they may buy. While extensive, the checklist does not explicitly discuss the scope or content of CIC governing provisions, although it does bring up “pet restrictions” as a line item for consideration. Aside from pet restrictions, however, the only reference to neighborhood covenants is a line item as to whether they are “good, average or poor” (whatever that means). For more information on the HUD homebuyer checklist and related documents, see Buying a Home, U.S. DEP’T OF HOUSING & URBAN DEV., http://portal.hud.gov/hudportal/HUD?src=/topics/buying_a_home (last visited...
CIC covenants and rules is likely not even a factor considered prior to purchase or, if considered, is a fairly unimportant detail in the home purchase calculus.

UCIOA and statutes in virtually every state mandate that a seller of real property disclose the details of a private governance regime prior to or at the closing of a real estate purchase. However the delivery of pages upon pages of legalese at or shortly before closing may do little to actually inform a buyer. Typically, homebuyers are not represented by counsel in home purchase negotiations, and legal counsel conducting real estate closings do not generally undertake to review and advise the buyer with respect to CIC obligations. Under these


112 E.g., UCIOA, supra note 22, § 4; COLO. REV. STAT. ANN. § 38-35.7-102 (West 2013); FLA. STAT. ANN. § 720.401 (West 2013); HAW. REV. STAT. ANN. § 508D-3.5 (West 2013); N.C. GEN. STAT. ANN. § 47E-4 (West 2013).

113 The quantity of disclosures made in connection with a real estate purchase diminishes the ability of the disclosures to truly inform. See, e.g., Melvin Aron Eisenberg, Comment, Text Anxiety, 59 S. CAL. L. REV. 305 (1986) (finding that “consumers who are faced with the dense text of form contracts characteristically respond by refusing to read”). Timing of disclosure in real estate conveyancing—in particular, disclosures made after a buyer has made an offer on a home—diminishes disclosure effectiveness as well. Stephanie Stern, Temporal Dynamics of Disclosure: The Example of Residential Real Estate Conveyancing, 2005 UTAH L. REV. 57. Recent studies of home mortgagors found that these buyers misapprehend or fail to read even the most basic parts of mortgage loan disclosure forms. Debra Pogrund Stark, et al., Ineffective in Any Form: How Confirmation Bias and Distractions Undermine Improved Home-Loan Disclosures, 122 YALE L.J. ONLINE 377, 379 (2013) (explaining that studies of consumers show that they have “miss[ed] the critical information that disclosure forms were designed to communicate”).

114 Most homeowners do not employ counsel to represent them in the conveyancing transaction. Debra Pogrund Stark et. al., Dysfunctional Contracts and the Laws and Practices That Enable Them: An Empirical Analysis, 46 IND. L. REV. 797, 801 (2013). Nor do buyers typically even have a realtor representing their interests because the agent working with a buyer is legally a seller’s sub-agent. The agent that works with the buyer is, in fact, often a seller’s subagent. Ann Morales Olazabal, Redefining Realtor Relationships and Responsibilities: The Failure of State Regulatory Responses, 40 HARV. J. ON LEGIS. 65, 66 (2003).

115 Lawyers who conduct residential real estate closings typically prepare
circumstances, it is highly unlikely that a buyer reads or understands CC&Rs prior to closing.

Finally, assent to the CIC terms does not even require a specific manifestation of acceptance thereof; rather, a party is deemed to have agreed simply by buying the land. Although this is true for any servitude, it is not the general rule for contract law, where a voluntary act manifesting intent to be bound is requisite to obligation. This simple fact further divorces true assent from legal obligation in the context of CIC covenants.

B. Servitude Damages and Duration

In the name of liberty and market freedom, our legal system generally eschews perpetual obligation and permits individuals to elect to walk away from their commitments (after payment of appropriate damages), and obtain a “clean slate.” The law reasonably protects a contracting party’s future autonomy from inescapable restraint by allowing exit via breach and

the deed and coordinate with the title company and mortgage lender, if applicable, with respect to recordation. Such counsel facilitates the closing, but does not actually advise the buyer or assist buyer in reviewing disclosure documents. Gary D. Beelen, Odds Are, It’s Not “Your” Closing Attorney, 21 DREW ECKL & FARNHAM, LLP J., no. 126, 2009, at 1, 1–5, available at http://www.deflaw.com/articles/odds-are-its-not-your-closing-attorney.


reimbursement of the non-breaching party’s expectation interest in nearly all cases.\textsuperscript{118} Although breach typically provides an exit from perpetual contract obligation, when contracts take on an \textit{in rem} character, attaching to real property as servitudes, that exit closes. Servitudes are generally enforced through “propriety” rules,\textsuperscript{119} meaning that the default remedy is specific performance.\textsuperscript{120} When an obligation is specifically enforceable, a party cannot opt out of the continuing affirmative requirement to comply.

Servitudes depart from contract law in another key aspect that impacts individual liberty—their potentially infinite duration. A servitude obligation—unlike a contract—presumptively exists in perpetuity, binding against current and future owners of the land, and cannot be terminated through breach.\textsuperscript{121} For servitudes,

\begin{itemize}
\item \textsuperscript{120} Winokur, \textit{supra} note 105, at 37 (“[T]he general availability of specific performance as a remedial alternative to damages precludes an owner’s unilateral election to breach the servitude and pay damages.”). Issuing a mandatory injunction is the typical way that restrictive covenants are enforced. See, e.g., Depeyster v. Town of Santa Claus, 729 N.E.2d 183, 190 (Ind. Ct. App. 2000); Metzner v. Wojdyla, 886 P.2d 154 (Wash. 1994). Servitude law draws a distinction between specifically enforceable equitable servitudes and real covenants that are enforceable through a grant of money damage, but this is a distinction without a difference. A given covenant-based servitude can be the subject of an action either in equity or in law at a plaintiff’s election, and it is easier to prove equitable grounds for recovery. See Runyon v. Paley, 416 S.E.2d 177, 182–83 (N.C. 1992); James L. Winokur et al., \textit{Property and Lawyerering} 642–43 (2002); Alfred L. Brophy, \textit{Contemplating When Equitable Servitudes Run with the Land}, 46 St. Louis U. L.J. 691, 698 (2002).
\item \textsuperscript{121} See Citizens for Covenant Compliance v. Anderson, 906 P.2d 1314,
contracting decisions today limit not only the contracting parties’ own future freedom but also the freedom of future generations of property owners. Problems of dead-hand control are thus endemic to covenants that run with the land. Under the common law, however, courts are generally empowered to strike down covenants that unduly restrain alienation on the basis of public policy. First-generation CICs created before widespread common interest ownership statutes were enacted in the 1970s and 1980s were cognizant of the common law’s hostility toward perpetual restrictions on land and contained expiration dates. Today, statutes in every state explicitly or implicitly authorize CIC ownership structures, and courts routinely uphold CIC covenants even without effective temporal limits. Because CIC


CIC covenants can be modified through supermajority vote of community members, but it is both cumbersome and practically difficult to amend CIC declarations.


See Welshire, Inc. v. Harbison, 91 A.2d 404 (Del. Ch. 1952) (30 years); Van Sant v. Rose, 103 N.E. 194 (Ill. 1913) (43 years); Easton v. Carebybrook Co., 123 A.2d 342 (Md. Ct. App. 1956) (8-year initial term, then continued until modification by vote of majority of owners).

Typically, CIC restrictions provide for automatic renewal after a given
Covenants have a virtually unlimited duration, their impact and effect is more expansive than contract law. Without durational restraints, substantive limitations are more justifiable. A CIC covenant that has an expansive or troubling scope—one that ties up land alienability or impacts personal freedoms, for example—will not eventually just disappear. If courts lack the tools to constrain the subject matter of covenants, it may be impossible to nullify the legal impact of such covenants, even if the covenant eventually contradicts the values of society as a whole or the impacted neighborhood in general.

Servitudes come in several flavors and have different, and evolving, legal formation requirements. Modernly, servitudes are generally grouped into easements and covenants. The variant closest to a conveyance is the easement—a right to make beneficial use of another’s land. Covenants running with the land, on the other hand, are closer in form and substance to contracts among neighbors, although of unlimited duration and specifically enforceable. Drawing the line between contracts that bind only the parties thereto and covenants that run with the land, thus binding on future owners is maddeningly difficult.

Initial term. Under the law of Louisiana, however, restrictions imposing affirmative obligations cannot exist in perpetuity. Diefenthal v. Longue Vue Found., 865 So. 2d 863, 882 (La. Ct. App. 2004), writ denied, 869 So. 2d 883 (La.).

Restatement (Third) of Prop.: Servitudes § 1.3 (2000) no longer uses the terms “real covenant” and “equitable servitude” to distinguish between types of covenants. Instead, the Restatement calls both covenants created in writing and enforceable at law and a servitude implied in equity “covenants.”

Restatement (Third) of Prop.: Servitudes § 1.2 (2000) defines “easement” as a “nonpossessory right to enter and use land in the possession of another and obligates the possessor not to interfere with the uses authorized by the easement.” A “profit” is closely related to an easement, except that it additionally gives the beneficiary the right to extract something from the burdened land. Id.

The current Restatement of Property departs from the use of the terms “real covenant” and “equitable servitude,” to refer to contracts that run with the land and therefore take on the character of property. Id. § 1.4.

The law of servitude formation has been progressing from a more formalistic approach that demanded strict adherence to formal requirements of privity and property relevance (the so-called “touch and concern” requirement) toward a more liberalized approach such as that advocated by the Restatement (Third) of Property.\textsuperscript{131} Under the Third Restatement’s approach, anything that a valid contract can achieve can now be achieved in perpetuity by a covenant. This approach offers nothing to constrain the content of covenants aside from public policy limits that apply to contracts generally. Once, the touch and concern rule for valid formation of real covenants operated to limit the scope of perpetually restraining covenants,\textsuperscript{132} but as this requirement has been watered down and in some cases (per the new Restatement’s approach) eliminated, little substantive control remains with respect to what types of obligations real covenants can impose.\textsuperscript{133} Some courts purport to limit covenant enforcement to obligations that are “reasonable,”\textsuperscript{134} but many courts only apply reasonableness restraint to CIC covenant amendments, not the original covenants. Furthermore, the test for reasonableness is not rigorously nor consistently applied.\textsuperscript{135}

Servitude restrictions on land use preserve the status quo.

\textsuperscript{131} Restatement (Third) of Prop.: Servitudes §§ 2.1, 3.1, & 3.7 (2000). See also supra notes 25–31 and accompanying text.

\textsuperscript{132} Scholars who argue that covenants should be completely analogized to contracts have been the most vocal critics of the touch and concern test in the context of common interest communities. E.g., Epstein, Covenants and Constitution, supra note 80.

\textsuperscript{133} Without substantive limits on the scope of CIC covenants, neighborhood private laws can “dictate basic aspects of a resident’s mode of living within the privacy of his or her own unit.” Armand Arabian, Condos, Cats, and CC&Rs: Invasion of the Castle Common, 23 Pepp. L. Rev. 1, 1 (1995).

\textsuperscript{134} E.g., White Egret Condo., Inc. v. Franklin, 379 So. 2d 346, 351 (Fla. 1979) (pre-FHA amendment case upholding age restrictions on condominium occupancy as “reasonable”); Nahrstedt v. Lakeside Vill. Condo. Ass’n, 878 P.2d 1275, 1283 (Cal. 1994) (“[O]ur Legislature has made common interest development use restrictions contained in a project’s recorded declaration ‘enforceable . . . unless unreasonable.’” (emphasis in original) (citation omitted)).

\textsuperscript{135} See supra notes 80–85 and accompanying text.
Although this may be the very goal sought by the authors of the servitude, perpetual real property stasis imposes future opportunity costs. Servitude rigidity is potentially problematic for all easements and covenants, but most recent scholarly debate on the costs of rigidity has focused on the context of conservation servitudes.\textsuperscript{136} Conservation servitudes restrain use of land indefinitely and are deliberately difficult to terminate.\textsuperscript{137} Placing perpetual burdens on land ignores the possibility of unexpected changes in land use needs.\textsuperscript{138} Even though today’s perfect

\textsuperscript{136} For a definition and overview of conservation easements, see Michael R. Eitel, Comment, \textit{Wyoming’s Trepidation Toward Conservation Easement Legislation: A Look at Two Issues Troubling the Wyoming State Legislature}, 4 WYO. L. REV. 57, 59 (2004). Conservation servitudes “present a difficult choice among conflicting social values. Although authorization of private conservation servitudes in gross reinforces freedom of contract, promotes the benefits of private initiative, and assists conservation of the natural environment, other important social policies suffer.” Korngold, \textit{Privately Held Conservation Servitudes}, supra note 27, at 435. The term “conservation easements” is a misnomer because such servitudes are not non-possessory use rights of a non-owner but instead are restrictions on an owner’s ability to use her own land. \textit{Id.} at 436–37. There are some key differences between CIC restraints on transfer and conservation servitudes, most importantly that the former involves a restriction on alienation and the latter only restrains use.\textsuperscript{137} \textit{Id.} at 439–43. Indeed, the whole point of conservation easements is to render future land development impossible. \textit{Id.} at 479, 453–54; see also Julia D. Mahoney, \textit{Perpetual Restrictions on Land and the Problem of the Future}, 88 VA. L. REV. 739, 767 (2002) (“[C]onservation servitudes can achieve their goals if and only if the future options of owners of burdened land are constrained.”).\textsuperscript{138} Korngold, \textit{Privately Held Conservation Servitudes}, supra note 27, at 479. Some scholars have advocated a periodic review by courts to determine whether the easement merits continued validity or should be stricken as a matter of fairness or efficiency. Gerald Korngold, \textit{Resolving the Intergenerational Conflicts of Real Property Law: Preserving Free Markets and Personal Autonomy for Future Generations}, 56 AM. U. L. REV. 1525 (2007) [hereinafter Korngold, \textit{Resolving the Intergenerational Conflicts}]; Gerald Korngold, \textit{Solving the Contentious Issues of Private Conservation Easements: Promoting Flexibility for the Future and Engaging the Public Use Land Process}, 2007 UTAH L. REV. 1039 [hereinafter Korngold, \textit{Solving the Contentious Issues}]. Another approach would be to make the beneficiaries of such easements public entities, constrained by the democratic process. See Korngold, \textit{Privately Held Conservation Servitudes}, supra note 27. Other scholars contend that the perpetual validity of conservation servitudes must be
candidate for conservation may be better allocated in the future to development, legally un-burdening land from servitude restraints may be impracticable. Future generations will bear the costs of today’s land restraints. These same worries regarding unlimited duration and specific enforceability that have engendered much debate in the context of conservation easements also apply to community CC&Rs. Such covenants impose a particular vision of community use and behavior that is resistant to change and difficult to avoid.

C. Covenant Predictability and Community Exit

Although any type of perpetual covenant may become onerous and undesirable over time, the content of community CC&Rs are less rigid than conservation servitudes and other types of easements and covenants. Unlike traditional servitudes, CIC covenants can be amended by community vote. This flexibility


140 Several scholars have focused on the issue of perpetual validity of conservation servitudes and have pointed out that the status quo may not give adequate weight to the costs of alienation restraints. E.g., Cheever, supra note 139; Korngold, Privately Held Conservation Servitudes, supra note 27; Mahoney, supra note 137; Jessica Owley, Changing Property in a Changing World: A Call for the End of Perpetual Conservation Easements, 30 STAN. ENVT'L. L. J. 121, 144 (2011).

141 CIC declarations can be amended by prescribed procedures, typically by supermajority vote of the owners. UCIOA provides that the declaration may be amended with a 67% affirmative vote unless the declaration specifies a different percentage or certain occupancy rules are impacted (threshold in that
mitigates some of the concerns otherwise posed by the unlimited duration of CIC servitudes. The ability to amend covenants is also a great advantage to the CIC structure compared with earlier neighborhood deed restriction schemes that provided no method for modification or termination of servitude restraints. But the benefit of flexibility is achieved at the cost of predictability. Because CC&R establish a dynamic association government, CICs are more adaptable than regimes that are controlled solely by the rigid provisions of a recorded document, but the changing nature of obligation in the case of CICs renders the obligation itself more difficult to justify on contractual reliance grounds.\textsuperscript{142} Thus, there are two opposite problems potentially posed by the possibility of changing CIC governing provisions. First, changes may be too difficult to achieve and may not in fact be forthcoming even when changing circumstances so warrant. Second, rule changes may be inspired by the whims of vocal neighborhood minorities and not actually reflect changing circumstances or new community values. If changes are non-unanimous (as is almost universally the case), then it is more difficult to justify the application of such changes to dissenting homeowners based on their supposed assent. Furthermore, unforeseen changes to community covenants may frustrate the reasonable expectations and desires of dissenting owners who bought into a community that was governed by a different set of substantive rules.

Theoretically, a community’s ability to amend covenant restrictions should provide a means to update neighborhood governance to reflect new cultural preferences and technological changes impacting property use. In reality, covenants are difficult to amend.\textsuperscript{143} Whether a given community is able to mobilize

\textsuperscript{142} See Brower, supra note 20, at 242 (noting that CIC enforcement is justified based on the unanimous assent of its members to covenant terms and explaining that later amendments “pose special problems”).

\textsuperscript{143} Amendments to CC&Rs are difficult to achieve in reality because of the generally low level of community engagement and participation coupled with the high levels of required assent. See generally Sterk, supra note 10.
sufficient votes for a given amendment turns on the idiosyncratic concerns of owners and the level of popular participation in the community. When restrictions in recorded declarations are difficult to modify, outdated laws will govern behavior and land use in the community. For example, many CC&Rs drafted in the 1970s and 1980s prohibit “satellite dishes,” based on the concern over blocked views and the unsightly nature of enormous satellite dishes such as those used at the time. Today’s satellite dishes are tiny and unobtrusive, yet covenants banning “satellite dishes” remain legally binding until they are removed by a supermajority. Other covenant restrictions that commonly persist, despite being criticized as obsolete, include prohibitions on trucks and laundry lines. Such blanket prohibitions seem unwarranted based on the modern trends of, respectively, driving a small pick-up truck as a passenger vehicle and air-drying of clothes in an effort to be more eco-friendly.

144 Stephen E. Barton & Carol J. Silverman, Common Interest Homeowners’ Associations Management Study (Cal. Dep’t Real Estate ed., 1987) (showing low levels of participation in community governance and concluding that many communities are not governed according to majority desires but rather the idiosyncratic concerns of a vocal minority).


146 Id. But see Portola Hills Cmty. Ass’n v. James, 5 Cal. Rptr. 2d 580, 583 (Cal. Ct. App. 1992) (striking down a restriction on satellite dishes as unreasonably obsolete), disapproved of by Nahrstedt v. Lakeside Village Condo. Ass’n, 878 P.2d 1275, 1290 (Cal. 1994) (reasonableness should be determined facially, not as applied to a particular circumstance).

147 In Bernardo Villas Mgmt. Corp. v. Black, a California court invalidated a restriction on trucks as unreasonable “as applied to clean, noncommercial pickup truck used by owners solely for personal transportation.” 235 Cal. Rptr. 509 (Cal. Ct. App. 1987), disapproved of by Nahrstedt, 878 P.2d at 1290 (reasonableness should be determined facially, not as applied to a particular circumstance).

148 “Virtually all” CICs ban outdoor clotheslines. Laura Thomas Gebert, Comment, A Survey of Selected Government-Sponsored Energy Plans and Recommendations for Florida’s Future Energy Policy, 8 Barry L. Rev. 149, 166 (2007). A typical covenant provides that “[n]o laundry or other clothes may be hung or displayed outside any Unit.” Mazdabrook Commons Homeowners’ Ass’n v. Khan, No. DC-011532-08, 2010 WL 3517030, at *12
however, that changing community opinions and mores is easier than changing community covenants.

Spotty public participation in community governance, hold-outs, and idiosyncratic vocal minorities make the flexibility of the CIC governance model haphazard. This unpredictable flexibility means that rules may change in unexpected ways. An association may enact a completely new restriction, never anticipated by members when they purchased property in the community.\textsuperscript{149} For example, in a 1978 California case, a mother and her two children were forced out of their home when their association passed a covenant amendment prohibiting occupancy by anyone under 18.\textsuperscript{150} More recently, a smoker who purchased a


\textsuperscript{149} There is a thread of case law that attempts to distinguish between changes to CIC covenant terms and the addition of new terms, with courts holding that amendment provisions in an original declaration authorize changes but not additions. \textit{E.g.}, Lakeland Prop. Owners Ass’n v. Larson, 459 N.E.2d 1164 (Ill. App. Ct. 1984); Boyles v. Hausmann, 517 N.W.2d 610, 616 (Neb. 1994). The reasoning in these cases has been criticized as logically flawed. \textit{See} Evergreen Highlands Ass’n v. West, 73 P.3d 1 (Colo. 2003). For example, changing a provision explicitly permitting leasing to explicitly prohibit leasing would be permitted under the reasoning of \textit{Boyles}, but adopting a leasing prohibition would not be permitted if the original declaration was silent as to an owner’s ability to lease. More recent cases have implicitly overruled or simply ignored these holdings. \textit{See}, \textit{e.g.}, Apple II, Condo. Ass’n v. Worth Bank & Trust Co., 659 N.E.2d 93 (holding that addition of leasing limitation was valid without even acknowledging the conflict with the Lakeland precedent).

\textsuperscript{150} Ritchey v. Villa Nueva Condo. Ass’n, 146 Cal. Rptr. 695, 700 (Cal. Ct. App. 1978). Age-based restrictions were not prohibited by statute in 1978, but a later amendment of the Fair Housing Act created a statutory basis for striking down such restrictions. \textit{See} Fair Housing Act of 1968, 42 U.S.C. §§
condominium unit in Colorado likewise did not anticipate that his association would later amend the declaration to prohibit smoking in any part of the building, including inside his home.\textsuperscript{151}

By purchasing property in a CIC, an owner is deemed to have agreed to be bound not just to the terms of the recorded declaration but also to any changes that a sufficient percentage of her neighbors may later enact.\textsuperscript{152} In many cases, the standard for judicial review of declaration amendments is a variant of the Business Judgment Rule—changes to owner obligations are deemed valid as long as the association acted in good faith and followed procedures enumerated in the governing documents.\textsuperscript{153}

\textsuperscript{151} After the judge’s ruling, the homeowner complained to the press that, “I can’t relax and have a cigarette in my own home.” Ann Schrader, \textit{Couple’s Smoking at Home Snuffed}, DENVER POST (Nov. 16, 2006), http://www.denverpost.com/news/ci_4667551; see also David B. Ezra, “Get Your Ashes Out of My Living Room!”: \textit{Controlling Tobacco Smoke in Multi-Unit Residential Housing}, 54 RUTGERS L. REV. 135, 139 (2001) (exploring the legal aspects of prohibiting smoking inside condominium units); Staci Semrad, \textit{A New Arena in the Fight Over Smoking: The Home}, N.Y. TIMES, Nov. 5, 2007, at A18 (detailing efforts within condominiums across the country to ban smoking inside units).

\textsuperscript{152} While oversight of amendments is minor, judicial review of restrictions contained in the original declaration is often even more cursory. See Noble v. Murphy, 612 N.E.2d 266, 270 (Mass. Ct. App. 1993); Brower, \textit{ supra} note 20, at 242. A complaining owner must prove that an amendment is “unreasonable” or it will be specifically enforceable. See Villa De Las Palmas Homeowners Ass’n v. Terifaj, 90 P.3d 1223, 1234. (Cal. 2004). Amendments are presumed enforceable against all owners “unless they are wholly arbitrary, violate a fundamental public policy, or impose a burden on the use of affected land that far outweighs any benefit.” \textit{Nahrstedt}, 878 P.2d at 1287. In some jurisdictions all amendments and rules properly enacted are clothed with a strong presumption of validity unless a plaintiff can show bad faith. See Arabian, \textit{ supra} note 133.

\textsuperscript{153} A typical approach is uphold any rules and regulations that have been enacted by the board, acting within the scope of its authority and not abusing its power or acting arbitrarily and capriciously. Unit Owners Ass’n of
Judges reason that by buying into a community, owners in a CIC have manifested their assent to the terms of the declaration, including the procedures for amending its terms.\textsuperscript{154} Courts conclude that by agreeing to amendment procedures, owners implicitly agreed to be bound to whatever restrictions a majority of their neighbors sees fit to impose in the future.\textsuperscript{155}

Because community restrictions are subject to majority-rule changes, they operate much like a social contract and unlike servitudes in the absence of an association or built-in amendment procedure.\textsuperscript{156} In a very real sense, CIC covenants are really

\begin{flushright}
\end{flushright}

\textsuperscript{154} See, e.g., Woodside Vill. Condo. Ass’n v. Jahren, 806 So. 2d 452, 461 (Fla. 2002) (finding that a recorded condominium declaration puts owners on notice that the restrictions governing the subject properties are “subject to change through the amendment process” and that owners have thereby agreed “that they would be bound by properly adopted amendments”); Kroop v. Caravelle Condo, Inc., 323 So. 2d 307, 309 (Fla. Dist. Ct. App. 1975) (upholding amendment prohibiting leasing because “[p]laintiff acquired title to her condominium unit with knowledge that the Declaration of Condominium might thereafter be lawfully amended”); Hill v. Fontaine Condo. Ass’n, 334 S.E.2d 690 (Ga. 1985) (an amendment restricting residence to adults only is enforceable on all owners); McElveen-Hunter v. Fountain Manor Ass’n, 386 S.E.2d 435, 436 (N.C. Ct. App. 1989), aff’d, 399 S.E.2d 112 (N.C. 1991) (holding that an amendment prohibiting leasing “does not infringe upon any legal right of the plaintiff’s; for she had notice before the units were bought that the declaration was changeable”).


\textsuperscript{156} Most CC&Rs amendments require approval by a supermajority of owners. Changes to rules, however, are made by the board of directors for the association. This board is elected by majority vote.
dynamic governing constitutions. The operation of a CIC therefore raises entity governance issues, such as how decisions are made, minority voting rights, and limits of governing power.

Group decision-making can be justified by showing that members of the group enjoy sufficient “voice” (or participation) and have the ability to “exit” (or leave) if unsatisfied with group decisions. At first blush, the CIC model seems to pose no problem on these grounds. Every owner has a vote (voice) in community governance. And although owners are bound by majority-enacted rules, this presents no real liberty concerns as long as owners can “vote with their feet” and leave if dissatisfied (exit). In the context of corporate governance, exit is the relatively simple matter of selling one’s stock. But CIC membership is bundled with homeownership and the only way to exit is to sell one’s home and move. This makes exit from a CIC tremendously burdensome. Real property is quite illiquid; it may take quite some time to find a buyer. In addition, it is personally and psychologically disruptive to relocate or divest one’s homeownership. Therefore, although exit is available in theory, market and psychological realities create a practical barrier to exit in CICs.

In some cases, restrictive covenants create legal barriers to CIC exit as well, by limiting an owner’s ability to sell or lease her property. Some CIC covenants may provide that property

---

157 The terms “voice” and “exit” are borrowed from the corporate governance classic, ALBERT O. HIRSCHMAN, EXIT, VOICE, AND LOYALTY (1970).

158 Brower, supra note 20, at 245 (noting that “[p]articipatory consent substitutes democratic decision-making and consensus building for state regulation over substantive terms”).

159 Id. at 242 (explaining the argument that assent exists even for amendments because dissatisfied owner members in a CIC are always free to leave the community if they disagree with its rules).

160 Id. at 224 (referring to the “financial and psychological stakes raised” by requiring a home sale to exit). Much of the impetus behind defaulting mortgagor rescue efforts has been the individual harms from forced home sales. See Julia Patterson Forrester & Jerome Michael Organ, Promising to Be Prudent: A Private Law Approach to Mortgage Loan Regulation in Common-Interest Communities, 19 GEO. MASON L. REV. 739, 739 (2012) (calling a forced sale of a home “clearly devastating to the homeowner”).
transfers can occur only with association consent. Others may grant the association a first right of refusal with respect to any proposed transfer. Restrictions on who can occupy a unit and prohibitions on leasing of a unit are even more common. When restrictions constrain an owner’s ability to exit a CIC regime, it no longer is valid to say that continued membership or occupancy in the private community is truly voluntary and necessarily manifests a continuing desire to be bound by the governance regime. This calls into question the continuing legitimacy of the CIC social contract.


163 Boyack, Community Covenant Alienation Restraints, supra note 33. Although limitations on occupancy based on race now are illegal and ineffective, for decades racial segregation was upheld as an acceptable way to promote the accepted policy goals of high property values and social harmony. HARRY GRANT ATKINSON & L.E. FRAILEY, FUNDAMENTALS OF REAL ESTATE PRACTICE 428–29 (1946); BROOKS & ROSE, supra note 34; MCKENZIE supra note 29, at 60–68; ROBERT C. WEAVER, THE NEGRO GHETTO 231 (1948).

D. Covenant Drafting: Authors and Influences

CIC “agreements” consist of non-negotiable covenants that have already been drafted and recorded by the developer to create a binding servitude on the land before homes are ever sold.164 This not only informs the reality of homeowner choice, it also reveals that none of the community residents actually authors the covenants that bind the community. Who, then, dictates these adhesive provisions?

At first blush, the answer seems to be that it is the developer who drafts the governing documents, forms the CIC association, and records the declaration, but the reality is more complicated.165 Some market theorists claim that the unilateral act of a developer in designing CIC covenants is not troubling because in choosing to create a CIC and in crafting the content of community CC&Rs, the developer takes into account consumer preferences as a way to maximize sale price.166 This makes sense in theory, but in reality, this has never been completely true. Instead, as a condition of zoning approval, local municipalities often require

---

164 See McKENZIE, supra note 29, at 127; supra Part I.B. Prior recordation is required to legally sell a condominium unit and is prudent in order to create a binding servitude on subsequent property owners. HYATT, supra note 4; Winokur, supra note 105. Some early CICs were established from existing neighborhoods, and in such cases, homeowners did theoretically have some input into a declaration’s content. McKENZIE, supra note 29, at 33–36.

165 See generally WAYNE S. HYATT, CONDOMINIUMS AND HOME OWNERS ASSOCIATIONS: A GUIDE TO THE DEVELOPMENT PROCESS (1985) (explaining the developer’s process of creating a CIC and explaining how home buyers are recipients of, rather than shapers of, the initial servitude regime). In a section titled “Developer-Appointed Boards Should Actively Lead the Owners,” Hyatt notes: “[M]ost people, by obvious logic, are followers in most aspects of their lives—some in virtually all respects. Social order would not be obtained without that condition.” See also McKENZIE supra note 29 at 21, 127 (describing the developer’s role in establishing CC&Rs and bemoaning lack of resident input into the governing terms); Franzese & Siegel, supra note 50, at 1127–30 (“CIC residents play no direct role in the critical decision-making process leading to the organization of the CIC.”); Winokur, supra note 105, at 58–60 (explaining the complete lack of homeowner input with respect to the content of CIC covenants).

166 Forrester & Organ, supra note 160, at 744–45.
that a new development be organized as a common interest community, and this factor drives CIC creation perhaps more than anything else. Furthermore, Fannie Mae and Freddie Mac (each a “Government Sponsored Enterprise” or “GSE”),\(^\text{167}\) and the Federal Housing Administration (“FHA”) indirectly determine the content of CIC covenants through their mortgage finance underwriting guidelines.\(^\text{168}\) Because of the influence of these government actors, CIC covenants have become standardized in the industry and may fail to represent developer marketing strategy or consumer preferences.\(^\text{169}\)

The vast majority of mortgage loans made today are insured by the FHA or earmarked for resale to Fannie Mae or Freddie Mac.\(^\text{170}\) The GSEs were at one time private entities but have always been heavily regulated at the federal level and were established with an implicit (later explicit) government

\(^{167}\) Fannie Mae (formerly the Federal National Mortgage Association) and Freddie Mac (the Federal Home Loan Mortgage Corporation) were chartered by Congress and regulated by federal agencies and since 2008 have been in conservatorship with the federal government. See Andrea J. Boyack, *Laudable Goals and Unintended Consequences: The Role and Control of Fannie Mae and Freddie Mac*, 60 AM. U. L. REV. 1489, 1499–1502 (2011) [hereinafter Boyack, *Laudable Goals*] (giving an overview of the market role and enumerated purposes of Fannie Mae and Freddie Mac).

\(^{168}\) See Boyack, *Community Collateral Damage*, supra note 2.

\(^{169}\) Franzese & Siegel, supra note 50; Steven Siegel, *The Public Role in Establishing Private Residential Communities: Towards A New Formulation of Local Government Land Use Policies That Eliminates the Legal Requirements to Privatize New Communities in the United States*, 38 URB. LAW. 859, 873–98 (2006) [hereinafter Siegel, *The Public Role*]; see also Dilger, supra note 106, at 38 (explaining that CICs are “increasingly offering fairly uniform levels and types of services.”).

guaranty. They exist in order to promote homeownership. But historically and today, the GSEs do more than funnel money into the residential mortgage market: through approval requirements and form documents, the GSEs and the FHA dictate the terms of housing arrangements at every level.

In order to qualify for resale to one of the GSEs, a mortgage must be secured by an acceptable property. In the CIC context, that generally means that the community in which the property is located must meet GSE underwriting mandates. The Department of Housing and Urban Development maintains a list of “Approved Condominium Projects,” and typically Fannie Mae and Freddie Mac will only purchase mortgages on units in condominiums that are on the approved list.

Communities with a high percentage of non-owner-occupied

---


172 Boyack, Laudable Goals, supra note 167, at 1495.

173 In crafting CIC declarations, developers lift language directly from government forms and model documents and mirror precisely GSE and FHA underwriting requirements. Winokur, supra note 105, at 59.

174 Both Fannie Mae and Freddie Mac prohibit any ownership concentration in condominiums, meaning that if one owner holds title to 10% or more of the units, no unit in the CIC may secure a GSE mortgage. Additional requirements include required community majority owner occupancy for loans to owner-investors, at least 10% of the association’s budget earmarked to fund reserves, and no more than 15% of the members being delinquent on paying their assessments. FREDDIE MAC CONDOMINIUM UNIT MORTGAGES (July 2013, available at http://www.freddiemac.com/learn/pdfs/uw/condo.pdf; EFANNIEMAE SEC INSTRUMENTS (2014), available at https://www.efanniemae.com/sf/formsdocs/documents/secinstruments.

units or a high percentage of members in default on assessment payments will not appear on the approved lists and thus will likely not qualify for GSE mortgage funds. The precise threshold percentages vary from time to time, and precise mandates of Fannie Mae and Freddie Mac may differ, but the GSEs typically preclude mortgage loans secured by properties in CICs where more than 15% of the owners are delinquent in their assessments or where more than 50% of units are non-owner-occupied. This latter provision justifies community restrictions on leasing. Because of the community owner occupancy requirement, standard form declarations provide for various levels of control over an owner’s ability to lease, ranging from complete or near-complete prohibition of leasing to nearly ubiquitous (and GSE/FHA-mandated) restrictions on short-term rentals. In some contexts it is tricky to comply with both GSE owner occupancy standards and the mandates of the FHA, however. The FHA views a complete ban on leasing as an unlawful restraint on alienation, but the GSEs require high community owner occupancy rates. Because of this, conventional wisdom in crafting CIC declarations is to prohibit most—but not all—units from being leased. This allows a CIC to walk the line

176 Boyack, Community Collateral Damage, supra note 2, at 105–06.
178 See supra notes 166–71 and accompanying text.
between running afoul of the FHA rules and disqualifying the
community from GSE investment.

Complying with the underwriting requirements of Fannie
Mae, Freddie Mac, and the FHA can make or break a CIC
project. Properties in qualifying communities have access to
vastly more mortgage capital, and liquidity bolsters property
values. Conversely, property in a community with too many
tenants or too many assessment-delinquent owners will be cut off
from mortgage funds, decreasing the property’s liquidity and
market price and perhaps even rendering the property
unsellable. Developers across the nation want their products
sold for the highest prices and therefore need their would-be
buyers to have access to the requisite funds. This requires that the
developers will frame the CC&Rs to match the guidelines of the
GSEs and the FHA whenever possible.

Because meeting FHA and GSE requirements is so vital to
community success, changes to entity and agency policies can
rapidly and effectively impact covenant content for future CICs.
One such example is how new policies of the GSEs and FHA
rapidly changed the use of private transfer fees (PTF) covenants
in CC&Rs. Over the past decade, many developers started
including PTF covenants in CC&Rs as a way to defer and
privatize payment of today’s development costs. PTF covenants

law.com/articles/fha-information; Jim Slaughter, FHA Guidelines &
Condominium Rental Restrictions, ROSSABI BLACK SLAUGHTER, PA (Mar. 1,
2013), http://www.lawfirmrbs.com/blog/fha-guidelines-condominium-rental-
restrictions/. It is literally possible to satisfy both GSE and FHA requirements
of CIC covenants provided that only one unit may be rented at any time.

180 This sets up a strange dichotomy: in communities with no-
leasing covenants, owners cannot legally rent, but in communities without such
covenants, too many neighborhood rentals will make it practically impossible
for an owner to sell. The existence of GSE guidelines on owner occupancy
thus necessarily restricts (practically if not legally) the owners’ ability to
transfer. For a more detailed discussion of this conundrum, see Boyack,
Community Collateral Damage, supra note 2.

181 See Winokur, supra note 105, at 59.

182 For example, between 2001 and 2006, Lennar Corporation included
PTF covenants into CC&Rs governing 13,000 homes in California. These
PTFs are payable to the Lennar Charitable Housing Foundation. See Robbie
Wheelan, Home-Resale Fees Under Attack, WALL ST. J. (July 30, 2010),
require that a fee equal to a percentage of the sale price be paid either to the association or to a designated third party as a condition of property resale. More than eleven million homes are currently encumbered by PTF covenants.

Innovators of such PTF covenants claim that these covenants keep housing affordable by temporally spreading the ballooning costs of development. Mimicking the traditional freedom of contract rationale in CIC oversight cases, PTF proponents argue that PTF covenants are not unfair because buyers in CICs, by the very act of purchasing the property in the first place, have agreed to pay these resale fees in the future. But many buyers, policy

http://online.wsj.com/news/articles/SB10001424052748703314904575399290511802382. In New York, Freehold Partners crafted a creative solution to building costs by entering into agreements with developers to buy the right to collect PTFs in exchange for upfront development fees. Freehold then securitized the obligations by pooling and selling shares in the aggregate income stream from PTFs. Id. (“Municipalities have long used similar fees, called transfer taxes, to raise revenues or recoup public subsidies for private development projects, but private transfer fees are relatively new.”). For an excellent and thorough discussion of PTFs, see R. Wilson Freyermuth, Private Transfer Fee Covenants: Cleaning Up the Mess, 45 REAL PROP. TR. & EST. L.J. 419 (2010).

Freyermuth supra note 182; see also Richard Mansfield, Private Transfer Fee Covenants: A Thing of the Past?, WORLDWIDE ERC (Feb. 7, 2011, 10:32 AM), http://www.worldwideerc.org/Blogs/MobilityLawBlog/Lists/Posts/Post.aspx?List=c020aee5%2D48ad%2D47b2%2D82952Da4cf71ba9e34&ID=57 (explaining how PTFs work and when they came into use).


The act supposedly manifesting assent to the PTF covenants included in recorded CC&Rs was the home purchase. Ward & Hopkins, supra note 184, at 902; see also Freyermuth, supra note 182 (explaining the problematic aspects of inferring consent in this way). One law review article considering the issue of PTF covenants contends that buyers would simply decrease their offer price when purchasing property burdened by PTF covenants in recognition of their obligation to pay in the future, but there is no indication
makers, and legislatures objected to an imposition of a long-term private tax on transfers of property, particularly when the proceeds of such fees went to private investors.\footnote{Id. The Coalition to Stop Home Resale Fees asserted that PTFs are “Wall Street lining their pockets while stealing equity from homeowners.” Wheelan, \textit{supra} note 182. Most PTFs are not designed to exist in perpetuity, but rather provide an expiration date, typically 99 years. Mansfield, \textit{supra} note 183; Ward & Hopkins \textit{supra} note 184.} At least thirty-six states responded to the advent of PTFs by passing laws limiting their validity,\footnote{By 2011, 36 states had passed some limiting legislation with respect to PTFs. Ward & Hopkins, \textit{supra} note 184, at 902.} typically channeling PTF proceeds to community associations and prohibiting payment of PTFs to private for-profit third parties.\footnote{Only a handful of regulating states prohibit PTFs. Most of the legislative focus has been on to whom the fees are paid, not whether the fees are payable. \textit{Id.} California’s statutory fix permits all types of PTFs but mandates special disclosures. Mansfield, \textit{supra} note 183.}

It is recognized that state governments can pass statutes to directly prohibit certain types of covenant restrictions. The federal government’s ability to control the content of CIC covenants, however, is less obvious. No federal agency has the authority to ban certain types of covenants from CIC declarations, but the Federal Housing Finance Agency (FHFA)\footnote{The FHFA was created by the Housing and Economic Recovery Act of 2008. See \textit{Pub. L. No.} 110-289, 122 Stat. 2654 (2008) (codified as amended in scattered sections of 12, 15, 26, 37, 38, and 42 U.S.C.A.). The primary purpose of FHFA is to regulate Fannie Mae, Freddie Mac, and the Federal Home Loan Banks.} can achieve this indirectly. The FHFA controls the actions of Fannie Mae and Freddie Mac. The FHFA regulates the GSEs, and GSE underwriting requirements drive CIC structuring. When the FHFA tells the GSE to refrain from purchasing mortgages secured by property burdened by certain types of restrictions, it indirectly—but tremendously effectively—mandates the content of

---

Id. The Coalition to Stop Home Resale Fees asserted that PTFs are “Wall Street lining their pockets while stealing equity from homeowners.” Wheelan, \textit{supra} note 182. Most PTFs are not designed to exist in perpetuity, but rather provide an expiration date, typically 99 years. Mansfield, \textit{supra} note 183; Ward & Hopkins \textit{supra} note 184.
CIC covenants.\footnote{191} In March 2012, the FHFA published a rule prohibiting Fannie Mae, Freddie Mac, and the Federal Home Loan Banks from purchasing mortgages on properties encumbered by PTFs payable to third parties.\footnote{192} This FHFA regulation has been tremendously effective, virtually wiping out privately directed PTF covenants in CICs formed after March 2012. This example provides an interesting glimpse into how federal agency action can directly impact the content of “private” CC&Rs.

Locally, municipalities can also impact covenant communities by requiring a CIC structure in exchange for granting zoning approval for new projects.\footnote{193} Since zoning approval is a prerequisite to creating a new community, local regulators’ preferences, with respect to the existence and content of CC&Rs, are incorporated whenever possible. In mandating covenant substance, many municipalities adhere to the FHA guidelines

\footnote{191} One commentator opined that the FHFA regulation “virtually guarantees that [PTFs] will be used no more.” Bobby Saadieh, \textit{FHFA’s Final Ruling Will Restrict Private Transfer Fees}, \textsc{PERTRIA} (Apr. 12, 2012, 11:04 AM), \url{http://www.pertria.com/2012/fhfas-final-ruling-will-restrict-private-transfer-fees-2/}.

\footnote{192} FHFA Restrictions, 12 C.F.R. § 1228 (2012). The FHFA rule does not address PTFs payable to a community association. The FHFA rule also excludes PTFs paid to certain tax-exempt organizations that use the PTF proceeds to benefit the property, but includes any fees not allocated to property improvement and upkeep. \textsc{Id.} The rule also applies only prospectively (from its announcement in 2011), and thus impacts CIC declarations recorded after that time, but not any of the previously recorded CC&Rs that included PTF provisions. \textsc{Id.; see also Mansfield, supra note 183} (explaining that this rule will not affect the thousands of existing mortgages for deeds containing a PTF covenant). For further discussion of the FHFA rule, see \textit{Announcement SEL-2012-05, Selling Guide} (Fannie Mae, Washington, D.C.), June 19, 2012, at 1, \url{available at https://www.fanniemae.com/content/announcement/sel1205.pdf}. Freehold Capital, however, estimates that over $600 billion worth of PTF securities are currently in commerce. \textsc{FREEHOLD CAPITAL, supra note 185}.

\footnote{193} \textsc{See Siegel, The Public Role, supra note 169, at 877–95} (calling the CIC ownership concept as “a form of ‘grand bargain’ between developers and municipalities” and citing to several local zoning statutes that require use of the CIC form).
with respect to CIC structuring.\textsuperscript{194} Financial realities motivate municipal requirements as well. Local governments have long realized that the CIC ownership structure can be used as a vehicle for privatizing traditional municipal functions.\textsuperscript{195} The greater the percentage of community amenities and upkeep that can be channeled to private community maintenance, the better for the municipal budget.

These external influences on the content of CIC covenants is obscured by continued judicial assertions that such covenants represent the private contractual choices of the residents in a given community.\textsuperscript{196} In reality, covenant terms do not necessarily represent homeowner will in any real sense.\textsuperscript{197} Rather, CIC covenants are more likely to reflect the extent to which a

\textsuperscript{194} The FHA prescribes numerous “initial” terms for CC&Rs and also strongly advocates the imposition of supermajority requirements to amend CIC governing documents. Such supermajority requirements attempt to promote predictability preferred by FHA insurers and “prevent owners from banding together.” McKENZIE, supra note 29, at 127. These “recommendations” are backed with the possibility of FHA mortgage insurance and have been widely followed. Franzese & Siegel, supra note 50, at 1114.

\textsuperscript{195} TREESE ET AL., supra note 42, at 3; Boyack, Community Collateral Damage, supra note 2, at 60; Siegel, The Public Role, supra note 169, at 879.

\textsuperscript{196} Franzese & Siegel, supra note 50, at 1112 (the “CIC phenomenon is, increasingly, the direct product of conscious and deliberate government policy . . . ”). The CIC covenant situation is an example of an adhesion contract drafted by neither party to the transaction, “where the terms are proffered by a third party and both contracting parties are reduced to the humble role of adherent.” Andrew A. Schwartz, Consumer Contract Exchanges and the Problem of Adhesion, 28 YALE J. ON REG. 313, 346 (2011).

\textsuperscript{197} To the contrary, numerous studies have shown that homeowners are dissatisfied with the content of their community covenants and, as a general rule, the provisions of CC&Rs diverge markedly from community preferences. Winokur, supra note 105, at 63 n.260–61; see also STEPHEN E. BARTON & CAROL J. SILVERMAN, CAL. DEPT. OF REAL ESTATE, COMMON INTEREST HOMEOWNERS’ ASSOCIATIONS MANAGEMENT STUDY (1987). A report published by the Urban Land Institute found that a majority of residents in CICs were greatly dissatisfied with their community. CAROL NORCROSS, TOWNHOUSES & CONDOMINIUMS: RESIDENTS’ LIKES AND DISLIKES 80 (1973). The report characterized residents as “unhappy, resentful, discouraged, and disillusioned about their associations,” with “[a] considerable number of families . . . so angry that they are selling their homes and moving away . . . to get away from what they think of as strait-jacket controls on their lives.” Id.
developer acquiesces to municipal requirements and follows FHA and GSE underwriting “guidance.”

III. A LEGAL-HYBRID APPROACH TO CICs

A. Refocusing Freedom of Contract Policy

CIC covenants are legal hybrids, not contracts. Servitude law determines their duration and enforceability, and their functions approximate association governance or even, to some extent, public local governments. Because CIC covenants are real property servitudes that create dynamic private community governance systems—not mere contracts—contract law should not create a basically un-rebuttable presumption of validity. The economic justifications for presumptive enforcement of contracts voluntarily entered into do not always apply to CIC covenants. Unlike voluntary contracts, CIC covenants are not necessarily freely chosen by owners who voluntarily elect to be bound by their terms. Autonomy and efficiency policy goals, therefore, are not necessarily promoted by CIC covenant enforcement. Promoting the underlying values that freedom of contract represents should inform the decision of whether to enforce CIC governing acts, but CIC covenants and regulations should not be upheld simply based on the rhetoric of freedom of contract as an end in and of itself. At a minimum, an owner’s overt act specifically manifesting assent should be prerequisite to being bound to the provisions of community CC&Rs. In addition, unlike contracts, courts and legislators should protect the public

198 NORMAN WILLIAMS, JR., & JOHN M. TAYLOR, AMERICAN PLANNING LAW: LAND USE AND THE POLICE POWER § 49.2 (rev. ed. 2013) (explaining that in a CIC, “the actual decisions on land use and building forms in the district, and perhaps also on density, are explicitly to be made, not by a general public policy adopted in advance, but by negotiation between the municipality and the developer”); Siegel, The Public Role, supra note 169, at 879–80.

199 Contracts voluntarily entered into should be enforceable notwithstanding unfairness created by their terms. Economic and liberty theory justifies this result. Treating CIC covenants as if they were contracts freely chosen by the members who are bound by their terms, however, does not necessarily promote autonomy and efficiency.
interest by limiting CIC governance’s permissible subject matter and scope. To summarily validate private community regulations as if they were mere contract provisions does not necessarily promote the values of autonomy and efficiency. To the contrary, in some cases, it threatens these same values.

Freedom to voluntarily obligate oneself in contract to terms of one’s choice is a paramount and protected legal right allocated to capable parties in our society. Each person with this freedom to contract has the power to be his or her own legislature and create binding obligations that will be enforced by the court. Freedom of contract is a universal concept, a key characteristic of almost every legal system. The principle that agreements are binding is


202 All common law jurisdictions have cases that reiterate the primacy of the principle of freedom of contract. See, e.g., News Ltd. v. Austl. Rugby Football League Ltd. (1996) 135 ALR 33 (Austl.) (explaining judicial hesitancy to “interfere with the general freedom of contract under the law”). Freedom of contract is a foundational piece of European contract law and the contract jurisprudence of all EU Member States of the European Union. See
the cornerstone of international law, and one of the fundamental precepts in our political philosophy.

Freedom of contract theory requires voluntary assent, and enforcement of private agreements is predicated on personal autonomy both with respect to choosing to be bound in obligation and with respect to choosing the parameters of that obligation. Freedom of contract meshes well with American primacy of personal freedom and capitalist economic theory of market self-regulation that considers each contracting party the best judge of his or her own interests. In addition, many commentators believe that allowing individuals the power to contract as they choose, substantially free from regulatory interference or oversight, advances liberty interests. Economic theory also

THE PRINCIPLES OF EUROPEAN CONTRACT LAW, art. 1:102 (2002).


204 Western capitalist countries, especially the United States, adhere more strongly to freedom of contract principles in their purest, least constrained form. See P. S. Atiyah, The Rise and Fall of Freedom of Contract 10 (1985).

205 See Reinhard Zimmerman, The New German Law of Obligations: Historical and Comparative Perspectives 205 (2005) (“[F]reedom of contract is not an end in itself. Rather, it must be regarded as a means of promoting the self-determination of those who wish to conclude a contract.”).


207 Richard Epstein calls freedom of contract an essential aspect of individual liberty, guaranteeing “to individuals a sphere of influence in which they will be able to operate, without having to justify themselves to the state or to third parties.” Richard A. Epstein, Unconscionability: A Critical
posits that optimal efficiency results when individuals may contract freely, and that judicial protection of the future expectations created by contracts increases societal wealth.

Even though freedom of a contract is an aspect of personal liberty, all contract enforceability presents a temporal autonomy paradox. An individual who exercises her freedom of contract today binds her future self, necessarily limiting her later freedom. Future freedom limitations are only justified because they are voluntarily chosen. Protections against involuntary contracting ensure that a party’s freedom is only restricted to the extent that she so chooses. Furthermore, the policy of allowing

Reappraisal, 18 J.L. & Econ. 293, 293–94 (1975) [hereinafter Epstein, Unconscionability]. According to theories of autonomy and individual will, it is empowering to grant contracting parties quasi-legislative powers inter se. See, e.g., Brian A. Blum, Contracts § 1.4.1 (6th ed. 2013) (“The power to enter contracts and to formulate the terms of the contractual relationship is . . . an integral part of personal liberty.”); E. Allan Farnsworth, The Past of Promise: An Historical Introduction to Contract, 69 Colum. L. Rev. 576 (1969) (drawing parallels between legislation and contract).

Although widely accepted throughout the twentieth century, the efficient market hypothesis has come under fire during the most recent financial crisis, with some theorists blaming free markets for creating the real estate bubble that sparked a global financial meltdown in 2008. Other theorists opine that it was the interference with the free market that created systemic volatility. For a brief overview of these competing viewpoints, see David Shay Corbett II, Free Markets and Government Regulation: The Competing Views of Thomas Woods and George Cooper, 14 N.C. Banking Inst. 547 (2010).


See Winokur, supra note 105, at 50 (explaining this concept in terms of Ulysses tying himself to his ship’s mast, deliberately robbing his future self of the freedom to react to the sirens’ song).

For example, the doctrines of duress, undue influence, unconscionability, incapacity, and fraud all protect a contracting party from
contractual non-performance in exchange for payment of compensatory damages ameliorates concerns about limitations of one’s future freedom.\textsuperscript{212} Efficiency policy supports the contract damages approach as well, justifying not only a party’s freedom to enter a contract but also her freedom to breach the contract upon paying the non-breaching party’s expectation interest.\textsuperscript{213} Freedom to breach a contract and pay damages is a widely touted American innovation that supports the dual values of efficiency and personal liberty, and mitigates the temporal autonomy paradox of contract law.\textsuperscript{214} Although continuing to be obligated to the financial effect of a contract, contracting parties typically can use breach to exit the contracting relationship.\textsuperscript{215} The voluntary

\textsuperscript{212} Courts generally award expectation damages for a breach of contract equal to the economic difference between what the non-breaching party expected to obtain from the breaching party’s performance and what actually was obtained (plus foreseeable costs resulting from the breach and less any cost savings from avoiding reciprocal performance and from mitigation). The theory behind expectation damages has been explained as best approximating the value of both retrospective and prospective reliance and as the economic equivalent of the bargained-for interest of the contracting parties. \textit{See} David W. Barnes, \textit{The Net Expectation Interest in Contract Damages}, 48 Emory L.J. 1137, 1139 (1999); L. L. Fuller & William R. Purdue, Jr., \textit{The Reliance Interest in Contract Damages (Pt. 1)}, 46 Yale L.J. 52, 57–62 (1936).

\textsuperscript{213} In the late nineteenth century, Oliver Wendell Holmes posited that breach of contract is viewed by the law as “amoral,” and is essentially an option purchased through payment of expectation damages. Holmes, \textit{supra} note 118, at 462. Theorists of the law and economics school have seized upon this concept and expanded it into the theory of efficient breach, holding that “it is uneconomical to induce completion of performance of a contract after it has been broken” and explaining that the law should encourage (or at least not discourage) any breach that is “efficient.” POSNER, \textit{supra} note 100, at 149–51.

\textsuperscript{214} The default remedy in contract breach actions in the United States is a monetary award of expectation damages, but under civil law, breach of contract is typically remedied by an order of specific performance rather than a monetary calculation of damages. \textit{See} Zapata Hermanos Sucesores, S.A. v. Hearthside Baking Co., 313 F.3d 385, 389 (7th Cir. 2002), \textit{cert. denied}, 540 U.S. 1068 (2003) (“[T]he civil law grants specific performance in breach of contract cases as a matter of course.”).

\textsuperscript{215} Breach as a tool for flexibility justifies other aspects of contract law such as judicial reluctance to excuse an obligation based on changed circumstances, judicial scrutiny of penalizing liquidated damages provisions,
manifestation of assent requirement coupled with contract law’s approach to damages adequately ensures both freedom and efficiency in a typical contract context.

The same values that underlie freedom of contract theory can only justify the enforcement of CIC covenants if there is a higher threshold of true assent. Because CIC covenant terms are more durable than contracts and are specifically enforceable, the possibility of breach and the passage of time do not ameliorate their effect. Actual informed assent is therefore even more vital in the context of private community covenants. In addition, because the key act of assent drives obligation under the dynamic governing process, and because individual opt-out is not possible in a CIC (absent sale of the home), the law should require that the amendment and rulemaking processes be specifically known by and explicitly agreed to by owners from the start. This can be accomplished through (a) requiring homeowners to demonstrate a separate manifestation of intent to be bound by the CIC, apart from the mere purchase of a parcel of real property located in a given community, and (b) through mandating a more effective (earlier, more accessible) disclosure of community covenants and rules.

Public policy restraints in contract law also offer some ideas about how to deal with covenants and rules that impact other important social policies. While courts generally uphold contracts regardless of their content, there is some degree of judicial suspicion with respect to certain contractual provisions such as and judicial reluctance to order specific performance. See John D. Wladis, Common Law and Uncommon Events: The Development of the Doctrine of Impossibility of Performance in English Common Law, 75 GEO. L.J. 1575 (1987).


For a discussion of how property purchase is deemed assent to current as well as future terms of community governing documents and association acts, see supra notes 50–93, 164–73, and accompanying text.

The Consumer Financial Protection Bureau has been pioneering efforts to increase the effectiveness of consumer disclosures in the context of mortgage lending. A similar effort should drive qualitative improvements of CIC disclosures to homebuyers.
limitations on a party’s autonomy with respect to future contracting or future breach, limits on free trade, and barriers to free alienation. For example, although parties might agree today that no modification to a contract will be binding unless that agreement is evidenced by a signed writing, if parties later agree to orally modify the contract, the later oral modification will still be enforced at common law. In spite of the general hands-off approach to the subject matter of contracts, courts do police contractual promises not to compete based on public policy concerns regarding market freedom and an individual’s right to earn a livelihood. Contractual promises designed to have the in terrorem effect of discouraging breach, in the form of penalizing liquidated damages clauses, are likewise subject to judicial restraint and invalidation. And limitations on property alienability have been legally suspect for hundreds of years.

While some contracts are deemed unenforceable on substantive public policy grounds, this is a rather exceptional result. See Restatement (Second) of Contracts § 179 (1981); E. Allan Farnsworth, Contracts § 5.1 (4th ed. 2004); see also Swaverly v. Freeway Ford Truck Sales, 700 N.E.2d 181 (Ill. App. Ct. 1998) (finding that public policy strongly favors freedom to contract and enforcement should only be avoided if a contract clearly contravenes articulated public policy).

Historically, covenants not to compete were held to be invalid restraints on trade. See Valley Medical Specialists v. Farber, 982 P.2d 1277, 1281 (Ariz. 1999). Courts will, however, enforce non-compete provisions that are determined to be reasonable in scope. See, e.g., Estee Lauder Companies, Inc. v. Batra, 430 F. Supp. 2d 158, 177 (S.D.N.Y. 2006); Ohio Urology, Inc. v. Poll, 594 N.E.2d 1027, 1031–32 (Ohio Ct. App. 1991).


See Michael D. Kirby, Comment, Restraints on Alienation: Placing A 13th Century Doctrine in 21st Century Perspective, 40 Baylor L. Rev. 413, 413 (1988) (“Without doubt, the concept of free alienability is a cornerstone of
Correctly interpreting and applying public policy constraints on CIC covenants is vital to ensuring the proper scope and role of private community governing rules.

Another issue involving CIC covenants is their presumptive specific enforceability. Specifically enforcing covenants regardless of their impact on community preferences and their economic costs is an unwarranted dilution of owners’ and, in some cases, non-owners’ liberty. Only in cases where parties have actually and voluntarily agreed to provisions that restrain important freedoms should courts specifically enforce these sorts of covenants. Over-reliance on the form of freedom of contract without requiring actual assent undermines both autonomy and efficiency—the very social values that freedom of contract is designed to promote.


225 Franzese & Siegel, supra note 50. C.f. Epstein, Covenants and Constitutions, supra note 80, at 922–25 (arguing that covenants should be presumptively enforceable against buyers with constructive notice because freedom of contract should be the lens through which to view a servitude
There has been much scholarship endorsing a hands-off judicial enforcement of CIC covenants based on the wholesale application of freedom of contract theory. But this approach is justified if, and only if, the “agreement” to CIC governance really fits the traditional concept of a voluntary assent. If a contact provision truly reflects party will and intent to be bound, and if the obligation only lasts a “reasonable time,” and if it imposes no unwarranted costs on third parties, then a freedom of contract justification is quite compelling. But in reality, many modern CC&Rs do not promote the autonomy and liberty values behind freedom of contract. They are not really products of party intent to be bound, they presumptively last in perpetuity, and they impact personal freedoms of contract parties and non-parties. Because of this covenant-contract disconnect, freedom of contract theory provides insufficient justification for the negative externalities that certain types of CIC restrictions impose.

B. Limiting Servitude Scope

In addition to setting a higher assent threshold in the context of CC&Rs, the law should revitalize the concept of a substantive regime).  


227 See Kirby, supra note 223, at 429 (finding that courts “have not adequately examined freedom of contract and its relationship to promissory restraints” and concluding that “if two parties contract that a particular property will not be subject to sale for some reasonable time” then such agreement should be upheld).

limit on CIC covenants beyond the outer limit of public policy. Traditional servitude law provided this sort of limitation on covenants scope: the touch and concern test. But years of stretching this test to address all manner of restrictions has deprived it of any real meaning, and several modern scholars have called for its abolition. Yet there remain compelling reasons to have some sort of more restrictive substantive limit in the law of real covenants beyond the public policy limitation of contract law. Opting out is not an option in CICs. Breach does not terminate obligation and a party cannot elect damages in lieu of performance. Changing covenant terms is cumbersome at best and impossible in some cases. Thus, there is a great need to have some initial control of the legitimate subject matter for regulations of private community covenants.

At the other end of the spectrum from those who call for hands-off enforcement of all CIC covenants in the name of freedom of contract are CIC naysayers who condemn this entire system of property ownership and private governance. But calling for elimination of condominiums, planned developments, and association governance goes much too far. CIC governance serves legitimate social functions. It provides a workable solution to the tragedy of the commons, allowing shared neighborhood amenities and common areas. It creates effective ways to combat community nuisances caused by use incompatibilities. And it

---

229 See supra notes 23–28 and accompanying text.

230 Nuisance law is notoriously difficult to apply and necessitates *ad hoc* decisions of reasonableness of a given use, leading to erratic results. Rose, supra note 30, at 5. Prosser famously called the law of nuisance an “impenetrable jungle.” W. PAGE KEETON ET AL., PROSSER AND KEETON ON THE LAW OF TORTS § 86 (5th ed. 1984). See also Boyack, *Community Covenant Alienation Restraints*, supra note 33; Winokur, supra note 105, at 37. Prior to the advent of association governance, restrictive covenants would only be enforced if an individual owner chose to sue for enforcement in court. Such owner would bear the costs of this lawsuit, but all owners in the community would benefit from having the covenant enforced. See McKENZIE, supra note 29, at 35; Marc A. Weiss & John W. Watts, *Community Builders and Community Associations: The Role of Real Estate Developers in Private Residential Governance*, in ADVISORY COMM’N ON INTERGOVERNMENTAL RELATIONS, RESIDENTIAL COMMUNITY ASSOCIATIONS 98 (1989). This was yet another manifestation of the freeriding problem and generally discouraged
can (perhaps only theoretically in some cases) foster engagement and involvement at a local, grassroots level in community problems, planning, and coalescence. In order to preserve the effectiveness and value of these functions, CIC intrusion into illegitimate spheres—such as those that impact important personal freedoms or are not justified by neighbors’ economic interests—should be disallowed.

The challenge comes in distinguishing justifiable realms of community governance from unwarranted incursions of private regulatory power. Because community servitudes can provide a workable solution to neighborhood nuisances, limitations on property use should be presumptively within the proper scope of CIC covenants and association governance, particularly with respect to uses that create cost externalities.

Other permissible areas of community governance relate to the valuable CIC function of solving two economic failures of common property: first, regulation of common areas to prevent overuse, and second, requiring affirmative contribution to common area upkeep to prevent freeriding. It is therefore legitimate for CIC covenants to address the uses of both common and individual property in the community. And CICs should also be empowered to mandate pro rata owner assessment contributions, take actions to collect these assessments, and ensure the upkeep of common areas.\textsuperscript{231} Solving the “tragedy of the commons”\textsuperscript{232} in terms of overuse and freeriding has been one of the tremendous contributions that CICs have made.\textsuperscript{233} CICs reap societal gains in encouraging community amenities, legal enforcement of such ungoverned covenant regimes.

\textsuperscript{231} Most associations’ governing documents explicitly provide for assessment funding of association obligations. See Hyatt, supra note 4, at 105, 108. Where covenants do not so provide, courts have liberally implied the power to collect assessments from owners who are benefitted by community amenities and upkeep. See, e.g., Evergreen Highlands Ass’n v. West, 73 P.3d 1 (Colo. 2003).

\textsuperscript{232} Garret Hardin, \textit{The Tragedy of the Commons}, 162 SCI. 1243, 1244–45 (1968).

\textsuperscript{233} See Treese et al., supra note 42, at 3–5 (noting that common upkeep also allows a community to take advantage of cost savings from economies of scale); Ellickson, supra note 49, at 1522–23 (discussing the equitable methods of assessments and distribution of costs amongst property owners).
providing for fair allocation of maintenance costs, and arbitrating between use incompatibilities. Covenants addressing use, upkeep, and owner maintenance contribution should therefore be considered justifiably within the substantive limits of servitude law.

A modern conception of “touch and concern” could draw the appropriate distinction, holding that how a property is used and the requisite maintenance of that property—and requisite contribution to common property—are aspects that are substantively related to the real property itself.234

Other types of community covenants and rules, however, fall beyond the permitted scope for governance by servitude. Controlling who resides in a property, for example, is not the same as controlling what the use of the property is.235 Occupancy limitations, leasing prohibitions, and transfer restrictions are not legitimate solutions to “commons” issues, but rather are unjustifiable attempts by members of a community to control their neighbors’ identity. Likewise, rules controlling behaviors that are completely contained within a home are difficult to justify on the basis of neighborhood externalities.236 Such covenants should be

---


235 The use of the property turns on how it is enjoyed and employed by the party in possession. For example, between a landlord and a tenant, it is the tenant’s use that defines the use to which the property is being put. Several courts have specifically held that renting a unit in a CIC (even short-term rentals) does not render the “use” of that unit “commercial” rather than residential. E.g., Kiekel v. Four Colonies Homes Ass’n, 162 P.3d 57 (Kan. Ct. App. 2007); Lowden v. Bosley, 909 A.2d 261, 266 (Md. 2006); Kaufman v. Fass, 756 N.Y.S.2d 247 (App. Div. 2d Dep’t 2003); Scott v. Walker, 645 S.E.2d 278 (Va. 2007). Leasing and occupancy restrictions are clearly restraints on alienation of the right to possess, not a restriction on property use. A residential occupant, no matter what her race and regardless of whether she holds legal title or a leasehold interest, possesses and uses the property in the same way as another residential occupant. To the extent leasing is a use, it is but a use of the landlord’s investment capital. The actual use of the property turns on how it is enjoyed and employed by the party in possession. This concept is explained in greater depth in Boyack, Community Covenant Alienation Restraints, supra note 33.

236 See Brower, supra note 20, at 204 (discussing the broad scope of CIC governing provisions, including behavior inside homes). There are in-home
limited by a modern substantive “touch and concern” requirement in servitude law.

The best method to sort out which covenants are proper and which are overreaching is to adopt a twenty-first century updated “touch and concern” test. This test would require economic justification for communal governance schemes rather than focusing on the amorphous concept of relating to the land. If a given covenant acts to remediate a cost externality—such as a nuisance or an aspect of the “tragedy of the commons”—then characterizing that provision as a servitude would be justified. An agreement among neighbors that does not address a cost externality, however, should not be elevated to the status of a real covenant running with the land, regardless of the authors’ intent. Rather, such neighborhood agreements that are not economically justified should be mere personal contracts, analyzed and enforced as such. Any non-covenant provisions of a neighborhood agreement may (if they meet the formation requirements of contract law) create in personam obligations among the contracting parties. And the breach of these obligations would give rise to a claim for contract damages. But these terms would not run with the land nor would they be specifically enforceable. This approach would preserve the value of community covenants without allowing either the CIC structure or the “touch and concern” limitations on covenant-making to unduly encroach onto residents’ autonomy.

C. Solving the Constitutional Conundrum

Private regulation of certain personal freedoms generates popular outrage. Courts have upheld association restrictions on behaviors that may generate cost externalities. One example is smoking. See, e.g., Ezra, supra note 151.

free speech, but public opinion backlash has been substantial. CIC restrictions on religious displays and practices have generated critical scholarship. And rights of persons to privacy and autonomy within their own homes have been fervently defended. Although “constitutional” violations are often asserted by discontented CIC members, absence of state action is usually fatal to such claims. Constitutional jurisprudence with respect to CICs is a bit of a mess—emotional outliers make for bad law—and Supreme Court precedents can be misleading. This has led to disparate state law treatment of personal freedoms in community covenant contexts. The tension in the law with respect to constitutional freedoms and CIC functions needs to be resolved.

Proposed solutions to the CIC constitutional conundrum fall into two general categories. One approach is to treat CIC associations as if they were public government units, thereby


238 Midlake on Big Boulder Lake Condo. Ass’n v. Cappuccio, 673 A.2d 340, 350 (Pa. Super. Ct. 1996) (holding that the owners “contractually agreed to abide by the provisions in the Declaration at the time of purchase, thereby relinquishing their freedom of speech concerns regarding placing signs on this property”).


242 See Hyatt & French, supra note 2, at 114–55; Hyatt, Common Interest Communities, supra note 67, 338–42.
giving residents protection through the First Amendment and other constitutional rights against community interference.\textsuperscript{245} This approach is problematic and creates worrisome precedents, as evidenced by the substantial judicial resistance to analogizing private groups to public actors.\textsuperscript{246} The second, and preferable, approach is a legislative solution—enact a “Bill of Rights” for homeowners in CICs.

Professor Susan French was among the first to suggest a homeowners’ bill of rights solution to the constitutional governance gap in CICs.\textsuperscript{247} Professor French conceived of this quasi-constitutional guaranty of personal freedoms as being a provision included in the governing documents of the CIC.\textsuperscript{248} More recently, groups have called for state legislatures to enact a homeowners’ bill of rights that would apply to all CICs in the state.\textsuperscript{249} The National Conference of Commissioners on Uniform


\textsuperscript{246} See, e.g., Hudgens v. NLRB, 424 U.S. 507, 516 (1976) (holding that private property can only be treated as if it were public “when the property has taken on \textit{all} the attributes of a town” (emphasis in the original)); Illinois Migrant Council v. Campbell Soup, 574 F.2d 374 (7th Cir. 1978) (illustrating how difficult it is to prove that private property has “all” aspects of a town); Flagg Bros. v. Brooks, 436 U.S. 149 (1978) (explaining that state regulation alone does not constitute state action); NCAA v. Tarkanian, 488 U.S. 179 (1988) (holding that regulatory power over an entity does not render acts of that entity susceptible to Constitutional scrutiny); S.F. Arts & Athletics, Inc. v. U.S. Olympic Comm., 483 U.S. 522 (1987) (same). For discussions of the limits of state action application to private communities, see G. Sidney Buchanan, \textit{A Conceptual History of the State Action Doctrine: The Search for Governmental Responsibility}, 34 \textit{HOUS. L. REV.} 333 (1997); Katharine Rosenberry, \textit{The Application of the Federal and State Constitutions to Condominiums, Cooperatives and Planned Developments}, 19 \textit{REAL PROP. PROB. & TR. J.} 1 (1984).


\textsuperscript{248} \textit{Id.}

\textsuperscript{249} The AARP is promoting a Bill of Rights for Homeowners Associations. The proposed Bill of Rights includes “the right to resolve
COMMUNITY COVENANT CONTRACT MYTH

State Laws has considered including a bill of rights for homeowners in CICs as part of its UCIOA revision. State legislators could add great value by undertaking to identify and guaranty important individual rights in the context of private CIC governance. Statutory protection could solve the issue of to what extent CIC governance can be analogized to public governance. Creating special legislative protection for owners in CICs would not only address the most emotionally charged topics of CIC regulation (and siphon off the hard cases that make bad law) but would also bring clarity to the contentious issue of constitutional applicability to CIC governance.

CONCLUSION

Commentators and courts routinely consider the purchase of a home in a CIC as a conscious, voluntary choice to be bound by the applicable neighborhood covenants. Based on this assumption, CIC covenants and rules promulgated thereunder are treated presumptively enforceable, just like any other contract. The realities of home-purchasing decisions and the CIC creation process cast significant doubts on this approach. Although courts claim that in enforcing CIC covenants they are upholding neighborhood desires, in fact, the terms of community covenants may not necessarily be expressions of community preference. The original form of community covenants are imposed by developers at the direction of municipalities and mortgage market actors, not elected by the residents themselves. Furthermore, CIC restrictive covenants are perpetual, mandatory, non-negotiable requirements of owning a home in the community. And even if buyers actually know, understand, and accept the content of recorded covenants at the time of purchase, the content of neighborhood rules may

disputes without litigation,” the right to be informed of any changes to the rules, and “the right to oversight of associations and directors.” For a summary of the proposed bill, see A Bill of Rights for Homeowners in Associations, IN BRIEF (AARP Pub. Policy Inst., Washington, D.C.), July 2006, at 1–2.

thereafter change in ways unforeseeable by a purchaser and essentially unconstrained by courts or constitutions. Members can opt out of this system of private regulation—but only by selling their home.

The solution to the contract-covenant disconnect is to recognize that recorded CC&Rs that impose neighborhood obligations are not, in fact, simple contracts. CIC governance is founded on and impacts three areas of the law: contracts, property, and constitutional governance. The proper judicial conception of CIC covenants and rulemakings, then, must draw upon all three of these areas by requiring a bona fide manifestation of assent to be bound, by appropriately limiting the substantive scope of neighborhood covenants, and by protecting homeowner rights from governmental overreaching.

First, a higher consent threshold is vital. In the context of CIC covenants, the contractual temporal autonomy paradox is augmented. Recorded declarations are non-negotiable contracts of adhesion, and as such, it is unlikely that buyers—by the mere act of purchase alone—have truly, voluntarily consented to the obligations. A CIC homebuyer is not a “Ulysses,” deliberately choosing to be bound in order to limit future action (for his own benefit).251 Rather, a CIC homebuyer is bound without her deliberate election and is subject to terms she has no hand in crafting and no choice but to accept. Her supposed manifestation of assent is the purchase of a home, and she cannot buy that particular piece of property without acquiescing to the imposed terms.

In other contexts, lack of buyer input with respect to adhesive contract terms is rendered less objectionable because market prices and market choices reflect general consumer preferences and values among varying options.252 But in the context of CICs,  

251 In The Odyssey, Ulysses tied himself to the ship’s mast in order to restrain himself from reacting to the sirens’ song. See Winokur, supra note 105, at 50 (explaining contractual obligation with reference to this metaphor).

252 While limitations on autonomy may be value-deterring, most theorists, courts and developers see a counterbalance in the ability of owners to have input into controlling the autonomy of their neighbors in turn. In addition, the CIC ownership structure permits shared amenity upkeep that makes such amenities, and perhaps homeownership in general, more
the lack of variation among CIC forms and the lack of non-CIC housing choices in several parts of the country undermine these market checks. In addition, a homebuyer usually comparison-shops with respect to the real property and not with respect to associated covenant terms. Furthermore, CIC covenant terms may not even be made available to or reviewed by a buyer until closing (if at all). Providing a copy of CC&Rs only at closing renders homebuyer “consent” specious. At residential home closings, the homebuyer lacks both time and the benefit of counsel to assist in navigating the often lengthy and complicated CIC declaration, bylaws, and associated rules. Even when a purchaser is aware of the content of the applicable CIC covenants prior to closing, it is still pure fiction to claim that the owner manifests her “choice” to be obligated thereunder when she closes the home purchase. A homebuyer chooses \textit{the property} and merely acquiesces to associated covenants, most likely without even knowing or understanding what these covenants require.

Combatting lack of true homeowner assent must be supplemented by limitations on CIC covenant scope and legislative protections of homeowner rights. These protections are necessary because (a) the homeowners are not the authors of community covenant content, and (b) the impact of overreaching covenants extends far beyond the impact of overreaching terms in contracts.

Unlike most contracts, governments and government-related entities shape the content of CIC declarations to a far greater degree than do preferences of the contracting parties—here the neighborhood residents. To obtain zoning approval, developers craft CC&Rs that address municipal priorities, such as creating affordable. \textit{See Nahrstedt v. Lakeside Vill. Condo. Ass’n, 878 P.2d 1275, 1282 (1994)} (“Notwithstanding the limitations on personal autonomy that are inherent in the concept of shared ownership of residential property, common interest developments have increased in popularity in recent years, in part because they generally provide a more affordable alternative to ownership of a single-family home.”).

\textit{253 See Franzese & Siegel, supra note 50, at 1121 (“[I]t is difficult to conceive of a more heavy-handed public interference in the private marketplace than a government rule or practice that mandates a highly particularized form of governance on new housing development.”).}
privately funded community amenities and upkeep. To create communities that will qualify for FHA insurance and GSE secondary market purchases, developers include provisions to meet enumerated underwriting criteria, such as limitations on the percentage of non-owner-occupants in a neighborhood. When the CC&Rs are crafted and recorded, it is the desires of these authorities that influence their content, not the theoretical and unarticulated preferences of unidentified future buyers. This fact alone argues for the implementation of some “bill of rights” type of protection for the parties who are thus governed.

In addition, unlike typical contracts, CIC covenants presumptively exist in perpetuity. The durability of covenants makes it vital to reconsider subject matter limitations on CIC governance and spheres of homeowner protection. Covenants should not be permitted to achieve in perpetuity every end that would be achievable among original contracting parties. To run with the land, a covenant should be justified by an economic need—the problem of incompatible uses, negative externalities, or free-riding, for instance. Only when covenant content supports the legitimate function of CIC governance should the covenant be enforceable as a servitude and not a mere personal contract.

Third, in addition to mandating a higher threshold for owner consent and judicially limiting the scope of servitude provisions, states should act to protect important owner and occupant rights through legislation. Consent alone cannot protect future

---

254 See id. at 1112 (asserting that “government policy aimed at load-shedding municipal functions and services onto newly created CICs” drives the content of CC&R); see also Siegel, The Constitution and Private Government, supra note 106 (claiming that governments dictate CIC formation and content); supra notes 64, 72, 149, 151 and accompanying text (discussing the concept of privatization of public function).

255 See supra notes 152–57 and accompanying text. The Restatement takes the position that the only permissible leasing restrictions should be those required by institutional lenders. RESTATEMENT (THIRD) OF PROP.: SERVITUDES § 6.7 cmt. b (2000).

256 Franzese & Siegel, supra note 50, at 1113; Grassmick, supra note 2, at 212.

257 Although some early-generation CICs and CICs in Louisiana do have expiration dates, most CIC covenants today continue indefinitely unless terminated by supermajority (sometimes unanimous) vote. See supra note 126.
generations of CIC owners from being bound by the value judgments of today.\textsuperscript{258} For example, Professor Korngold explained that even though proponents of perpetually enforceable servitudes argue that dead hand control is rendered unobjectionable by adequate notice,\textsuperscript{259} “this begs the question of whether the deprivation of individual opportunity and autonomy is itself ‘fair.’”\textsuperscript{260} For example, notice of a racial segregation covenant would not justify its enforcement.\textsuperscript{261} Similarly, notice that a covenant regime exists prior to purchase of a property in a neighborhood should not necessarily justify the enforcement of private regulations that impinge on individual rights or are unjustifiable based on economic exigency. Notice is not synonymous with choice.

Subject matter constraint is also warranted because servitudes are specifically enforceable; an owner cannot choose to pay expectation damages rather than comply.\textsuperscript{262} A breach, even numerous breaches, of an obligation does not terminate the restriction. And although a supermajority of owners can amend or perhaps even terminate CIC restrictions, these options are cumbersome and practically difficult to achieve.\textsuperscript{263} When it comes

\textsuperscript{258} Dead hand control is perhaps the “most compelling reason” that courts should be wary of treating freedom of contract as dispositive in determining servitude enforceability. Korngold, \textit{Privately Held Conservation Servitudes}, supra note 27, at 457.


\textsuperscript{261} Korngold, \textit{Privately Held Conservation Servitudes}, supra note 27, at 457.

\textsuperscript{262} \textit{See supra} notes 195–99 and accompanying text.

\textsuperscript{263} \textit{See} Winokur, \textit{supra} note 105, at 35–37 (explaining the practical difficulties involved in amending CIC covenants). Several state enabling statutes provide that a CIC can only be dissolved through unanimous vote of the members. \textit{E.g.}, ARIZ. REV. STAT. ANN. § 33-556 (Supp. 1964); MASS.
to CIC obligations, opting out of particular covenants is not a possibility and neither is exit by breach. The only way to escape obligations imposed by a CIC regime is to transfer ownership or mobilize a sufficient number of community members to vote for covenant revisions.\footnote{CICs are plagued with participation problems that transcend even issues of lack of participation in democracies generally. See generally David C. Drewes, Note, \textit{Putting the “Community” Back in Common Interest Communities: A Proposal for Participation-Enhancing Procedural Review}, 101 \textit{COLUM. L. REV.} 314, 315 (2001); Ross Thomas, Note, \textit{Ungating Suburbia: Property Rights, Political Participation, and Common Interest Communities}, 22 \textit{CORNELL J.L. & PUB. POL’Y} 205 (2012).} Some CICs require near unanimity to change or eliminate the governance regime, and this poses a collective action problem that grows with the size of the subject community.\footnote{See Steven A. Ramirez, \textit{The Special Interest Race to CEO Primacy and the End of Corporate Governance Law}, 32 \textit{DEL. J. CORP. L.} 345, 383–84 (2007) (concluding that the collective action problem increases with group size); Sterk, \textit{supra} note 10, at 617 (explaining the problem of holdouts and collective action costs in the context of CIC amendment).}

There is a clear disconnect between freedom of contract ideals and the realities of CIC covenant formation. Reflexive enforcement of CIC governing provisions based on contract principles perpetuates the myth of knowing consent by owners to be bound to these provisions. The reality of CIC covenant creation suggests that courts and legislatures should take a more proactive approach to protect owners from covenant overreaching and balancing competing public policies. True manifestation of knowing assent to CIC governance—covenant terms \textit{and} governing processes—should be prerequisite to buying into a community. And the law should impose subject matter limitations on the scope of CIC governance, both through limiting what obligations can become servitudes and by legislatively protecting important individual rights.
ACCOMMODATING EMPLOYERS' INTERESTS INTO THE DISCUSSION OF EMPLOYMENT PROTECTIONS FOR VICTIMS OF DOMESTIC VIOLENCE

Timothy John Durbin*

INTRODUCTION

The landscape of employment law is shifting as states increasingly pass legislation that requires employers to afford special treatment to employees who are victims of domestic violence (“victim-employees”).¹ New Jersey and California provide salient examples of this shifting landscape. On July 17, 2013, New Jersey Governor Chris Christie signed the New Jersey Security and Financial Empowerment Act ("NJ SAFE Act"),²

* J.D. Candidate, Brooklyn Law School, 2015; B.A., Virginia Commonwealth University, 2009. I would like to thank my friends, family, and roommates for their unwavering support and encouragement while I wrote and edited this Note. I am grateful to the editors and members of the Journal of Law and Policy for their wisdom, suggestions, and revisions.

¹ See Beth P. Zoller, Domestic Violence Victims Emerging As A New Protected Class, JDSUPRA (Mar. 15, 2013), http://www.jdsupra.com/legalnews/domestic-violence-victims-emerging-as-a-42895/. This Note considers only employment protections for employees who are themselves victims of domestic violence. In fact, the question is far more complicated than that. Many of the laws herein discussed have been amended to apply to employees whose family member is a victim of domestic violence. See, e.g., Or. Rev. Stat. § 659A.270(2) (2011) (as amended by 2013 Or. Laws Ch. 321 (H.B. 2903) (June 6, 2013) (originally applying to a “parent or guardian of a minor child or dependent who is a victim of domestic violence”)). Also, many of the laws herein discussed apply with equal force to victims of a broader swath of intimate partner violence. See, e.g., Haw. Rev. Stat. § 378-71 (2014) (defining covered employees as victims of “domestic abuse, sexual assault, or stalking.”). These distinctions likely have large impacts on the employers who are covered by these laws, but this Note does not seek to address those issues.

² New Jersey Security and Financial Empowerment Act of 2013, P.L.
which requires employers to provide any victim-employee up to twenty days of unpaid leave within a calendar year. On October 10, 2013, California Governor Jerry Brown signed California Senate Bill 400 (“S.B. 400”), which prevents employers from discriminating against victim-employees and requires employers to reasonably accommodate victim-employees. The New Jersey and California bills exemplify the three current state-level approaches to providing employment protections to victim-employees: (1) requiring a statutorily defined amount of unpaid leave (“the statutory leave approach”); (2) preventing discrimination on the basis of status as a victim of domestic violence (“the antidiscrimination approach”); or (3) requiring employers to reasonably accommodate victim-employees (“the reasonable accommodation approach”).

2013, c. 82.


5 S.B. 400, 2013–14 Cal. S., Reg. Session (Cal. 2013) [hereinafter Cal. S.B. 400]. Cal. S.B. 400 added to the Labor Code that “[a]n employer shall not discharge or in any manner discriminate or retaliate against an employee who is a victim of domestic violence,” and “[a]n employer shall provide reasonable accommodations for a victim of domestic violence.” Id.

Thirty-four states have passed legislation under one of these three approaches. As such, victims of domestic violence are rapidly becoming another subset of the population that the law sees as deserving of special treatment. The commonality between each of the three approaches is that they each require employers to provide some form of employment protection—or special treatment—to victim-employees. The primary goal of these statutes has been to provide job security for victims of domestic violence so that they can become financially independent from, and eventually, leave their abusers. While this goal is noble, it begs the question: do these statutes impose too great of a cost on employers?

To date, employers and their advocates have opposed these employment protections. Many commentators have criticized this wholesale opposition because domestic violence is not a purely domestic problem and instead directly affects employers’ bottom line. Although employers do have a valid reason to

---

Domestic Violence at Work Creating A Societal Response by Making Businesses A Part of the Solution, 74 Brook. L. Rev. 377, 392–95 (2009) (discussing trends in state laws). Each of the current state-level approaches was modeled off of existing federal law: the statutory leave approach is modeled after the Family Medical Leave Act; the antidiscrimination approach is modeled after Title VII of the Civil Rights Act; and the reasonable accommodation approach is modeled after the Americans with Disability Act. See Deborah A. Widiss, Domestic Violence and the Workplace: The Explosion of State Legislation and the Need for a Comprehensive Strategy, 35 Fla. St. U. L. Rev. 669, 700, 707, 709 (2008).

7 See LEGAL MOMENTUM, STATE LAW GUIDE, supra note 6, at 1–36.
8 Zoller, supra note 1.
9 See infra notes 26–29 and accompanying text.
10 See, e.g., Letter from California Chamber of Commerce et al. to Members, California Senate Committee on Judiciary (Apr. 19, 2013) [hereinafter California Chamber of Commerce, Letter in Opposition], available at http://www.calodging.com/images/uploads/pdfs/SB_400_Sen_Jud_Oppose.pdf (opposing SB 400 on the behalf of employers). See also Karin, supra note 6, at 398 (discussing specific opposition to this type of legislation).
11 See, e.g., Karin, supra note 6, at 383–85. One possible explanation for employer opposition is that it is simple path dependence. This theory asserts that because employers have opposed employment protections in the past—like unpaid leave for pregnant workers—they blindly continue to do so now. Professor Deborah Widiss provides two examples of how path dependence
oppose employment protections, since such protections can increase the cost of business, overall those protections are in fact beneficial. When protections are drafted with employers’ interests in mind, they can trade small-term costs for long-term gains.

Employers should also recognize that wholesale opposition is no longer an effective strategy for a number of reasons. First, there has been an “explosion” of state-level statutes that provide employment protections for victim-employees. Over two thirds of the states have some form of this statute. Second, the federal government has recently signaled two ways that it supports providing protections for victims of domestic violence: (1) on February 15, 2013, the Office of Personal Management (“OPM”) required all federal agencies to prohibit discrimination of federal employees on the basis of their status of domestic violence; and (2) federal agencies have issued guidance for how existing federal laws can provide protection to victim-employees in limited factual scenarios. Rather than continue their strategy of wholesale opposition, employers may be operating in this area of law. See Widiss, supra note 6, at 705, 708–09. First, Widiss argues that “proposing domestic violence victim status as an additional protected class predisposes businesses to oppose such bills.” Id. at 708. Second, Widiss describes how the invocation of a statutory leave requirement raises the specter of the Family and Medical Leave Act (“FMLA”) and “triggers an assumption” that, like the FMLA, these leave laws should exempt smaller businesses. Id. at 702–05.

12 Hearing on S.B. 229 Before the S. Comms. on Labor & Pub. Emp’t and Hum. Servs., 2011 Leg. (Haw. 2011) (letter testimony of Poka Laenui, Exec. Dir., Wai’anae Coast Comm. Mental Health Ctr.), available at http://www.capitol.hawaii.gov/session2011/testimony/SB229_SD1_TESTIMONY_LAB-HUS_03-22-11_.PDF. In his letter opposing a 2011 Hawaii Bill that amended the state labor code to require employers to afford victims of domestic violence reasonable accommodation, Poka Laenui, Executive Director of the Wai’anae Coast Community Mental Health Center, argues that “this bill [will be used to] transfer upon employers cost of doing business the burden of underwriting what is essentially a social-criminal and financial societal issue.” Id.

13 Widiss, supra note 6, at 669, 698.

14 See LEGAL MOMENTUM, STATE LAW GUIDE, supra note 6, at 1–36.

15 See infra Part I.B for a discussion of OPM’s new policy.

16 Part II of this Note discusses how the U.S. Department of Labor and the Equal Employment Opportunity Commission have issued fact sheets
opposition to employment protections, employers and their advocates should support limited protections for victim-employees. Such protections can and should be drafted in a manner that ensures that they address employer concerns, while imposing only minimal costs on those employers.

The New York State Senate provides an excellent example of how, on the state level, employers could shape the debate around employment protections for victims of domestic violence. During its 2013 legislative session, the New York Senate considered two bills.  

The first, Bill 2509, would require employers “[t]o permit victims of domestic or sexual violence to take [up to 90 days] unpaid leave . . . to address on-going domestic or sexual violence issues.” The second, Bill 3385, on the other hand, would only require employers to provide a reasonable amount of leave to victim-employees, provided that it does not cause an undue burden on their operation. It seems likely that employers would rather provide unburdensome leave than three months of leave. This single legislative session presents a lucid example of how employer advocacy could mean the difference between providing...
reasonable leave that is not overly burdensome or being required to provide up to three months of leave.

This Note creates a blueprint for a fine-tuned law that decreases the economic impact of domestic violence, normalizes the cost of providing protection, and requires only low-cost solutions. While other scholars have taken the position that some statutory approaches present a better solution for employees and employers, none have thus far endeavored to catalog the state-level statutes to show which statutes are deserving of employer support. Examination of how state statutes have dealt with the issue so far demonstrates that a federal law could be drafted so as to alleviate many of employers’ concerns while accommodating their economic interests.

Part I describes how domestic violence affects employers’ bottom line, and how consequently, employers have an economic interest in employment protections for victims of domestic violence. Part I also asserts that internal policies are insufficient to deal with the costs of domestic violence because such policies unfairly create extra costs for proactive employers. Part II shows how existing federal law ignores employers’ interests while providing victim-employees with protection. Part III explores the statutory leave approach and discusses how key statutory provisions can solve employer concerns without great cost. Part

—

20 See, e.g., Widiss, supra note 6, at 718–23. Widiss advocates for a targeted and comprehensive approach that pairs sets of legislation to address the goals of employees, employers, and the general public. Id. Widiss provides three examples of such pairings: first, coupling a law that would make status as a victim of domestic violence a protected classification with workplace restraining orders, id. at 720; second, drafting unpaid leave and unemployment issuance statutes to provide protections based on employer size, id. at 720–21; and third, using unemployment issuance statutes in conjunction with reasonable accommodation statues so that when an accommodation would represent an undue burden on the employer, the employee is still protected, id. at 721. See also Elissa Stone, Comment, How the Family and Medical Leave Act Can Offer Protection to Domestic Violence Victims in the Workplace, 44 U.S.F. L. Rev. 729, 736 (2010) (advocating for the statutory leave approach); but cf. Lisalyn R. Jacobs & Maya Raghu, The Need for A Uniform Federal Response to the Workplace Impact of Interpersonal Violence, 11 Geo. J. Gender & L. 593, 607 (2010) (arguing that the statutory leave approach is inadequate on its own).
IV discusses the weaknesses of the antidiscrimination approach. Part V explores the reasonable accommodation approach and discusses the ways the approach inherently addresses employers’ interests and concerns. Part VI concludes with a summary of the statutory provisions that employers should consider supporting.

I. EMPLOYERS HAVE AN ECONOMIC INTEREST IN EMPLOYMENT PROTECTIONS FOR VICTIM-EMPLOYEES AND INTERNAL POLICIES ARE INADEQUATE TO CORRECT THE PROBLEM.

Prior to a discussion about which statutory approaches would best address employer interests, two questions must be answered: First, do employers have an interest in addressing domestic violence in the lives of their employees? Second, why are employers unable to address this problem internally?

A. Domestic Violence Directly and Indirectly Affects Employers’ Bottom Line

The first question—should employers really support employment protections for victims of domestic violence?—may seem callous, but the answer should not be treated as a foregone conclusion. Clearly, domestic violence affects every aspect of a victim’s life.21 Domestic violence is—at root—about the abuser controlling, dominating, and coercing the victim.22 Domestic violence can take the form of economic control,23 or it can start at


22 Schneider, supra note 21, at 356.

23 Id.
home and then spill into the victim’s work life. 24 It is estimated that “a staggering twenty-nine percent of male and forty percent of female workers report[,] having been subjected to intimate partner violence at some point in their lives.”25

While criminal codes have long protected victims’ physical safety, there have been fewer legal solutions for the ways that domestic violence affects a victim’s ability to attain or maintain a job.26 Over time, states have adopted laws to address the economic instability in the lives of those affected by domestic violence, primarily by providing victims with employment protections.27 A driving theory behind these statutes is that by facilitating job retention, victims are more likely to become financially independent,28 and thereby more likely to end the cycle of violence in their lives.29

24 Jessie Bode Brown, The Costs of Domestic Violence in the Employment Arena: A Call for Legal Reform and Community-Based Education Initiatives, 16 Va. J. Soc. Pol’y & L. 1, 21–24 (2008). Brown discusses how domestic violence can spill into the workplace in the following ways: the victim-employee misses work “because of the abuse [he or she] faces at home;” the abuser calls the victim-employee while he or she is at work; the abuser is a coworker who abuses the victim-employee at work; or the abuser shows up at the victim-employee’s workplace and perpetrates violence there. Id. at 17. There is also ample evidence that domestic violence affects a victim-employee’s productivity while at work. See infra notes 42–60 and accompanying text.

25 Jacobs & Raghu, supra note 20, at 597.

26 See Brown, supra note 24, at 2.


28 See Widiss, supra note 6, at 675–76.

29 See, e.g., WASH. REV. CODE § 49.76.010 (2013) (stating, in the legislative findings of the Washington domestic leave law, that “[o]ne of the best predictors of whether a victim of domestic violence, sexual assault, or stalking will be able to stay away from an abuser is his or her degree of
While it is clear that violence directly and indirectly affects victims’ lives, it is less intuitive that employers are similarly affected. In fact, domestic violence can directly affect employers, as evidenced by the tragic murder of Zina Haughton. When Zina’s husband, Radcliffe Haughton, became increasingly violent at home, and even came to her work and slashed her tires, she filed a restraining order. Four days later, Radcliffe again came to Zina’s work and started a fire inside the workplace with a propane tank. He then shot his estranged wife and six other women, killing a total of four, including himself. While some scholars assert that incidents like these are “relatively uncommon,” when polled, the overwhelming majority of corporate security officials report concern that domestic violence directly threatens the safety of their workplaces. Workplace incidents like this can lead to bad press and massive tort liability

---

33 Id.
34 Id.
35 See, e.g., Widiss, supra note 6, at 686. But see Stacy M. Downey & Amy Johns, New Study Examines the Role of Intimate Partner Violence in Workplace Homicides Among U.S. Women, Ctrs. for Disease Control & Prevention, http://www.cdc.gov/niosh/updates/upd-05-03-12.html. (citing a National Institute for Occupational Safety and Health (NIOSH) study finding that the second cause of violent death for female employees while working is homicide perpetrated by an intimate partner or relative); Swanberg & Logan, supra note 21, at 14 (“The majority of women from this study reported that their abusers had shown up at work at some point during the last 2 years.”).
Domestic violence can also directly increase the employer’s business costs even when the violence occurs elsewhere. An abuser will often attempt to weaken the victim’s economic independence by disrupting his or her job performance. According to one study fifty-six percent of victims report being harassed while at work by their abuser. It has been observed that an abuser’s controlling behavior may actually impact victim-employees’ productivity more than the violence itself. Qualitative studies of how domestic violence affects the workplace reveal examples of abusers “physically restraining a partner from going to work . . . , making a car unavailable, or cutting up work clothes” and “inflict[ing] visible injuries, reneg[ing] on promises to provide child care, or keep[ing] the

37 See John E. Matejkovic, Which Suit Would You Like? The Employer’s Dilemma in Dealing with Domestic Violence, 33 CAP. U. L. REV. 309, 312–34 (2004). Essentially, claims against the employer would sound in negligence, id. at 313, under the theory that employers owe employees and patrons a duty to protect them from foreseeable danger, id. at 314. These actions can yield six-digit price tags for employers. Id. at 313 (“Because the employer had been warned of the husband’s threats and the employer did not beef up security, the jury awarded the plaintiffs $5 million.”).


40 Swanberg & Logan, supra note 21, at 14 (“Women’s responses implied that stalking at work caused significant levels of stress and psychological discomfort that in turn significantly affected job performance. In fact, the content of respondents’ discussions that focused on the ramifications associated with being stalked while at work led us to surmise that the abusers’ stalking behavior produced more anxiety and stress for women than actual physical actions taken by abuser prior to work.” (emphasis added)). This can occur for one of two reasons. First, the employee is concerned that his or her abuser will terrorize him or her at work. Id. Second, the employee’s productivity is lessened because of the residual stress from an earlier incident. See Carole Warshaw et al., Mental Health Consequences of Intimate Partner Violence, in INTIMATE PARTNER VIOLENCE: A HEALTH BASED PERSPECTIVE 147, 150, 161 (Connie Mitchell & Dierdre Anglin, eds., 2009).

41 Goldscheid, supra note 30, at 75 (citing Swanberg & Logan, supra note 21, at 6–8).
victim up late at night the day before a critical event like an exam or a meeting.”

These kinds of activities not only hurt the employee, but the bottom line of the victim’s employer as well. The Bureau of National Affairs estimated that, as a result of domestic violence, employers lose between $3 and $5 billion due to lost employee productivity. Others estimate that employers lose up to $13 billion per year in profits because of domestic violence.

Domestic violence also has indirect effects on an employer’s business costs. These indirect effects include detrimental employee health, absenteeism, and turnover. The 2010 National Intimate Partner and Sexual Violence Survey, promulgated by the Center for Disease Control (“CDC”), found that domestic violence victims suffered health effects beyond the scrapes and bruises of violence, including “frequent headaches, chronic pain, difficulty with sleeping, ... asthma, irritable bowel syndrome, and diabetes.” Studies document that victims suffer escalating emotional trauma and thus require psychiatric care. Two commentators estimate that employers pay up to hundreds of millions of dollars in health care costs because of the adverse health effects caused by domestic violence. In 2003, the CDC found that female victims “lose nearly 8.0 million days of paid

42 Calaf, supra note 38, at 171 (citations omitted).
44 Id.
45 See Brown, supra note 24, at 24; see also AM. INST. ON DOMESTIC VIOLENCE, http://www.aidv-usa.com/ (last visited Mar. 27, 2014).
47 See Warshaw et al., supra note 40, at 148–50. It is also important to note the escalating nature of emotional damage; the crippling mental and emotional effects of domestic violence in turn make the victim more susceptible to worsening emotional abuse as well as less likely to seek available resources. Id. at 149–50, 155–56.
work each year,” which “is the equivalent of 32,114 full-time jobs each year.” This pervasive absenteeism disrupts the workplace and the employer’s operations.

Domestic violence also causes high employee turnover. This can occur in several ways. First, “homicide is a leading cause of death in the workplace.” Over a quarter of these murders are committed by an intimate partner or close relative.

Second, domestic violence affects every aspect of the victim’s life that makes it difficult to maintain employment. Third, managers see victim-employees’ pervasive absenteeism as disruptive to the workplace, which, without employment protections in place, can lead to the employee’s termination.

The high rates of employee turnover associated with domestic violence increase employers’


50 Congress reached the same conclusions after hearings for the Violence Against Women Act. See Brief for Arizona and Thirty-Seven Other States as Amici Curie Supporting Petitioner, United States v. Morrison, 529 U.S. 598 (1999) (No. 99-5), 1999 WL 1032809, at *5 (“Congress found that violence against women imposes significant costs on employers by increasing absenteeism, lowering productivity, increasing health care costs, and creating higher turnover.”); id. at *7 (“Congress also found that employers have responded to the effect on ‘such bottom line issues as tardiness, poor performance, increased medical claims, interpersonal conflicts in the workplace, depression, stress and substance abuse’ by directly addressing domestic violence in order to reduce their costs and protect their employees.”) (quoting Domestic Violence: Hearing on S. 596 Before the Sen. Comm. on the Judiciary, 103d Cong. 15 (1993) (statement of James Hardeman, Polaroid Corp.)).

51 See id. at *6.

52 Id. at *6 n.3 (citing AFSCME & AFL-CIO, Hidden Violence Against Women at Work, WOMEN IN PUB. SERV., Fall 1995, at 1).


54 See Randel & Wells, supra note 48, at 823.

business costs as “well-trained employees are a valuable asset and . . . training new employees is more costly than retaining a productive and knowledgeable existing workforce.”

Domestic violence’s effect on employers extends beyond the disruption it causes to the particular victim; it also affects the productivity of victims’ coworkers. For example, coworkers surveyed in Pennsylvania reported that victim-employees more often came to work late, left early, and took frequent breaks. About half of these coworker employees report that they felt compelled to cover the work of victims and/or abusers, and a similar percentage felt that “company resources . . . [were used] to deal with or deliver . . . abuse.” Thus, domestic violence causes decreased productivity in the victim-employees and their coworkers.

In sum, domestic violence has powerful direct and indirect economic effects on the victims’ workplace. Due to these economic effects employers have an economic interest in solutions that provide employees with the opportunity to end the cycle of violence in their lives.


Many employers are already aware of the effect that domestic violence has on their workplace. In 2007, the Corporate

56 See Randel & Wells, supra note 48, at 823.
58 Id. at 7.
59 Id. at 7–8.
60 Id. at 8.
61 See Randel & Wells, supra note 48, at 823.
62 820 ILL. COMP. STAT. 180/5(24) (2012). In passing the Illinois Victims’ Economic Security and Safety Act, the General Assembly found, “[49%] of senior executives recently surveyed said domestic violence has a harmful effect on their company’s productivity, 47% said domestic violence negatively affects attendance, and 44% said domestic violence increases health care costs.” Id. See also Randel & Wells, supra note 48, at 826, 829–34
Alliance to End Partner Violence documented corporate opinions of the effect of domestic violence on the workplace and the economy. The study revealed that sixty-three percent of polled CEOs considered domestic violence a major problem, and forty-three percent of CEOs and ninety-one percent of employees reported an effect on the company’s bottom line.

Recognizing the negative impact domestic violence has on their businesses, some employers have begun to adopt internal policies to assist their employees who are victims. This includes the nation’s largest employer—the federal government. On April 12, 2012, President Barack Obama issued a Memorandum calling on the federal government to provide protection for federal victim-employees. This Memorandum directed the Office of Personnel Management (“OPM”) to issue guidance to each federal agency to modify its internal policies so as to address the effects of domestic violence. Accordingly, on February 15,
2013, the OPM directed federal agencies to develop or modify internal policies to address the effects of domestic violence on federal agencies. OPM’s directive included guidance on how these policies should be developed and enacted. OPM directed every policy to consider the need to provide leave, the importance of nondiscrimination against victim-employees, the need for training and awareness, and the role of safety and building security. The directive also directed that “to the greatest extent possible, agencies should work in collaboration with the employee to provide leave and/or other workplace flexibilities to help the employee remain safe and maintain his or her work performance.” The new federal protection is in line with similar protections for employees of state and local governments. President Obama stated that these protections should “act as a model” for private sector employers.

Not all employers need the federal government to provide a model; many companies have adopted internal policies that provide protections to victim-employees. These policies include facility safety improvements, internal counseling, and external referrals for the victims. Verizon Wireless in their Employee against or by employees, guidelines for assisting employee victims, leave policies relating to domestic violence situations, general guidelines on when it may be appropriate to take disciplinary action against employees who commit or threaten acts of domestic violence, measures to improve workplace safety related to domestic violence, and resources for identifying relevant best practices related to domestic violence.”

69 U.S. OFFICE OF PERSONNEL MGMT., GUIDANCE FOR AGENCY-SPECIFIC DOMESTIC VIOLENCE, SEXUAL ASSAULT, AND STALKING POLICIES 1–3, 8 (Feb. 2013).

70 Id. at 8–10.

71 Id. at 11–27.

72 Id. at 11.

73 See generally LEGAL MOMENTUM, STATE LAW GUIDE, supra note 6.

74 See Presidential Memo on Domestic Violence, supra note 67.

75 See, e.g., Runge, Employment Needs Update, supra note 21, at 13–14.

Code of Business Conduct, for example, declares that domestic violence is a workplace issue and encourages employees to come forward with their domestic violence issues.\textsuperscript{77} This simple policy change can serve as “the starting point for fostering trust.”\textsuperscript{78}

Other examples of employer policies include State Farm permitting flexible work hours and special time-off policies for victims of domestic violence\textsuperscript{79} and Cigna holding an annual “Worksite Violence/Partner Violence Month.”\textsuperscript{80} The Corporate Alliance to End Partner Violence surveyed effective employer policies and suggested policies that “[d]efine a policy for flexible work hours” and “[c]onsider what special accommodations may be able [available] for victims.”\textsuperscript{81} In addition to adopting internal policies, industry leaders have also formed advocacy groups such as Employers Against Domestic Violence\textsuperscript{82} and the Corporate

\begin{footnotesize}
\begin{enumerate}
\item Id.
\item 2007 Corporate Alliance Survey, supra note 63.
\item Employers Against Domestic Violence is a Massachusetts alliance of corporate partners, victim rights organizations, and governmental agencies, “[c]ommitted to proactively addressing the causes and effects of violence in the workplace.” Who We Are, Emp’rs. Against Domestic Violence, http://employersagainstdomesticviolence.org/about/who-we-are/ (last visited Mar. 27, 2014). It includes business leaders, such as Blue Cross Blue Shield of Massachusetts, the Boston Red Sox, John Hancock Financial Services Liberty Mutual Group, Massachusetts General Hospital, Verizon Wireless, and many others. See Membership Organizations, Emp’rs. Against Domestic Violence, http://employersagainstdomesticviolence.org/member-organizations/ (last visited Mar. 27, 2014).
\end{enumerate}
\end{footnotesize}
Alliance to End Partner Violence\textsuperscript{83} to collectively advocate for protections for victims of domestic violence.

These internal policies and advocacy groups indicate that some employers are beginning to recognize that they have an economic interest in ensuring there are adequate protections for victim-employees.\textsuperscript{84} One such solution, which some employers are enacting, is accommodations for victims.\textsuperscript{85} The new protections for federal employees should be viewed as a shift in federal policy, which could have important ramifications for businesses.\textsuperscript{86} Given the growing consensus in corporate and government sectors, it is not difficult to imagine a federal law that treats domestic violence as a workplace issue and requires employers to bear additional costs.\textsuperscript{87} It would therefore be shrewd of employers to decide now which employment protection best accommodates their interests without imposing greater costs.

\textsuperscript{83} The Corporate Alliance to End Partner Violence is comprised of business leaders, such as Chase Bank, Lincoln Mutual Life Insurance Co., the National Football League, Prudential, State Farm Insurance Companies, and the Target Corporation. Our Members, CORP. ALLIANCE TO END PARTNER VIOLENCE, http://www.caepv.org/about/members.php (last visited Mar. 27, 2014).

\textsuperscript{84} Dougan & Wells, supra note 76, at 2. Of course, employers are not unanimous in their recognition that domestic violence is a workplace issue. See Meg Hobday, Domestic Violence Comes to Work: The Need for A Work-Related Response, 67 BENCH & BAR OF MINN., no. 3, Mar. 2010, at 20, 22 (“According to a 2005 Survey of Workplace Violence Prevention, only 29.1 percent of businesses have policies addressing workplace violence generally, and less than half of those address domestic violence specifically.”). Many employers still oppose treating domestic violence as a workplace issue. Parts III, IV, and V address some of the arguments that employers have raised in opposition to treating employment.

\textsuperscript{85} Both the State Farm policy and the recommendation provided by the Corporate Alliance to End Partner Violence discussed above are examples of this. See 2007 CORPORATE ALLIANCE SURVEY, supra note 63; State Farm Best Practices, supra note 79.

\textsuperscript{86} See Parker & Dorn, supra note 17.

\textsuperscript{87} See generally Karin, supra note 6 (discussing how reframing the issue as a workplace issue is more likely to garner public support, the support of the business community, and the support of the federal government).
C. Legislation Would Spread the Cost of Protecting Employees.

The prevalence of these organizations and internal policies begs the salient question: do employers need a law to tell them to do what they are already doing? Allowing employers to solve domestic violence issues with internal policies alone presents several problems. First, the majority of employers have no internal policies.\(^{88}\) Second, studies show that even when internal policies exist, employees are often unaware of them.\(^{89}\) Third, there is compelling evidence that employers tend to underestimate the prevalence of domestic violence in their staff.\(^{90}\) Fourth, adopting a policy may cause the employer to incur costs that might decrease its competitiveness in the market.\(^{91}\) Finally, smaller employers are often without the resources to adopt internal policies to deal with the problems of domestic violence.\(^{92}\)

The last two issues presented by only having internal policies are of greatest concern for employers. First, an employer who provides her employees a protection that is not required by statute is incurring an additional business cost.\(^{93}\) As is discussed below, this cost will generally be minor,\(^{94}\) but it should not be ignored. At times, employers adopt internal policies out of a natural sense of wanting to protect employees.\(^{95}\) An employer that adopts a


\(^{89}\) See 2007 CORPORATE ALLIANCE SURVEY, *supra* note 63, at 6 (finding that “72\% of executives say their companies offer programs and services that address domestic violence but less than half of employees (47\%) are even aware of this fact”).

\(^{90}\) See Widiss, *supra* note 6, at 682.

\(^{91}\) See Goldscheid, *supra* note 30, at 120.

\(^{92}\) See Karin, *supra* note 6, at 418.

\(^{93}\) See Goldscheid, *supra* note 30, at 120 (discussing the costs associated with accommodation statutes).

\(^{94}\) See id. (“[These] costs generally will be modest.”); see also infra notes 238–48 and accompanying text.

\(^{95}\) See, e.g., Betsy Weintraub, *The Hidden Safety Hazard—Domestic Violence*, FISHER & PHILLIPS LLP (Nov. 1, 2012),
policy that affords extra accommodations to victim-employees incurs a cost that is in addition to industry-wide costs of business. As such, the employer is potentially penalized in the market for what is essentially a moral decision.

The second concern—that smaller resource-poor employers are unable to introduce domestic violence policies—is related. Smaller employers have a double bind: either adopt internal policies, which they cannot afford, or continue to allow domestic violence to affect their bottom line. Thus, where there is no statutory employment protection for victims of domestic violence, employers who adopt an internal domestic violence policy and small employers are penalized by the unregulated marketplace. Professors Marcy Karin and Deborah Widiss argue that the issue should be framed as a public health issue in which all of society has a stake. If you look at domestic violence through this lens it allows you to view a putative federal employment law as a common cost of business that all employers should pay, not merely those that elect to or can afford to. Legislation would redistribute the cost of protecting victims and make employment protections a cost of doing business.

II. EXISTING FEDERAL EMPLOYMENT LAWS ONLY HAVE A LIMITED APPLICATION TO THE ISSUES RAISED BY DOMESTIC VIOLENCE’S EFFECT ON THE WORKPLACE.

The problem should now be clear: domestic violence is not merely a domestic problem; instead, it directly and indirectly

(advising employer clients to adopt policies because “[a]s a human being, [they] have an even bigger responsibility to watch out for dangers to your employees . . .”).

96 See Goldscheid, supra note 30, at 120.
97 See Karin, supra note 6, at 418.
98 Karin, supra note 6, at 399–400 (“Making this change would broaden the conversation on this issue beyond the current focus on expanding protections for victims . . .”); Widiss, supra note 6, at 693–94 (suggesting an alternative to a cost-benefit analysis).
99 Widiss, supra note 6, at 685–86.
affects employers’ bottom line.\textsuperscript{100} While some employers are able to address domestic violence with internal policies, those policies represent an extra cost of business.\textsuperscript{101} Further, resource-poor businesses are unlikely to assume the cost of broad employment protections in light of an unregulated market.\textsuperscript{102} There are also likely shortsighted employers that would rather avoid the short-term costs of an employment policy than reap the benefit of ameliorating the effect of domestic violence on their workplace. Given these concerns, scholars have asserted that employers have an interest in federal legislation that normalizes the cost of providing victim-employees with some form of protection.\textsuperscript{103} The question then becomes how should the law address these issues?

There is a valid question as to whether this issue should be resolved at the state or federal level.\textsuperscript{104} So far, legislation has only been passed at the state level.\textsuperscript{105} In recent years federal bills have been introduced into both houses of Congress, but these attempts have not progressed.\textsuperscript{106} If employers and their advocates were to begin to advocate for limited employment protections for victims of domestic violence, they would need to determine whether their

\begin{footnotesize}
\begin{itemize}
\item \textsuperscript{100} See supra Part I.A.
\item \textsuperscript{101} See supra notes 88–96 and accompanying text.
\item \textsuperscript{102} See, e.g., Widiss, supra note 6, at 726 (arguing that in an unregulated market—one without a statute that requires an employer to provide some form of protection to employees—“certain costs are too great for at least some employers to bear”); see also supra notes 97, 98 and accompanying text.
\item \textsuperscript{103} Widiss, supra note 6, at 726 (“Recognizing the larger public interests at stake, legislatures should consider public funding, or other cost-spreading mechanisms, to supplement costs that they deem unreasonable for either individual employers or individual employees to bear as a result of a perpetrator of domestic violence’s criminal actions.”).
\item \textsuperscript{104} See Karin, supra note 6, 379 (arguing that a federal statute is better suited to address the problem).
\item \textsuperscript{105} Parts III, IV, and V of this Note describe the statutes that have been passed at the state level.
\item \textsuperscript{106} See Robin R. Runge, The Evolution of A National Response to Violence Against Women, 24 HASTINGS WOMEN’S L.J. 429, 453 (2013) (describing early efforts to have the Violence Against Women Act include employment protections for victims of domestic violence); Widiss, supra note 6, at 703 n.113 (describing more recent federal bills that would require employers to grant employment protections for victims of domestic violence).
\end{itemize}
\end{footnotesize}
interests are better served by a federal or state statute. This Note avoids that question and instead focuses only on the issue at a federal level. As the 2013–14 New York state legislative session shows, no matter the forum, because of the different approaches for dealing with the issue, employers have room to advocate for some laws instead of others.

Federal law can address the issue either by applying existing law to victims of domestic violence or with new legislation. The first solution—to apply or amend existing federal employment protection laws so that they cover victims of domestic violence—is inadequate. Over the last several years, scholars and commentators have theorized how existing federal law such as Title VII of the Civil Rights Act of 1964 (“Title VII”), the Americans with Disabilities Act (“ADA”), and the Family Medical Leave Act (“FMLA”) can be interpreted or amended to provide protections for domestic violence. Scholars have generally concluded that as these laws are currently drafted, they are incapable of addressing the myriad of issues presented by domestic violence in the workplace.

107 Cf. Karin, supra note 6, at 379–80, 397–98, 399–400, 428. Karin advocates that a “federal law would set a national standard to address a national problem.” Id. at 397.

108 See supra notes 17–19 and accompanying text.


112 See, e.g., Denise R. J. Finlay, Employment Discrimination Against Domestic Violence Survivors: Strengthening the Disparate Impact Theory, 88 N.D. L. REV. 989 (2012) (arguing discrimination against victims of domestic violence is actionable under disparate treatment theory of Title VII because domestic violence is gendered violence); Stone, supra note 20, at 736 (2010) (calling for an amendment to FMLA that would trigger the same entitlement to leave for an incident of domestic violence as the birth or adoption of a child).

113 See, e.g., Goldscheid, supra note 30, at 123 (arguing that even when Title VII is seen as protecting victims of domestic violence from gender-stereotype based discrimination, Title VII as it has been interpreted is incapable of protecting against subtle forms of gender bias which drive many of the adverse employment actions taken against victim-employees).
A survey of case law shows that when courts attempt to apply existing federal law to the issues presented by domestic violence, those laws are inadequate. For example, in *O’Donnell v. Gonzales*, the U.S. District Court of Massachusetts permitted an employee of the Federal Bureau of Prisons to survive summary judgment on an ADA claim that alleged the employer’s failure to accommodate the victim’s Post-Traumatic Stress Disorder (“PTSD”) and depression. Plaintiff-employee had produced documentation both of her status as a victim of domestic violence and her diagnosis of PTSD and depression. Her abuser was a coworker and ex-paramour. When the Bureau found out about the violence, management assigned the victim-employee to a 6:00 AM-to-2:30 PM shift and the abuser-employee to a 4:00 PM-to-midnight shift. Considering this insufficient, and after an incident where her abuser spray-painted threats and derogatory statements outside of plaintiff’s work area, plaintiff requested accommodation for her depression and PTSD in the form of either an unspecified period of leave or the removal of her ex-paramour from the work force. The court held that an unspecified period of leave was unreasonable, but that plaintiff had established a question of material fact as to whether, under the ADA, it was reasonable for one employee to request the removal of another employee.

The *O’Donnell* case provides an example of how incompatible the ADA may be with issues of domestic violence. Without a statute that expressly limits a court’s discretion of what constitutes a reasonable accommodation in the particular context of domestic violence, employers could be found liable under the ADA for failure to drastically reorganize their workforce.

---

115 *Id.* at *11.
116 *Id.* at *2.
117 *Id.* at *1.
118 *Id.* at *2.
120 *Id.* at *8.
121 *Id.* at *8, *10–11.
122 *Cf.* CAL. LAB. CODE § 230(f)(2) (West 2014) (requiring state
reasonable accommodation statute better serves employers when it is drafted to clearly outline what employers are required to do to reasonably accommodate victims of domestic violence.

More recently, the agencies that enforce federal employment statutes—the Equal Employment Opportunity Commission (“EEOC”)123 and the United States Department of Labor (“U.S. DOL”)124—have issued specific guidance for when existing federal laws may trigger employer liability.125 Neither of these agency Fact Sheets creates new law or extends existing equal employment law; instead, they discuss the application of existing doctrine to fact patterns that include victims of domestic violence.126

On November 1, 2012, the EEOC issued a Fact Sheet describing certain employer actions that could incur liability under Title VII and/or the ADA.127 The EEOC cautions against actions such as termination of a woman because of the “drama [she may] bring to the workplace” or rejection of a qualified male applicant who, as a victim, is viewed as weak.128 The Fact Sheet next provides examples of how workplace sexual harassment of and sexual violence against victims of domestic violence could

---

123 The federal agency that enforces both the ADA and Title VII of the Civil Rights Act of 1964.

124 The federal agency that enforces FMLA.


127 EEOC Fact Sheet, supra note 125.

128 Id. at *1–2.
incur employer liability under Title VII.\textsuperscript{129} The Fact Sheet also describes how certain employer actions can violate the ADA, such as passing over an applicant after learning of her counseling because of prior violence, or failing to intervene when coworkers tease an employee on account of his violence-related scars.\textsuperscript{130} Finally, the Fact Sheet describes how failure to provide reasonable accommodation to a victim-employee may incur liability under the ADA. Examples include when an employee “requests a schedule change or unpaid leave” on account of depression resulting from former violence.\textsuperscript{131} However, the Fact Sheet fails to address whether courts may, as the \textit{O'Donnell} court did, consider the scope of reasonable accommodation and find fault with companies who fail to provide large-scale accommodations, such as staff reorganization. As it is written, the EEOC Fact Sheet provides only factual instances that may give rise to a claim, but it avoids the more difficult question of what constitutes the reasonable accommodation of victim-employees.

The U.S. DOL has issued similar guidance for when and how the FMLA can be used by victims of domestic violence in limited factual scenarios.\textsuperscript{132} This guidance states, “FMLA leave may be available to address certain health-related issues resulting from domestic violence.”\textsuperscript{133} However, only “serious health conditions” are likely to trigger entitlement to leave,\textsuperscript{134} and thus the FMLA

\begin{footnotesize}
\begin{enumerate}
\item\textsuperscript{129} Id. at *2. This application comes as no surprise, as Title VII has long been held to prohibit sexual harassment of and sexual violence against any employees. \textit{See, e.g.}, \textit{Meritor Savs. Bank v. Vinson, 477 U.S. 57, 73 (1986).}
\item\textsuperscript{130} EEOC Fact Sheet, supra note 125, at *2.
\item\textsuperscript{132} See DOL FAQ, supra note 125, at 10.
\item\textsuperscript{133} Id. Just as with the EEOC fact sheet, this provides examples of applicability in the context of victim-employees: “[A]n eligible employee may be able to take FMLA leave if he or she is hospitalized overnight or is receiving certain treatment for post-traumatic stress disorder that resulted from domestic violence.” \textit{Id.}
\item\textsuperscript{134} See 29 U.S.C. §§ 2612 (a)(1)(A)–(D) (2009) (listing the reasons for leave, with “serious health condition” being the only applicable reason). Serious health conditions include only injuries that require hospitalization or
\end{enumerate}
\end{footnotesize}
only has a limited applicability to the issues raised by domestic violence.\textsuperscript{135} Domestic violence affects employers in direct and indirect ways, and FMLA’s application to domestic violence will likely cover only the most egregious examples of direct effects.\textsuperscript{136} It is unlikely that even “broken wrist[s] or black eye[s]” will be covered by the most liberal interpretation of FMLA.\textsuperscript{137} Furthermore, the FMLA cannot be easily read to permit leave to separate from or bring criminal actions against an abuser.\textsuperscript{138} Even at its most broad, the FMLA appears incapable of providing leave to attend court hearings or stay at a domestic violence shelter.\textsuperscript{139} Thus, the FMLA is incapable of providing leave for the employee determined to break the cycle of abuse. Employers looking for a federal solution to the indirect effects of domestic violence—like absenteeism and diminished productivity—must look beyond the FMLA.

Review of these Fact Sheets shows two things. First, employers should take note that there appears to be a shift in federal policy toward greater protections for victims of domestic violence. Second, these Fact Sheets create no new law and merely apply existing doctrine to limited factual scenarios. For years commentators have agreed that the application of existing federal laws to domestic violence will fail to address many issues caused by domestic violence in the workplace.\textsuperscript{140} The Fact Sheets issued by the agencies that enforce these statutes confirm these concerns because the Fact Sheets provide solutions to only a very limited number of factual scenarios. Instead of allowing

\textsuperscript{135} See Hobday, supra note 84, at 22 (arguing that “none of the current federal employment laws directly applies to [victim-employees]”).

\textsuperscript{136} See Stone, supra note 20, at 737 (positing that even “a broken wrist or black eye” may not be serious enough to meet the statute’s requirement of a serious medical condition).

\textsuperscript{137} Id. at 737.

\textsuperscript{138} Id.

\textsuperscript{139} See, e.g., Stone, supra note 20, at 736–37 (arguing that, because the only way for a victim to trigger a right to leave under FMLA is with a serious medical condition, “employees are not entitled to the leave that could allow them the time to take the first corrective steps in leaving their abusive partner”).

\textsuperscript{140} See supra notes 107, 113 and accompanying text.
employers and domestic violence victims to continue to suffer under current inadequate federal law, Congress should pass new legislation that is more directly targeted toward limiting domestic violence’s effect on the workplace.\textsuperscript{141}

This legislation could be modeled after one of the current statutory approaches that exist at the state level: the leave approach, the antidiscrimination approach, and/or the reasonable accommodation approach.\textsuperscript{142} The remainder of this Note explores the strengths and weaknesses of these approaches, as well as specific ways that statutes can be drafted so as to ensure that employers’ concerns are being addressed without too great of a cost.

III. THE STATUTORY UNPAID LEAVE APPROACH

A. The Approach Generally

At the state level there are a variety of leave laws that may apply to victim-employees. By far the most ubiquitous of these statutes are “crime leave laws.”\textsuperscript{143} These laws, present in thirty-three states, require employers to permit leave to victims of crimes.\textsuperscript{144} These laws, which were not enacted to specifically

\textsuperscript{141} See Widiss, supra note 6, at 672; see also Runge, Employment Needs Update, supra note 21, at 23 (chronicling legislative attempts to craft new federal legislation that would apply directly to victims of domestic violence).

\textsuperscript{142} This Note explores only these three approaches. For a discussion of other approaches, including employer protection orders, unemployment benefits for employee-victims, and the common-law exception to the at-will doctrine finding that the termination of a victim-employee because of her status violates public policy, see generally Karin, supra note 6.

\textsuperscript{143} Runge, Employment Needs Update, supra note 21, at 13, 15.

\textsuperscript{144} These states are Alabama, ALA. CODE § 15-23-81 (2014), Alaska, ALASKA STAT. § 12.61.017 (2012); Arizona, ARIZ. REV. STAT. ANN. § 13-4439 (2012); Arkansas, ARK. CODE ANN. § 16-90-1105 (LexisNexis 2014); California, CAL. LAB. CODE 230.2(b) (West 2014); Colorado, COLO. REV. STAT. § 24-4.1-303(8) (2014); Connecticut, CONN. GEN. STAT. ANN. § 54-85(b) (2012); Delaware, DEL. CODE ANN. tit. 11, § 9409 (2014); Florida, FLA. STAT. § 741.313 (2013); Georgia, GA. CODE ANN. § 34-1-3 (2014); Hawaii, HAW. REV. STAT. § 621-10.5 (2014); Iowa, IOWA CODE § 915.23 (2012); Maryland, MD. CODE ANN. CRIM. PROC. § 11-102 (2011);
protect victims of domestic violence, are imperfect. Crime leave laws only provide victim-employees with leave to attend court, and therefore cannot be used to secure leave to deal with other domestic violence issues such as seeking medical care or finally new housing.

Currently, thirteen states offer specific protections to victims of domestic violence by requiring their employer to permit victims to take leave to address economic, social, and psychological problems created by domestic violence situations. The unpaid leave statutes vary in their applicability to employers and employees. Some statutes apply, by their terms, to any conceivable employer. Other statutes limit jurisdiction to either


145 Runge, Employment Needs Update, supra note 21, at 13, 15.
146 See generally LEGAL MOMENTUM STATE LAW GUIDE, supra note 6.
147 Runge, Employment Needs Update, supra note 21, at 15.
149 See generally LEGAL MOMENTUM STATE LAW GUIDE, supra note 6.
150 See generally id.
151 See, e.g., N.M. STAT. ANN. § 50-4A-2(D) (2012) (“Employer’
a broad group of employers\textsuperscript{152} or only larger employers.\textsuperscript{153} Some statutes limit their applicability to employees who have been employed for a certain amount of time, though the majority of statutes include most employees within the jurisdiction of their leave laws.\textsuperscript{154} Other statutes include no jurisdictional hooks, and therefore appear to apply to every employer-employee relationship.\textsuperscript{155}

Unpaid leave statutes also vary greatly with regard to the maximum amount of time an employer must grant a victim-employee. Some statutes require employers to only grant three days,\textsuperscript{156} while others require close to two weeks.\textsuperscript{157} Maine, North...
Carolina, Oregon, and Washington do not place a specific cap on the amount of unpaid leave they require an employer to permit but rather require the employer to grant a “reasonable amount of time.” Kansas, on the other hand, sets neither a numerical nor a reasonable cap on the amount of leave an employee may take.

B. Potential Concerns with the Approach and Possible Solutions

The statutory leave approach presents several potential problems for employers but there are rejoinders to each of these problems.

The most intuitive concern is that employees will abuse the law and take leave for unrelated reasons. Some statutes have effectively addressed this concern with two interconnected provisions. The first permits leave only for a finite list of tasks such as: obtaining relief from the legal system, seeking medical

hours in one day” for “up to fourteen days in any calendar year”).

158 See ME. REV. STAT. tit. 26 § 850(1) (2007) (requiring “reasonable and necessary leave from work”); N.C. GEN. STAT. § 50B-5.5 (2012) (prohibiting discrimination against an employee who “took reasonable time off from work”); OR. REV. STAT. § 659A.272 (2014) (“a covered employer shall allow an eligible employee to take reasonable leave” for enumerated reasons); WASH. REV. CODE § 49.76.030 (2013) (requiring “reasonable leave from work, intermittent leave, or leave on a reduced leave schedule”); see also GUAM CODE ANN. 22-3-3401–3405 (2013) (requiring a reasonable amount of time). The benefits of this approach are discussed in the next section.

159 See KAN. STAT. ANN. § 44-1132 (2013) (“An employer may not discharge or in any manner discriminate or retaliate against an employee who is a victim of domestic violence or a victim of sexual assault for taking time off from work to [list of permissible reasons].”). The lack of statutory clarity in how much time is required begs the question of what an employer would be required to provide.

160 See Robin R. Runge, Redefining Leave from Work, 19 GEO. J. ON POVERTY L. & POL’Y 445, 480 (2012) (“One of the primary complaints of employers in opposition to the FMLA was the cost of hiring, training, and maintaining staff to ensure that the reasons that employees were requesting to take FMLA leave were permitted under the statute.”); see also Haase, supra note 55, at 351 (describing how the “argument is unrealistic in the case of unpaid leave” (emphasis added)).

161 See, e.g., FLA. STAT. § 741.313(2)(b)(5) (2013) (permitting leave to
attention related to injuries, seeking services from a victims’ services organization, seeking new or safer housing, and receiving counseling. The list is essentially the same in each law that enumerates permissible reasons. The second provision requires the employee to provide certification that the leave was for a permissible reason. Where an employee attempts to abuse

“seek legal assistance in addressing issues arising from the act of domestic violence”); KAN. STAT. ANN. § 44-1132(a)(4) (2013) (permitting leave to “make court appearances in the aftermath of domestic violence”); N.M. STAT. ANN. § 50-4A-2(B) (2012) (permitting leave “to obtain an order of protection or other judicial relief . . . or to meet with law enforcement officials, to consult with attorneys or district attorneys’ victim advocates or to attend court proceedings”).

See, e.g., COLO. REV. STAT. § 24-34-402.7(1)(a)(II) (2013) (permitting employee to “[o]btain[] medical care or mental health counseling . . . to address physical or psychological injuries.”); OR. REV. STAT. §659A.272(2) (2014) (permitting employee to “seek medical treatment for or to recover from injuries”).


See, e.g., COLO. REV. STAT. § 24-34-402.7(1)(a)(III) (2013) (permitting the employee leave to either seek new housing or “mak[e] his or her home secure from the perpetrator”); 820 ILL. COMP. STAT. 180/20(a)(1)(D) (2012) (permitting “participating in safety planning, temporarily or permanently relocating, or taking other actions to increase . . . safety”); OR. REV. STAT. §659A.272(3) (2014) (permitting leave “to relocate or take steps to secure an existing home”).


See, e.g., CAL. LAB. CODE § 230.1(b) (West 2014) (permitting employers to require notice before leave, or upon employee’s return, a police report, court order, or certification by medical, legal, or rights organization professional, that tends to show that advanced notice was not feasible); HAW. REV. STAT. §§ 378-72(b)–(d) (2014) (permitting employers to require
accommodating employers’ interests

statutory leave laws, the employer will likely not be prohibited from disciplinary action.\textsuperscript{168} As long as a statute only grants leave for an enumerated list of tasks and requires the employee to prove that the leave was correctly used, employers should not expect extensive employee dishonesty.

Another concern that employers have with the leave approach is that it forces employers to lose productive hours from their employees.\textsuperscript{169} But scholar David Haase correctly points out that, for family and medical leave, in the long run the opposite may actually be true.\textsuperscript{170} When no leave requirement is in place, managers make decisions regarding leave requests based on short-term disruptions of work.\textsuperscript{171} However, when a leave requirement is in place the potential for shortsighted management decisions is taken off the table, and employees are granted leave for important life events, increasing their loyalty and productivity over time.\textsuperscript{172} This argument is particularly salient when discussing domestic violence. As discussed earlier, domestic violence causes loss of productivity in victim-employees who are absent from or stalked at work.\textsuperscript{173} The ability to take leave permits victim-employees to secure economic independence from their abusers,\textsuperscript{174} and therefore increases their long-term productivity and full potential over time.\textsuperscript{175} Further, leave laws can be drafted so as to require

certification from a medical care professional, attorney, or employee of victims’ rights organization).

\textsuperscript{168} For example, in \textit{Sustata v. Shannon}, 966 N.E.2d 365 (Ill. App. Ct. 2012), the Appellate Court of Illinois for the Second District held that an employer could terminate a victim-employee after she was unable to provide documentation that her leave was for one of the statutorily enumerated reasons. It is important to note that the court reached this holding even in the face of arguably vague statutory language. \textit{Id.} at 371.

\textsuperscript{169} A similar criticism was made against the FMLA. Haase, \textit{supra} note 55, at 349 (noting that “it may not make economic sense for an employer to make sacrifices for employees”).

\textsuperscript{170} \textit{Id.}

\textsuperscript{171} \textit{Id.}

\textsuperscript{172} \textit{Id.}

\textsuperscript{173} \textit{See supra} Part I.A.

\textsuperscript{174} \textit{See supra} Part I.A.

\textsuperscript{175} \textit{See} Reynolds v. Fraser, 781 N.Y.S.2d 885, 889 (N.Y. Sup. Ct. 2004) (“The ability to hold on to a job is one of a victim’s most valuable weapons in
victim-employees to provide advanced notice\textsuperscript{176} so that the employer can adjust resources. Thus, there are several ways that leave laws can be drafted so as to alleviate employers concern about lost productive hours.

Another concern is that in order to comply with unpaid leave statutes, employers must expend additional administrative resources to keep more thorough records in order to document: which employees are victims, how many days those employees have taken, and what the leave was used for.\textsuperscript{177} There are two rejoinders to this concern of administrative costs. First, empirical evidence suggests that the costs of administrating leave laws are minimal.\textsuperscript{178} In 1991, when President George H.W. Bush first vetoed the FMLA based, in part, on his concern over increased administration costs,\textsuperscript{179} critics correctly countered that the concern was not compelling because in states that had already enacted

\textsuperscript{176} See, e.g., COLO. REV. STAT. § 24-34-402.7(2)(a) (2013) (permitting employers to condition leave on “the appropriate advance notice of such leave as may be required by the employer’s policy”). Some statutes waive the notice requirement when leave is not foreseeable. See, e.g., OR. REV. STAT. § 659A.280 (2014). Colorado and Florida will not allow employers to require notice “in cases of imminent danger to the health or safety of the employee.” COLO. REV. STAT. § 24-34-402.7(2)(a) (2013); FLA. STAT. § 741.313(4)(a) (2013). An even more refined solution to the issue can be found in the newly enacted California Labor Code. See CAL. LAB. CODE § 230.1 (West 2014). California requires advance notice, see id. § 230.1(b)(1), unless advance notice is not feasible at which point the employer may subsequently require certification that the unscheduled leave was appropriate, see id. § 230.1(b)(2). This certification can be in the form of a (1) police report, (2) a court order, or (3) documentation from a medical professional. See id. §§ 230.1(b)(2); 230(b)(2)(A)–(C).

\textsuperscript{177} See Haase, supra note 55, at 348 (describing employer’s similar concern with the passage of the FMLA). In fact, employers’ concerns about the costs that they would incur as the result of a leave requirement prompted President Bush to veto the Family and Medical Act of 1990. Maria L. Ontiveros, The Myths of Market Forces, Mothers and Private Employment: The Parental Leave Veto, 1 CORNELL J.L. & PUB. POL’Y 25, 25 (1992). Nevertheless, the law was passed in 1993 over these objections. 29 U.S.C. §§ 2601–2654 (1993).

\textsuperscript{178} See Ontiveros, supra note 177, at 31.

\textsuperscript{179} Id. at 25.
family leave requirements, only six percent of employers reported an increase in administrative costs and only four percent reported an increase in training and compliance costs.\footnote{Id. at 31 n.27.} Consider, too, that in terms of a leave law for victim-employees, the administration costs would likely be less because the FMLA requires employers to provide up to twelve weeks of leave,\footnote{29 U.S.C. § 2612(a)(1) (2012). Note, this provision has been held unconstitutional when used in a civil suit against a state employer for monetary damages when the state has not consented to such a suit. \textit{See} Coleman v. Court of Appeals of Md., 132 S. Ct. 1327 (2012).} whereas most leave laws for domestic violence only require between three and fourteen days.\footnote{See, e.g., COLO. REV. STAT. §§ 24-34-402.7(1)(a) (2013) (requiring employers to permit up to three days of leave); CONN. GEN. STAT. ANN. § 31-51ss(b) (2011) (permitting “employer to “limit unpaid leave . . . to twelve days during any calendar year”); FLA. STAT. § 741.313(2)(a) (2013) (requiring “up to 3 working days of leave from work in any 12-month period”); N.M. STAT. ANN. § 50-4A-2(B) (2012) (requiring leave “up to eight hours in one day” for “up to fourteen days in any calendar year”).} It is even possible that the additional administrative costs may be nil, because employers are already required to keep records of leave under the FMLA,\footnote{Haase, \textit{supra} note 55, at 348.} and therefore already have a system for documenting leave.

The second rejoinder to administrative costs is that these laws can be drafted in a way that permits employers to require certification that the employee is a victim and documentation that the leave was for a statutorily defined reason,\footnote{See, \textit{e.g.}, Sustatia v. Shannon, 966 N.E.2d 365, 371–72 (Ill. App. Ct. 2012) (holding that Illinois leave law permitted the employee to certify that she was a victim as well as require her to provide documentation that she used leave for one of the statutorily enumerated reasons).} as well as to require the employee to provide advance notice of the leave.\footnote{See \textit{supra} note 176 and accompanying text (describing notice requirements). California’s approach appears most tailored to the emergent nature of domestic violence and employers’ interests in notice and record keeping). \textit{See} CAL. LAB. CODE §§ 230(d)(2)(A)–(C) (West 2014).} These notice requirements should defer some administrative costs by shifting the record-keeping requirements onto the victim-employee, and thus deal with employer concerns.

Another concern is that smaller employers will be
disproportionately affected by statutory leave requirements. Exempting smaller employers from the leave requirement will solve this issue. The FMLA, for example, exempts employers who employ less than fifty employees. Many of the states that have adopted statutory leave laws have similarly limited their applicability to larger employers, on the basis that these resource-rich employers are more capable of assuming the cost of granting leave. Other states take a different approach and require employers of different sizes to grant different periods of leave. It should also be noted that the “cost” imposed by domestic violence leave laws is almost always less than the twelve-week “cost” imposed by the FMLA. As such, it may not be as important for resource-poor, smaller businesses to advocate for the small-employer exception, as it was when employers raised the argument against the more “costly” FMLA. Regardless, statutory leave laws prove amenable to the concern of disproportionate impact by either exempting smaller employers or by scaling required leave to employer size.

---

188 See supra Part III.A.
189 See Widiss, supra note 6, at 700–05.
190 See, e.g., HAW. REV. STAT. § 378-72(a) (2014) (requiring employers of fifty or more to grant up to thirty days and smaller employers to grant up to five). The problem with this approach is that it seems arbitrary that one employee would create such a difference in requirements. The “reasonable leave” approach, described shortly, may provide a similar, but less arbitrary approach.
191 Only Illinois requires up to 12 weeks of leave. 820 ILL. COMP. STAT. 180/20(a)(2) (2009). As discussed most states require between three days and two weeks. It should be noted that the New York Senate and Assembly are considering companion bills that would require employers to provide up to 90 days of leave to victims. See Nurse, supra note 17; see also A. 7029, 2013 N.Y. Assemb. (Apr. 30, 2013); S. 2509, 2013 N.Y. Senate. (Apr. 15, 2013).
192 Mattis, supra note 186, at 1338. Indeed, this is another example of how path dependence may be guiding employers’ opposition to employment protections for domestic violence even where there are distinct differences between these laws and their predecessors.
Some states have adopted a statutory leave statute that requires employers to grant a reasonable period of leave instead of a statutorily defined period. This hybrid approach was adopted by Hawaii, Oregon, and Maine. The reasonable leave approach presents an alternative solution to the cost-bearing problem. Oregon, for example, requires covered employers to grant “reasonable leave” but also permits employers to “limit the amount of leave if such leave creates an undue hardship on the employer’s business.” As such, Oregon permits an employer’s size to factor into how much leave it is required to provide. Hawaii resolves the proof of reasonableness issue by requiring the employee to submit a statement by a professional that certifies the need for and reasonableness of the requested time.

In sum, the statutory leave approach has proven amenable to employers’ cost-bearing concerns first because often the leave laws are paired with an exemption for smaller resource-poor businesses. Second, by requiring the employee to provide documentation that the leave was taken for an enumerated reason, a leave statute can ensure that the employer is only providing leave for employees who are attempting to resolve the issue of domestic violence in her life. Third, while a leave statute would certainly require the employer to assume the cost of providing leave, the potential long-term gain of permitting leave—allowing the employee to become independent from her abuser and thus

---

196 In fact, the Oregon leave law is a hybrid of the statutory leave approach and the reasonable accommodation approach and includes all the cost-detriment balancing benefits associated with the reasonable accommodation approach described infra Part V.
197 See Haw. Rev. Stat. § 378-72(b) (2014) (“Reasonable period of time’ . . . means: (1) Where due to physical or psychological injury . . . the period of time determined to be necessary by the attending health care provider . . . ; and (2) Where due to an employee’s need to take legal or other actions . . . the period of time . . . [determined necessary] by the employee’s or employee’s minor child’s attorney . . . .”); id. (c)–(d) (permitting employer to require certified documentation from the involved medical professional).
more productive—greatly overshadows the short-term costs. Furthermore, as is the case in Oregon, statutes can provide that an employer is only required to provide reasonable leave that does not present an undue burden. The reasonable leave approach ensures that the short-term costs of leave are less than the long-term benefits. Thus, employers should consider supporting a federal leave requirement. Provided that a leave requirement statute is drafted as described above, a leave requirement would benefit employers without imposing excessive costs.

IV. THE ANTIDISCRIMINATION APPROACH

A. The Approach Generally

The antidiscrimination approach essentially treats victim of domestic violence as a status, and prohibits employers from discriminating against victim-employees in a term or condition of their employment. Only five states, California, Hawaii, Illinois, New York, and Oregon have adopted the antidiscrimination approach. The Illinois Victims’ Economic Security and Safety Act (“VESSA”) provides a detailed example of how the antidiscrimination statutes operate:

An employer shall not fail to hire, refuse to hire, discharge, constructively discharge, or harass any individual, otherwise discriminate against any individual with respect to the compensation, terms, conditions, or privileges of employment of the individual, or retaliate against an individual in any

198 See, e.g., N.Y. EXEC. LAW § 296 (McKinney 2010) (“It shall be an unlawful discriminatory practice: (a) [f]or an employer . . . because of an individual’s . . . domestic violence victim status, to refuse to hire or employ or to bar or to discharge from employment such individual or to discriminate against such individual in compensation or in terms, conditions or privileges of employment.”).

199 CAL. LAB. CODE §§ 230.1(a)–(c) (West 2014).
201 820 ILL. COMP. STAT. 180/30 (2012).
202 N.Y. EXEC. LAW § 296 (McKinney 2010).
ACCOMMODATING EMPLOYERS’ INTERESTS

form or manner . . . because: (1) the individual involved: (A) is or is perceived to be a victim of domestic or sexual violence . . . . 204

However, antidiscrimination statutes are not well suited to addressing the issue of domestic violence. As a preliminary matter, it is worthwhile to note that New York is the only state to have adopted only the antidiscrimination approach; each of the other four states (California, Hawaii, Illinois, and Oregon) buttress their antidiscrimination statute with either a leave requirement or an accommodation statute. 205

B. Problems with the Antidiscrimination Approach

The clearest problem with the antidiscrimination approach is that unlike other protected categories, like race or gender, the status at issue (being a victim of domestic violence) “is a descriptive statement regarding a certain kind of criminal or controlling behavior to which an individual has been subjected.” 206 Having the status of a domestic violence victim is very different than having a status associated with an immutable characteristic such as Native American, woman, Latino, etc. Normally the antidiscrimination approach seeks to prevent employers from allowing certain immutable classifications to determine employment outcomes, even where disparate treatment would be profitable for the employer. 207 Unlike other protected classifications, which are immutable, the status of “victim” is mutable. A victim-employee is capable of attaining financial and

204 820 ILL. COMP. STAT. 180/30(a) (2012).
205 In fact, California, Hawaii, and Illinois have statutes that use all three approaches. See CAL. LAB. CODE §§ 230, 230.1 (West 2014); HAW. REV. STAT. § 378-2 (2014); 820 ILL. COMP. STAT. 180 (2012). Oregon has a reasonable leave requirement in addition to its antidiscrimination statute. OR. REV. STAT. ANN. §§ 659A.290, 659A.885 (2014). See also supra notes 17–19 and accompanying text for a discussion of how New York is considering buttressing its antidiscrimination statute with more functional statutes.
206 Widiss, supra note 6, at 706–07 (referring to the approach as a “strange fit”).
personal independence from her abuser. Indeed, encouraging an employee to become economically independent should be employers’ ultimate goal in advocating for these statutes. Thus, it is contrary to employer interests to advocate for a statute that is tied to status.

An interconnected problem with the antidiscrimination approach is that it is only triggered by an employer’s intentional act of discrimination motivated by animus. A criticism against this narrow focus, raised by employee advocates, is that it allows “savvy employers” to avoid liability as long as they dress their policies in a facially neutral manner. Often, therefore, employers support this approach because it, as compared to other approaches, bears them little to no cost. In the case of domestic violence, however, this “no cost” leads to “no gain.” Recall two earlier discussions: first, domestic violence reduces employee

208 Widiss, *supra* note 6, at 707. In fact, the goal of employment protections is to provide the victim-employee with an opportunity to break the cycle of violence and undo the classification. See, e.g., WASH. REV. CODE § 49.76.010 (2013) (stating, in the legislative findings of the Washington domestic leave law, “[o]ne of the best predictors of whether a victim of domestic violence, sexual assault, or stalking will be able to stay away from an abuser is his or her degree of economic independence”).

209 Goldscheid, *supra* note 30, at 67. Usually this argument is raised in a discussion of the antidiscrimination approach’s inability to deal with subconscious and implicit bias against immutable characteristics such as race or gender. See, e.g., id.


211 Widiss, *supra* note 6, at 696. Consider also that under the current standards of proof articulated by the Supreme Court’s decision in *St. Mary’s Honor Center v. Hicks*, 509 U.S. 502 (1997), an employer may dodge liability for discrimination, even when a former or current employee presents a prima facie case for discrimination, if the employer rebuts “merely by offering reasons which, if true, are nondiscriminatory.” Raymond Nardo, *Évidentiary Issues in Employment Discrimination Litigation*, 9 J. SUFFOLK ACAD. L. 139, 148 (1994). Of course, even if employers have easier standards of proof at trial, this does not avoid the litigation cost of going to trial.
productivity and thus increases costs; and second, employers should support a legal protection because it will externalize the cost of internally assisting the victim in ending the cycle of violence. The antidiscrimination approach is not in line with employers’ long-term interests because it does not require affirmative action and merely prohibits discriminatory acts. Instead, employers should embrace legislation that normalizes the cost of affirmatively assisting employees, which can assist those employees in recognizing their full potential, rather than supporting legislation tied to classification.

Another criticism of the antidiscrimination approach is that it may require employers to illegally infringe on the privacy of their workers. Unlike many statutory leave laws, which require the employee to submit certain documentation, many of the antidiscrimination statutes provide no explanation of how an employer is supposed to know which of their employees are victims. This becomes a problem in states that recognize the constructive knowledge theory of discrimination liability. In the context of domestic violence, an employer may be liable under

---


213 Professor Deborah Widiss takes this argument even further and notes that because the antidiscrimination approach essentially “[e]ncourages employers to treat everyone the ‘same’ . . . [it] could have the unintended effect of discouraging employers from providing employees who are victims of domestic violence with the flexibility they need.” Widiss, supra note 6, at 707.

214 See, e.g., California Chamber of Commerce, Letter in Opposition, supra note 10.

215 See supra note 176 and accompanying text (describing notice requirements).

216 Compare, e.g., N.Y. Exec. Law § 296 (McKinney 2014) (lacking any form of notice or documentation requirement), with the notice provisions described supra Part III.B.

the constructive knowledge theory if there is an adverse employment action (termination, demotion, failure to promote, etc.) after signs that should have put the employer on notice that domestic violence was present in the employee’s life. In other words, under this theory, an employer who should have known that an employee suffered from domestic violence and takes an adverse employment action against that employee can be found just as liable as an employer who intentionally takes the adverse employment action because of the victim’s status. Consider also that under some states’ labor laws, an employer may not inquire into or discriminate against an employee because of his off-duty activities or aspects of his personal life. Essentially this means that when an employer suspects that an employee is a victim, they have two options: discipline the employee for their erratic behavior or attempt to confirm their suspicion to ensure that the employee is properly noted as a member of a protected classification. The first option will incur liability under a constructive knowledge theory and the second will incur liability under a privacy law. This dilemma provides further evidence that the very nature of domestic violence makes it incompatible with an antidiscrimination approach.

218 Id.

219 See id. The California Chamber of Commerce argues that under established case law, “explicit statements regarding discrimination are unnecessary if surrounding circumstances are sufficient to place employer on notice.” Id. (citing Yanowitz v. L’Oreal USA, Inc., 116 P.3d 1123, 1133 (Cal. 2005)). The constructive knowledge theory of employment discrimination is also established in sexual harassment claims. See, e.g., Splunge v. Shoney’s, Inc., 97 F.3d 488, 490 (11th Cir. 1996) (holding that employers only need constructive notice of sexual harassment).

220 See, e.g., N.Y. LAB. LAW §§ 201-d(2)(a)-(d) (McKinney 2014).

221 California Chamber of Commerce, Letter in Opposition, supra note 10.

222 Id.
V. THE REASONABLE ACCOMMODATION APPROACH

A. The Approach Generally

As of January 2013, only four jurisdictions, Illinois,\(^ {223}\) Hawaii,\(^ {224}\) New York City,\(^ {225}\) and Westchester County, New York,\(^ {226}\) had statutes that require employers to reasonably accommodate victims of domestic violence.\(^ {227}\) On October 11, 2013, California became the fifth jurisdiction to adopt the approach.\(^ {228}\) Essentially, reasonable accommodation statutes require an employer to alter working conditions to accommodate an employee who is a victim of domestic violence, provided that the accommodation is reasonable.\(^ {229}\) Hawaii’s Labor Code provides an example of how these jurisdictions have crafted domestic violence reasonable accommodations statutes.\(^ {230}\) Section 378-81(a) provides “[a]n employer shall make reasonable accommodations in the workplace for an employee who is a victim of domestic or sexual violence . . . provided that an employer shall not be required to make the reasonable accommodations if they cause undue hardship on the work operations of the employer.”\(^ {231}\)

\(^{223}\) 820 ILL. COMP. STAT. 180/20 (2012).

\(^{224}\) HAW. REV. STAT. § 378-81(a) (2014).

\(^{225}\) N.Y.C. ADMIN. CODE § 8-107.1(3)(a) (McKinney 2013).

\(^{226}\) WESTCHESTER CNTY. CODE § 700.03(a)(8) (2007).

\(^{227}\) See generally LEGAL MOMENTUM STATE LAW GUIDE, supra note 6.


\(^{229}\) See Vande Zande v. Wis. Dep’t of Admin., 44 F.3d 538, 542 (7th Cir. 1995) (Judge Richard Posner discussing the reasonable accommodation requirement of the ADA).

\(^{230}\) HAW. REV. STAT. § 378-81 (2014).

\(^{231}\) Id. § 378-81(a). Subsection (a) also includes a list of possible reasonable accommodations:

1. Changing the contact information, such as telephone numbers, fax numbers, or electronic-mail addresses, of the employee;
2. Screening the telephone calls of the employee;
3. Restructuring the job functions of the employee;
4. Changing the work location of the employee;
5. Installing
employer to “verify” that the requesting employee is in fact a victim of domestic violence. Subsection (c) defines *undue hardship* as “an action requiring significant difficulty or expense on the operation of an employer” and lists several factors to be considered in the determination of whether a burden is undue. Practically, the way that the reasonable accommodation statutes typically function is thus. First, the employee will verify or document that he is a victim of domestic violence by presenting a restraining order or police incident report. Second, the employee will request a reasonable accommodation, such as a change in work location or a change in phone number because his abuser is harassing him at work. Third, the employer must provide that reasonable accommodation unless it would impose an undue burden on her business.

Employers have raised several arguments against the reasonable accommodation approach, but in fact, this approach is the most amenable to employers’ interests.

locks and other security devices; and (6) Allowing the employee to work flexible hours.

*Id.*  

232 *Id.* § 378-81(b).

233 *Id.* § 378-81(c). The factors to be considered are:

1. The nature and cost of the reasonable accommodation needed under this section;
2. The overall financial resources of the employer; the number of employees of the employer; and the number, type, and placement of the work locations of an employer; and
3. The type of operation of the employer, including the composition, structure, and functions of the workforce of the employer, the geographic separateness of the victim’s work location from the employer, and the administrative or fiscal relationship of the work location to the employer.

*Id.*

234 See *supra* note 176 (discussing documentation requirements).

235 See *infra* notes 251–55, 261 (discussing the process of requesting the accommodation).
B. Problems with the Approach and Possible Solutions

1. Bearing the Cost of Third-Party Behavior

The greatest concern with a reasonable accommodation statute is that it requires the employer to assume costs because of the behavior of a third party, the abuser. However, this cost-bearing issue is actually present in any statute that affords employment protections for employee-victims. Employer costs may seem more salient with a reasonable accommodation statute because the employer is required to enact long-term changes to their business as opposed to quick periods of leave. In reality, the inverse may be true. Statutes that require the reasonable accommodation of victims of domestic violence may be best suited to address employers’ cost-bearing concerns. This is because these statutes can be drafted so as to enumerate specific low-cost accommodations, because these statutes require a dialogue between the victim-employee and the employer, and because these statutes are best able to end the violence in the victim-employee’s life, thus returning him or her to full productivity. Thus, a reasonable accommodation statute can be drafted so as to trade the short-term costs of unburdensome accommodations with the long-term benefit of productivity.

First, unlike the other approaches, the reasonable accommodation approach necessarily considers costs imposed on employers. Reasonable accommodation statutes do this by their very nature “since ‘reasonableness’ is determined by examining the hardship to the employer of providing the accommodation.” Furthermore, reasonable accommodation statutes, including the ADA provision discussed in O’Donnell, often include an

---


237 Id. (suggesting the same with regard to the ADA).

238 Brown, supra note 24, at 34.

239 See 42 U.S.C. § 12112(b)(5)(A) (2012) (requiring employer to provide accommodations “unless such covered entity can demonstrate that the accommodation would impose an undue hardship . . . .”); see also supra notes 114–22.
“undue hardship” exception which prohibits courts and agencies from requiring accommodations that place unreasonable or extensive costs on the employer.\textsuperscript{240} Undue hardship provisions thereby function so as to ensure “focus[ ] on the resources and circumstances of the particular employer in relationship to the cost or difficulty of providing a specific accommodation.”\textsuperscript{241}

One reasonable accommodation statute, the Illinois Victims’ Economic Security and Safety Act (“VESSA”), for example, enumerates several factors that are to be considered in the reasonableness/undue burden analysis: (i) “the nature and cost of the accommodation” requested; (ii) “the [reasonable accommodation’s] effect on expenses and resources, or the impact . . . on the operation of the facility;” (iii) “the overall financial resources of the employer;” and (iv) “the administrative or fiscal relationship of the facility to the employer[s] . . . .”\textsuperscript{242} A provision such as this may help defer some of the fears that reasonable accommodation requirements will disparately impact small businesses. Hawaii, California, and Westchester County, New York provide lists of similar factors to consider in the determination of whether an accommodation is unduly burdensome.\textsuperscript{243} California’s recently amended Labor Code

\textsuperscript{240}See, e.g., HAW. REV. STAT. § 378-81(a) (2014) ("An employer shall make reasonable accommodations in the workplace for an employee who is a victim of domestic or sexual violence, including [list of potential accommodations] provided that an employer shall not be required to make the reasonable accommodations if they cause undue hardship on the work operations of the employer.").


\textsuperscript{242}820 ILL. COMP. STAT. 180/30(b)(4)(B)(i)–(iv) (2012).

\textsuperscript{243}See HAW. REV. STAT. § 378-81(c) (2014) (requiring consideration of: “(1) [t]he nature and cost of the reasonable accommodation needed . . . ; (2) [t]he overall financial resources of the employer; the number of employees of the employer; and the number, type, and placement of the work locations of an employer; and (3) [t]he type of operation of the employer, including the composition, structure, and functions of the workforce of the employer, the geographic separateness of the victim’s work location from the employer, and the administrative or fiscal relationship of the work location to the employer”); see also CAL. LABOR CODE § 230(f)(6) (West 2014) (relying on the definition
provides an additional and important guiding provision: “an undue hardship also includes an action that would violate an employer’s duty to furnish and maintain a place of employment that is safe and healthful for all employees . . . .”244 Such a provision ensures that employers are not required to grant accommodations that would make the workplace unsafe for coworkers of victim-employees.245

In sum, as evidenced by accommodation statutes in Illinois, Hawaii, California, and Westchester, accommodation statutes can protect against unreasonably expensive accommodations provided that they are clearly drafted. If employers are faced with a bill that would require a reasonable accommodation, they should advocate for a provision that, like the corresponding provision in VESSA, requires consideration of the size of the employer as well as the cost of providing the accommodation. Further, employers should advocate for accommodation statutes, like the recently enacted California Labor Code, that clearly provide that they are not required to provide accommodations that would jeopardize the safety of their workplace.

Second, short-term accommodation costs are likely to be slight. Even without a statute requiring employers to reasonably accommodate employee-victims, many employers have adopted internal policies that require management to make certain changes in employment policies for victims of domestic violence.246 The main goal of these internal policies has been to ensure safety on-site and during travel to work, and include “low-to-no cost modifications such as changing phone numbers, worksite locations, shift adjustments, or application of leave policies to domestic or sexual violence related appointments.”247 Essentially, the reasonable accommodation approach is a legal codification of

---

244 CAL. LABOR CODE § 230(f)(6) (West 2014).
245 Id. (exempting any accommodation that would require the employer to violate Section 6400 of California’s Labor Code). See also CAL. LABOR CODE § 6400(a) (West 2014) (“Every employer shall furnish employment and a place of employment that is safe and healthful for the employees therein.”).
246 See, e.g., Randel & Wells, supra note 48, at 829–32.
247 Goldscheid, supra note 30, at 120.
the policies that many employers intuitively adopt.\footnote{248}

Furthermore, employers should advocate for a reasonable accommodation statute that only requires minimal accommodations that are narrowly tailored toward ensuring the employee’s safety and economic independence. This would be accomplished by defining reasonable accommodations to include a finite list of actions. For example, the California’s revised labor code now provides:

[R]easonable accommodations may include the implementation of safety measures, including a transfer, reassignment, modified schedule, changed work telephone, changed work station, installed lock, assistance in documenting domestic violence, sexual assault, or stalking that occurs in the workplace, an implemented safety procedure, or another adjustment to a job structure, workplace facility, or work requirement in response to domestic violence, sexual assault, or stalking, or referral to a victim assistance organization.\footnote{249}

Employers should advocate that a future federal accommodation law would instead read, “Reasonable accommodations shall only include . . . [section] 230(f)(2)’s enumerated list.” This would ensure that employers are only required to assume lost-cost accommodations.

Third, the reasonable accommodation approach, because it requires communication between the victim-employee and his or her employer, is best suited to lessen domestic violence’s effect on employee-productivity. Generally employees are afraid to disclose incidents of domestic violence to their superiors.\footnote{250}

\footnote{248} See, e.g., Randel & Wells, \emph{supra} note 48, at 833 (describing Liz Claiborne’s policy that “allows for flexible hours and time off for employees who need to seek safety and protection, arrange new housing, attend court appearances, or take care of other such matters . . . [as well as] provid[es] secure work areas, special parking spaces, [and] escorts to and from transportation”).

\footnote{249} \textsc{Cal. Labor Code} § 230(f)(2) (West 2014) (emphasis added).

Without legal protection, an employee likely equates disclosure with the risk of termination. If a reasonable accommodation statute requires the employee to request the accommodation and provide documentation of their domestic violence, then the employee is provided with an incentive to break the silence. This is important because then there can be a discussion about the needs of that particular employee and what steps they can take to better their situation. For example, California Labor Law section 230 is amended so as to require the employer to engage “in a timely, good faith, and interactive process with the employee to determine effective reasonable accommodations.”

The interactive process that is part and parcel to the reasonable accommodation approach provides three benefits. First, if an employer is aware of the possibility of violence in the workplace, they can take proactive steps toward avoiding that violence. These steps would likely cost less than dealing with an incidence of violence after it has occurred. Second, when victim-employees inform their supervisors about their situation,

252 It is not unreasonable to predict that employers would be successful if they were to advocate for reasonable accommodation statutes to include provisions that employees prove with certified documents that they are victims of domestic violence. As described supra Part III.B, many of the statutory leave approach laws permit employers to condition the granting of leave on the employees’ ability to certify their status as a victim of domestic violence.
253 See Goldscheid, supra note 30, at 116–17 (“[A]n employee will be more likely to disclose in an environment with a clearly articulated policy that explains the employer’s . . . support for victims . . . .”).
254 Karin, supra note 6, at 408.
255 CAL. LABOR CODE § 230(f)(4) (West 2014). This statutory provision appears to have been inspired by the Federal Code of Regulations’ guidance for ADA reasonable accommodation requirements. See 29 C.F.R. § 1630.2 (2012). The federal code provides “it may be necessary for the covered entity to initiate an informal, interactive process with the individual.” Id. § 1630.2(o)(3).
256 See Jacobs & Raghu, supra note 20, at 604 (“It is in the employer’s interest to have as much information as possible about a potentially disruptive situation, so that it can take steps to avoid such a situation, instead of having to respond to an actual incident.”).
257 See id.
the supervisor can refer the victim-employee to external services. The workplace is an “effective vehicle for disseminating information about available resources.” A qualitative study by Jennifer E. Swanberg and T. K. Logan found that once a supervisor is aware of the situation in the employee’s life, they often go out of their way to offer support. The following quotes from victim-employees in the Swanberg and Logan study provide salient examples of how managers responded to the information by providing slight accommodations and referrals to victim-services:

Yeah, I told my boss. She uhm, that was really like the first incident and she let me take a couple days off from work. She was supportive . . . I could not have [moved to a shelter] without her.

. . .

[T]here was another lady in HR, and I’ve talked to her and told her. The woman that I report to . . . talked with the county attorneys and explained to her what was going on. They have me working at another location now and I don’t mind going to that location. I told her that I really needed to work . . . she mentioned to me that they would set up a laptop [at the shelter] for me and a phone . . . I said that I would be willing to work in the warehouse or work anywhere just so I can work. And they had me, I have to go to [another city] every day and they pay me for the driving time.

. . .

I have a manager . . . [S]he’s taking me to a seminar on domestic violence . . . She helps me . . . because her daughter’s going through it,

258 Hobday, supra note 84, at 21. Consider also that employer-side attorneys often advise employers to collaborate with employees in drafting solutions. See, e.g., Veena A. Iyer, Commentary, Intimate Partner Violence at Work, 27 WESTLAW J. EMP., no. 18, Apr. 2013, at *1, *7.

259 Swanberg & Logan, supra note 21, at 12 (“Among all of the women who confided in supervisors or managers . . . a strong majority (86%) received formal or informal support from the workplace.”).
she can relate. 260

These examples illustrate that when management learns of the employee’s situation, they can involve external services. 261 When an employee seeks these external services, they are more likely to end the cycle of abuse and return to their full level of productivity. Therefore, by opening the lines of communication, the reasonable accommodation approach likely hastens the resolution of the issue. Third, there is a potential benefit to workplace morale when the workplace is more open. Swanberg and Logan’s study also concluded that employees were more focused on their work just knowing that they had their supervisors’ support. 262 Thus, the interactive process that is part and parcel to the reasonable accommodation approach involves a trade of short-term accommodation costs with long-term productivity gains. 263

In sum, while reasonable accommodation statutes do unquestionably require employers to assume the costs of third-party behavior, these statutes can be drafted to ensure that the costs employers bear are no greater than the gains. This is true because the approach necessarily considers the reasonableness of an accommodation and the burden it requires, the statutes can be drafted so as to only require a finite list of low-cost accommodations. Further, there are long-term benefits from an interactive process that allows for communication between the employer and a victim-employee. Thus, accommodations can require only short-term costs and yield long-term productivity

---

260 Id. at 12–13. It is also worth noting that the supervisor’s solution in most of these examples was to provide the victim-employees with a reasonable accommodation. See id. In the study the managers were under no legal requirement to provide victim-employees with accommodation, yet in the second example, the manager intuited that the best solution was to change the victim-employee’s worksite. Id. As such, the reasonable accommodation approach most closely tracks the intuitive behavior of management.

261 See id. at 13.

262 Id. at 12–13.

263 Studies have also shown that in the context of accommodating employees with disabilities, employers report “higher productivity,” “greater dedication,” “fewer insurance claims,” and “improved corporate culture.” Michael Ashley Stein, The Law and Economics of Disability Accommodations, 53 DUKE L.J. 79, 105 (2003).
2. Ad Hoc Judicial Determinations

Another potential criticism of the reasonable accommodation approach is that a court may later hold that what management determines to be an undue burden is in fact a reasonable accommodation. This issue of ad hoc judicial determinations is a valid concern for any statute with a reasonableness standard, especially a statute that seeks to control employer discretion and employment policy. While this concern is certainly valid, several aspects of the reasonable accommodation approach should diminish this concern.

First, some statutes, like Illinois’ VESSA and California’s SB 400, include extensive explications and definitions of both reasonable accommodation and undue burden. The clearly articulated boundaries of reasonable and undue burden constrain judicial discretion in advance.

---


265 See id. (raising the argument in the context of the ADA).


268 See, e.g., Haw. Rev. Stat. §§ 378-81(a)(1)–(6) (2014) (defining reasonable accommodation by way of example, providing the following list of examples: “[c]hanging the contact information . . . of the employee; [s]creening the telephone calls of the employee; [r]estructuring the job functions of the employee; [c]hanging the work location of the employee; [i]nstalling locks and other security devices; and [a]llowing the employee to work flexible hours”); Or. Rev. Stat. Ann. § 659A.275 (2014) (“‘Undue hardship’ means a significant difficulty and expense to a covered employer’s business and includes consideration of the size of the employer’s business and the employer’s critical need for the eligible employee.”).

269 A convincing criticism of the ADA is that “it contain[ed] many ill-defined terms” and thus permitted wide judicial discretion that would increase employer’s litigation costs. McGraw, supra note 264, at 539–40. History has proven this criticism correct. Carrie L. Flores, Note, A Disability Is Not A
terms, employers’ attorneys are able to draft employment policies that are most likely to be determined compliant with the law. Thus, where reasonable accommodation statutes are drafted with clarity, they decrease the likelihood of courts and management reaching different conclusions of the reasonableness of an accommodation.

Second, some courts have already begun to investigate the fine line between reasonable accommodation and undue burden. In 2004, the New York County Supreme Court interpreted the reasonable accommodation requirement of the New York City Human Rights Law (“NYCHRL”) for the first time in Reynolds v. Fraser. There, a former employee of the Department of Corrections brought an action accusing her employer of discrimination against her due to her status as a victim of domestic violence. In this case, after increasing violence at home, plaintiff left her husband and began intermittent bouts of homelessness and refuge in shelters. While at the shelter, plaintiff required surgery and a period of convalescence. Defendant-employer approved the sick leave but later terminated plaintiff due to her violation of a requirement to

Trump Card: The Americans with Disabilities Act Does Not Entitle Disabled Employees to Automatic Reassignment, 43 VAL. U. L. REV. 195, 207 (2008) (observing that the various ambiguous terms of the ADA resulted in numerous Circuit splits). Flores goes on to note that the Supreme Court “has interpreted and qualified some of the statutory text, shedding light on mystifying terms and nuances.” Id. at 209. Employers should learn from the history of the ADA and request that legislatures be clear in their definitions of covered employees, reasonable accommodations, and undue burdens.

Cf. The Americans with Disabilities Act: Great Progress, Greater Potential, 109 HARV. L. REV. 1602, 1615 (1996) (arguing that under the vagueness of the ADA, “[e]mployers [were forced to] choose between making the (perhaps needless) accommodations requested by disabled employees or applicants and risking costly litigation and potential liability for discrimination by refusing such requests”).

Reynolds v. Fraser, 781 N.Y.S.2d 885 (N.Y. Sup. 2004).
Id. at 885, 887.
Id. at 885–87.
Id. The court does not describe why the plaintiff required surgery. See Reynolds, 781 N.Y.S.2d at 885.
be present at her home during sick leave.276 The court in Reynolds held that the Defendant-employer failed to reasonably accommodate the plaintiff as a victim-employee,277 and thus violated NYCHRL section 8-107.1(3)(a)’s requirement that “any [covered employer] . . . shall make reasonable accommodation to enable a person who is a victim of domestic violence . . . to satisfy the essential requisites of a job.”278 The court in Reynolds interpreted the reasonable accommodation provision of the NYCHRL to require that an employer waive a technical requirement (such as a requirement to remain home during sick leave) when the employer is on notice that the employee is a victim.279 The Reynolds case represents the only occasion that a court has considered what a reasonable accommodation for the victim-employee is. That said, its holding appears fairly predictable and thus cuts against employer concerns that accommodation statutes will permit courts to widely second guess the decisions of management.

Finally, reasonable accommodation statutes with clearly defined requirements actually allow employers to predict judicial outcomes and avoid litigation costs. An example of this is discussed in the American Civil Liberties Union (ACLU) of Hawaii’s testimony in support280 of Hawaii’s 2011 Senate Bill 229 (“SB229”).281 SB229 was introduced in 2011 so as to require employers to reasonably accommodate victims of domestic violence.282 In its testimony in support of SB229’s reasonable

276 Id. at 891.
277 Id. at 887.
282 S.B. 229 was eventually passed and became effective January 1, 2012. See HAW. REV. STAT. § 378-81(a) (2014).
accommodation requirement, the ACLU highlighted the success of New York City’s reasonable accommodation statute.\(^{283}\) The ACLU of Hawaii recounts how the national ACLU represented a plaintiff who had been employed by New York City public school systems and needed to take time off of work to “attend court proceedings and seek medical treatment.”\(^{284}\) The plaintiff, “Kathleen,” was reprimanded for excessive absenteeism.\(^{285}\) Later, the employee requested to transfer to another school “for safety reasons,” and “[s]hortly after this conversation, she was fired.”\(^{286}\) The ACLU of Hawaii reported that, in the face of liability under NYCHRL, the school system settled the case and amended its internal policy to cover victims of domestic violence, “acknowledging that reasonable accommodations must be offered to these survivors, and publicizing its new policies throughout the school system.”\(^{287}\) The school system’s decision to settle may indicate that it felt it was clear that it had failed to provide a reasonable accommodation that was required by law. If that is the case, then Kathleen’s case, as presented by the ACLU of Hawaii, may indicate that reasonable accommodation statutes can provide employers with the ability to predict their likelihood of liability.\(^{288}\)

While employers are justifiably concerned that reasonable accommodation laws will permit judges to second-guess management decisions, that concern should not guide them to wholesale reject accommodation laws. First, domestic violence accommodation statutes can be tailored to limit judicial discretion. Second, while the case law on the subject is very limited, the Reynolds case and Kathleen’s case indicate that the jurisprudence of reasonable accommodation of domestic violence victims could be predictable and limit lawsuits.

\(^{283}\) Testimony of the ACLU of Hawaii in Support of S.B. 229, supra note 280, at 1.
\(^{284}\) Id. at 2.
\(^{285}\) Id.
\(^{286}\) Id.
\(^{287}\) Id.
\(^{288}\) See id.
Another concern with the reasonable accommodation approach is that an employer, who provides reasonable accommodations to some employees and not others, may be in violation of other employment discrimination laws. Consider the following hypothetical: an employer institutes a policy that provides accommodations for victims, and all covered employees are women. Would a male employee prevail on a Title VII action sounding in sexual discrimination? In *Muhammad v. Walmart*, the federal district court for the Western District of New York addressed this question and referred to a potential gender discrimination claim as “completely frivolous.” Mr. Muhammad claimed gender discrimination based on the disparate treatment between his “angry outburst” and an earlier violent incident involving a female employee. The female employee had arrived at work in the company of her intimate partner; they had a violent altercation wherein he struck her and shoved her through the entrance of the store; they then “yelled profanities at each other.” Walmart had declined to take disciplinary action

---


290 The overwhelming majority of victims are women. Goldscheid, supra note 30, at 61, 63, 66. It is therefore perfectly reasonable to imagine that for many employers, their only covered employees would be women.

291 Wal-Mart terminated Mr. Muhammad’s employment after Mr. Muhammad angrily confronted his supervisor, threw his identification on the floor, and left work in the middle of his shift, and in front of customers. *Muhammad*, 2012 WL 5950368, at *2. Mr. Muhammad, acting pro se, filed a complaint sounding in racial and disability discrimination. *Id.* at *4. During the pendency of litigation Mr. Muhammad’s attorney purported that there was a claim for gender discrimination. *Id.* at *5. The district court sanctioned the attorney for improperly raising what he termed a “completely frivolous” and unplead cause of action. *Id.* at *6. On appeal, the Second Circuit vacated the sanctions but did not speak to the merits of the claim. *Muhammad*, 2013 WL 5539924.


293 *Id.* at *3.
against the female employee because “it viewed her as being the victim of domestic violence.” Judge Siragusa ultimately decided the gender discrimination claims were improperly pled but, in dicta, opined that “even if they were pleaded, are completely frivolous.” This dictum may allow for the inference that if an employer reasonably accommodates by waiving the applicability of a rule for violence in the workplace, that accommodation is unlikely to create liability under Title VII even when it applies a different disciplinary standard to a male employee during the same shift.

In fact, a reasonable accommodation approach may actually limit employer liability under other federal laws. Consider the following example: a manager who approaches a male employee about his absenteeism learns that the employee’s partner is terrorizing him. Without proper guidance, the manager allows gender bias to enter into her assessment of and response to the situation. She perceives the male employee as weak and terminates him. This fact pattern, which was included in the EEOC Fact Sheet, would expose the employer to liability under Title VII because the manager’s action is based on gender stereotypes. However, a reasonable accommodation law can be drafted so as to require that a victim-employee who requests an accommodation must first document that they are a victim of an incident of domestic violence. As such, in this fact pattern, the manager and the employee would have already had a conversation about the violence, and the manager would be on notice that they are required to reasonably accommodate the employee. She would therefore be less likely to make a snap decision, which could expose the employer to liability. An accommodation statute that requires an interactive process “ensure[s] that any adverse employment action is based on legitimate reasons, as opposed to

294 Id.
295 Id. at *6.
296 This example is a modification of an example given by the EEOC. See EEOC Fact Sheet, supra note 125, at *2.
297 See Goldscheid, supra note 30, at 62.
298 See EEOC Fact Sheet, supra note 125, at *2; see also supra note 129.
299 See supra statutes cited in note 176 and accompanying text.
300 Schwab & Willborn, supra note 236, at 1258–60.
subtle biases.” The reasonable accommodation approach may therefore avoid employer liability under other employment legislation.

In sum, the reasonable accommodation approach is best tailored to employer’s interests. An accommodation statute can include limiting provisions that ensure that employers only incur small accommodation costs. Further, an accommodation statute can ensure that there is a dialogue between the victim-employee and management. Qualitative studies also demonstrate that this dialogue may help the victim-employee gain access to community resources and begin the project of leaving their abuser. Ultimately, this would achieve employers’ greatest interest: to limit the way that domestic violence affects the bottom line by encouraging victim-employees to take affirmative steps to ending the violence in their lives and returning to their full level of productivity.

CONCLUSION

As the employment landscape shifts around them, employers should reconsider their position on employment protections for victims of domestic violence. To date employers have opposed employment protections at the state level, but there are two convincing reasons why this position is untenable. First, domestic violence negatively impacts employers’ business because it cuts into their bottom line; domestic violence decreases productivity, increases turnover, and can even result in violence at the workplace. While some employers have taken steps to address domestic violence internally, legislation that would require all employers to provide employment protections would better serve the business community. Without a law, smaller, resource-poor employers are unable to avail themselves of internal policies.

301 Goldscheid, supra note 30, at 62. See also id. at 114–18.
302 See supra notes 10–12, 84 and accompanying text.
303 See supra notes 30–61 and accompanying text.
304 See supra notes 75–83 and accompanying text.
305 See generally supra Part I.A.
306 See supra notes 92, 97–99 and accompanying text.
Further, employers that do address the problem internally are penalized in the market.\textsuperscript{307} Instead, legislation would serve the business community because it would normalize the cost of accommodating victim-employees in the workplace; thus ameliorating domestic violence’s effect on all employers’ businesses, not merely those that can afford to address the problem internally.\textsuperscript{308} Second, the federal government has signaled that it supports employment protections for victims of domestic violence.\textsuperscript{309} If a federal law is imminent, now is the perfect opportunity for employers to enter into the debate and shape the way that the statute is drafted.\textsuperscript{310}

If employers do reconsider their position and begin to advocate for limited legislation, they should look to state statutes for guidance. As has been seen at the state level, domestic violence statutes, particularly the statutory leave and reasonable accommodation statutes, can be drafted so as to trade short-term costs with long-term benefits. The antidiscrimination approach, on the other hand, is not deserving of support because it is fundamentally reliant on classifications and not aimed at solutions.\textsuperscript{311} Employers have more to gain from a statute that requires affirmative actions that correct the problem.

The statutory leave approach has thus far been the most popular approach\textsuperscript{312} at the state level and is deserving of employer support. There are good reasons for this. First, where a statutory leave law is drafted so as to only provide leave for enumerated reasons and require documentation that the leave is used for that reason, employers can be sure that employees will not misuse the leave.\textsuperscript{313} Second, statutory leave laws can be drafted so as to require reasonable notice of leave; this ensures that employer

\begin{itemize}
  \item \textsuperscript{307} See supra notes 93–96 and accompanying text.
  \item \textsuperscript{308} See supra notes 88–96, 186–97 and accompanying text.
  \item \textsuperscript{309} See supra notes 66–74, 125, 127–39, and accompanying text.
  \item \textsuperscript{310} See supra notes 17–19, 104–08 and accompanying text (discussing how the 2013 session of the New York Senate provides an example of how employer advocacy can mean the difference between a law that requires providing reasonable leave or three months of leave).
  \item \textsuperscript{311} See supra Part IV.B.
  \item \textsuperscript{312} See statutes cited supra note 148.
  \item \textsuperscript{313} See supra notes 160–68 and accompanying text.
\end{itemize}
operations are not disrupted by the leave. Third, the statutory leave approach includes only minor administrative costs. Finally, statutory leave laws have proven amenable to the concern that smaller employers will be disparately impacted; this is accomplished either by exempting the smallest employers or by providing different requirements for employers of different sizes. In sum, employers should consider supporting a statutory leave law, provided that it:

(1) clearly enumerates the reasons for leave so as to permit employers to discipline employees who attempt to abuse the law;

(2) shifts administrative costs onto employees by requiring advance notice of leave or documentation that the leave was taken because of imminent physical injury;

(3) permits only a reasonable amount of time that does not create an undue burden on the employer; and

(4) requires certification of reasonableness by either a medical, legal, or social work professional.

While the reasonable accommodation approach has thus far

314 See statutes cited supra note 176 (discussing notice requirements).
315 See supra notes 177–85 and accompanying text.
316 See supra notes 186–97 and accompanying text.
317 For examples of how such a provision should be drafted, see the statutes cited supra at notes 161–64.
318 Such a provision would permit employers to require advance notice of leave except when the leave is not foreseeable, such as “in cases of imminent danger to the health or safety of the employee.” See, e.g., COLO. REV. STAT. § 24-34-402.7(2)(a) (2013). If the employee takes emergency leave without advance notice, the employer would be permitted to require the employee to provide certification that the leave was taken for an appropriate reason. See CAL. LAB. CODE § 230.1(b)(1)–(2) (West 2014). This documentation could be a police report, a court order, or documentation from a medical professional. See id. § 230(b)(2)(A)–(C). See generally statutes cited supra note 176.
319 Reasonable leave statutes are a hybrid of the statutory leave and reasonable accommodation approaches, and provide the benefits of each. For a discussion of reasonable leave statutes, see supra notes 158, 193–97 and accompanying text. For a discussion of the benefit of clear statutory definitions of reasonableness and undue burden, see supra notes 238–45.
320 Hawaii provides an example of how this requirement would work. See HAW. REV. STAT. § 378-72(b) (2014); see also supra note 197 and accompanying text.
been the least popular approach, \textsuperscript{321} it is, in fact, best suited to achieving employers’ interests. First, the reasonable accommodation approach necessarily considers the costs that employers bear.\textsuperscript{322} Second, reasonable accommodation statutes can be drafted so as to only require low-cost accommodations.\textsuperscript{323} Third, the long-term benefits of the reasonable accommodation approach far outweigh the slight initial costs.\textsuperscript{324} Finally, reasonable accommodation statutes can be drafted so as to require a dialogue—or interactive process—between the victim-employee and management.\textsuperscript{325} This interactive process has a positive impact on employee loyalty and productivity. In sum, employers should consider supporting a reasonable accommodation law provided that it:

(1) limits putative accommodations to a finite list, such as changing an employee’s telephone number or workstation and/or installing safety devices;\textsuperscript{326}
(2) clearly defines factors to be considered in making a determination of undue burden, such as cost of the accommodation and/or size of the employer;\textsuperscript{327}
(3) clearly articulates that any accommodation that would require the employer to jeopardize the safety of their workplace is unduly burdensome;\textsuperscript{328}
(4) requires that a requesting employee document their status as a victim of domestic violence;\textsuperscript{329} and
(5) requires an interactive process between the employer and the victim-employee as to what a reasonable

\textsuperscript{321} See statutes cited \textit{supra} notes 223–28.
\textsuperscript{322} See \textit{supra} notes 238–45 and accompanying text.
\textsuperscript{323} See \textit{supra} notes 246–49 and accompanying text.
\textsuperscript{324} See generally \textit{supra} Part V.B.
\textsuperscript{325} See \textit{supra} notes 248–63 and accompanying text.
\textsuperscript{326} See, e.g., \textit{CAL. LAB. CODE} § 230(f)(2) (West 2014). As discussed above, employers should advocate for a statute that requires only a finite list of accommodations. See \textit{supra} note 249 and accompanying text.
\textsuperscript{327} See, e.g., \textit{ILL. COMP. STAT.} 180/30(b)(4)(B)(i)–(iv) (2012); see also \textit{supra} notes 233, 238–41 and accompanying text.
\textsuperscript{328} See, e.g., \textit{CAL. LAB. CODE} § 230(f)(6) (West 2014); see also \textit{supra} notes 242–45 and accompanying text.
\textsuperscript{329} See, e.g., \textit{HAW. REV. STAT.} § 378-81(b) (2014).
accommodation for that particular employee would be.  

Domestic violence is a workplace problem. In 2009, “twenty-nine percent of male workers and forty percent of female workers reported having been victims of domestic violence . . . .” Domestic violence disrupts employers’ businesses, affects their bottom line, and costs between $3 and $13 billion per year in lost profits. The time has come for employers and their advocates to support legislation that addresses this workplace problem. The time has come for employers to support legislation that provides employment protections to victim-employees without imposing too great a cost on business.

330 See, e.g., CAL. LAB. CODE § 230(f)(5) (West 2014); see also supra note 255 and accompanying text.
331 For other scholars reaching the same conclusion from the perspective of the employee and society see generally Widiss, supra note 6; Karin, supra note 6; and Randel & Wells, supra note 48.
332 Jacobs & Raghu, supra note 20, at 597.
333 See supra notes 43–44 and accompanying text.
MOVING PAST A “POCKET CHANGE” SETTLEMENT: THE THREAT OF PREEMPTION AND HOW THE LOSS OF CHANCE DOCTRINE CAN HELP NFL CONCUSSION PLAINTIFFS PROVE CAUSATION

John Guccione*

I. INTRODUCTION

On August 29, 2013, retired Judge Layn R. Phillips1 announced a “historic” $765 million settlement proposal between the National Football League (“NFL” or the “League”) and over 4,500 retired football players.2 The plaintiffs, former NFL

* J.D. Candidate, Brooklyn Law School, 2015; B.A., SUNY College at Geneseo, 2010. I would like to thank my wonderful parents, Tom and Karen Guccione, as well as my friends and family for their amazing support. I also wish to thank the members of the Journal of Law and Policy for their excellent suggestions and generous sacrifices in time and energy. Special thanks to Tiffany Colón, whose love and incredible encouragement, patience, and kindness has made this note, and all else, possible for me.


2 The retired players who sued the league will be referred to as “concussion plaintiffs,” “plaintiffs,” and “former players/athletes.” The agreement is not final, as it is still pending the preliminary approval of Judge Brody. The settlement specifically allocates $75 million for baseline medical exams, $675 million for cognitive injury compensation, and $10 million for research and education, along with monies for the costs of notice to the class, settlement administrator compensation, and legal fees. Press Release, Alternative Dispute Resolution Ctr., NFL, Retired Players Resolve Concussion Litigation (Aug. 29, 2013) [hereinafter ADR Press Release].
athletes, accused the League of being aware of, and actively concealing, evidence linking football to mild traumatic brain injuries and their resulting “pathological and debilitating” neurological effects. The plaintiffs alleged “intentional tortious misconduct” by the NFL, “including fraud, intentional misrepresentation, and negligence,” and sought “a declaration of liability, injunctive relief, medical monitoring, and financial compensation for the long-term chronic injuries” the plaintiffs sustained during their NFL careers.

Reactions to the proposed agreement varied greatly. For a number of those closely involved in the litigation and settlement process, there was an initial attitude of satisfaction on both sides. For example, NFL Commissioner Roger Goodell remarked, “this [settlement is] best for the game going forward,” and “best for the players, and that’s what’s important.” NFL Executive Vice President Jeffrey Pash reiterated the League’s apparent commitment to the well-being of athletes and their families: “This agreement lets us help those who need it most and continue our work to make the game safer for current and future players.” Judge Phillips, who oversaw the parties’ negotiations, stated the proposed settlement would ensure retired NFL athletes received necessary financial support, at a time when they most needed it.

On the players’ side, Kevin Turner, a former running back and a lead plaintiff in the litigation, assured the public that the benefits of the agreement would make a difference for thousands

\[4\] Id.
\[5\] See, e.g., What People Are Saying, NFL Concussion Litig., http://www.nflconcussionmdl.org/what-people-are-saying/ (last visited Mar. 28, 2014) (listing testimonials of former NFL players, reporters, and legal experts expressing relief and gratitude as a result of the NFL settlement).

\[7\] ADR Press Release, supra note 2.
\[8\] Id.
of former athletes, both now and in the future. Christopher Seeger, co-lead plaintiffs’ attorney, reiterated this message, stating that the agreement “will get help quickly to the men who suffered neurological injuries . . . faster and at far less cost, both financially and emotionally, than could have ever been accomplished by continuing to litigate.”

Despite such positive responses, other commentators and former players expressed immediate dissatisfaction. A number of experts and former NFL players spoke out against the settlement, highlighting a number of terms that clearly favored the League. In addition, simply by settling (regardless of the final terms) the NFL was afforded a number of protections they would have lost had the litigation continued. For example, as with most settlements, the proposed terms expressly articulated that the agreement in no way represented an admission of liability on the part of the NFL. Many commentators also noted that by agreeing to settle, the NFL avoided an extremely damaging discovery process. Should the case have moved forward, plaintiffs’ counsel likely would have deposed the Leagues’ staff and obtained access to internal documents and e-mails through the discovery process, revealing exactly what the NFL knew and

---

9 Id.
10 Id.
11 Former running back Leroy Hoard, for example, expressed concern regarding the fact that the second half of settlement monies are distributed over a long 17-year period, while punter Chris Kluwe and former linebacker Aaron Curry worried that the settlement, while helpful, would not provide sufficient compensation. Reaction to the Concussion Deal, ESPN (Aug. 30, 2013), http://espn.go.com/nfl/story/_/id/9612672/reaction-nfl-concussion-settlement.
12 ADR Press Release, supra note 2.
13 See, e.g., LaMar C. Campbell, Opinion, NFL Concussion Settlement Raises Questions, CNN (Sept. 9, 2013), http://www.cnn.com/2013/09/08/opinion/campbell-nfl-lawsuit/ (“[T]he league does not have to face the discovery and deposition process and therefore leaves many questions unanswered.”); Daniel Engber, Opinion, NFL Concussion Settlement Doesn’t Show Us How Dangerous Football Really Is, THE BUFFALO NEWS (Sept. 8, 2013), http://www.buffalonews.com/opinion/nfl-concussion-settlement-doesnt-show-us-how-dangerous-football-really-is-20130908 (“[T]hey would have been forced to put a huge library of internal documents on the record.”).
allegedly concealed from players and the general public.\textsuperscript{14} Owners would have faced “continuing accusations of abusing players,” as highly skilled and motivated plaintiffs’ counsel would have conducted “nothing less than a strip search of NFL records.”\textsuperscript{15} Escaping an admission of liability also meant the NFL avoided one of the Plaintiffs’ key allegations: that the League knew the dangers and risks of repeated concussions, that it voluntarily undertook the responsibilities of studying NFL head injuries, and ultimately concealed their long term effects.\textsuperscript{16} In the words of former NFL Players Union President and Pro-Bowler, Kevin Mawae, while the settlement was great for older players in need of immediate help, it constituted “$700 million worth of hush money that [the NFL] will never be accountable for.”\textsuperscript{17}

Issues with the proposed settlement extend beyond the League’s ability to avoid admitting liability and evade discovery process disclosure. A number of critics have also expressed doubt with regard to the adequacy of the underlying settlement amount, going so far as to call it “barely a drop in the bucket.”\textsuperscript{18} Indeed, for an organization that currently generates approximately $9 billion a year in revenue, the $765 million settlement amount reflects “less than half of what ESPN alone pays the League annually.”\textsuperscript{19} Some early commentators predicted that the suit

\textsuperscript{14} See Campbell, supra note 13.
\textsuperscript{15} Lester Munson, Mediation Could Be the Answer, ESPN (July 9, 2013), http://espn.go.com/espn/otl/story/_/id/9462264/questions-answers-judge-decision-send-nfl-concussion-lawsuit-mediation.
\textsuperscript{16} See Plaintiffs’ Master Complaint, supra note 3, at 23, 32.
\textsuperscript{19} Patrick Hruby, Q&A: The NFL’s Concussion Deal, THE ROTATION (Aug. 30, 2013), http://therotation.sportsonearthblog.com/qa-the-nfls-
could be worth as much as $10 billion, assuming each injured player and their family received an award of $500,000. In January 2013, Paul M. Barrett of Businessweek hypothesized a $5 billion agreement. Even the more grounded figures initially sought by the plaintiffs were in excess of $2 billion; over 260% more than the proposed settlement amount. As former Minnesota Vikings player and current plaintiff Brent Boyd lamented, “$765 Million? The breakdown is $1.2 million over 20 years per team. What is that, a third of the average salary? There is no penalty there. It’s pocket change.”

Presiding U.S. District Judge Anita Brody ultimately validated these concerns on January 14, 2014 by refusing to grant the settlement preliminary approval. Before the proposed class action settlement agreement could take effect, Judge Brody had to give her approval pursuant to Federal Rule of Civil Procedure 23(e). In class actions such as this, the court “must assure ‘to the greatest extent possible that the actions are prosecuted on behalf of the actual class members in a way that makes it fair to


20 See Glenn M. Wong, SN Concussion Report: NFL Could Lose Billions in Player Lawsuits, SPORTING NEWS (Aug. 22, 2012), http://www.sportingnews.com/nfl/story/2012-08-22/nfl-concussion-lawsuits-money-bankrupt-players-sue-head-injuries. This prediction was made when only 3,000 former players were involved. Applying this $500,000 per player award to the number of plaintiffs ultimately involved in the settlement would result in even greater damages. See id.


23 See Wilner, supra note 17.


25 FED. R. CIV. P. 23(e); In re NFL, 961 F. Supp. 2d at 713–14.
Because class action settlements can bind absent class members who did not participate in the litigation, Rule 23(e) requires judges to ensure the agreement is “fair, reasonable, and adequate” irrespective of the parties’ approval. While the Judge noted that Plaintiffs’ counsel “believed” the aggregate sum of the settlement was sufficient based on “analysis conducted by the independent economists or actuaries retained by the parties,” she had concerns about the settlement’s “fairness, reasonableness, and adequacy,” as such analyses were not actually provided to the Court. Brody refused to grant preliminary approval until documentary proof of the settlement’s fairness was provided.

Judge Brody also expressed concern that the funds would be insufficient to compensate all class members who received a “Qualifying Diagnosis” (such as dementia, Alzheimer’s, or Parkinson’s Disease), resulting in the settlement’s largest payouts. Brody noted that the settlement “contemplates a 65-year lifespan,” and was expected to cover a class of around 20,000 individuals. She found it “difficult to see how the Monetary Award Fund would have the funds available over its lifespan to pay all claimants at these significant award levels.” As of April 2014, the settlement remains on hold.

Given the apparent inadequacies of the proposed settlement, and assuming that such amounts may not be increased in the future, litigation may be the only way that former-NFL players can be assured of an adequate remedy. However, there are serious problems with the plaintiffs’ claims. First, plaintiffs

---

26 In re NFL, 961 F. Supp. 2d at 713 (citation omitted).
28 In re NFL, 961 F. Supp. 2d at 716.
29 Id.
30 Id. at 715.
31 Id.
33 One such issue this Note will not discuss is class certification, which has not occurred due to acceptance of the proposed settlement. Though NFL athletes could still bring individual suits for their injuries, certification will
must be wary of the possibility that their claims will be preempted by federal law, forcing them to be resolved through arbitration pursuant to the NFL-NFL Players’ Association collective bargaining agreement (“CBA”), rather than a court proceeding. If the plaintiffs are forced to arbitrate, they lose the benefit of a trial where potentially sympathetic jurors might favor the many badly injured and allegedly misled plaintiffs, instead facing a more neutral decision-maker less likely to award significant damages.

Second, players must be able to prove causation—that is, that the NFL caused their injuries. This is no easy task when some athletes may already have had a predisposition to diseases like Alzheimer’s, and may have sustained brain injuries in non-NFL football activities, such as high school and college football, and in their personal lives. To help resolve this issue, the courts and the plaintiffs should look to extend the “loss of chance doctrine,” traditionally applied only in medical malpractice lawsuits,34 to the NFL. The doctrine allows injured parties to recover damages for the “reduction in odds of recovery” caused by a defendant’s negative contributions, even if plaintiffs cannot show that the alleged injuries were “caused in fact by the defendant’s negligence.”35 As long as an injured party can demonstrate that a defendant’s actions lessened their ability to recover, the defendant may be held liable for that reduction.36

Through the loss of chance doctrine, plaintiffs can argue that pose a major barrier for players, especially their medical monitoring claims. Since “liability turns on the specific facts of each class member’s claimed exposure,” and class members may not share identical risks of harm, some argue such claims are not “indivisible,” and that class certification would be denied. Sheila B. Scheuerman, The NFL Concussion Litigation: A Critical Assessment of Class Certification, 8 FIU L. REV. 81, 105 (2012); see also TIMOTHY LIAM EPSMEN, SMITHAMUNDSN LLC, NFL CONCUSSION CLASS ACTION LITIGATION, available at http://www.dri.org/DRI/course-materials/2012-AM/pdfs/39b_Epstein.pdf.


36 Id.
despite possible neurological disease predispositions and brain injuries arising outside of the NFL, the League conflated these risks and should therefore be held liable for the plaintiffs’ resulting reduced changes of recovery. Given the unfavorable terms of the proposed settlement\(^{37}\) and the risk it will ultimately be rejected, an extension of loss of chance to nonmedical malpractice torts (though it must be limited, and has its risks\(^{38}\)) provides a great opportunity for plaintiffs to succeed on their merits and hold the NFL accountable.

Fortunately, even if Judge Brody ultimately approves the settlement of \textit{In re National Football League Players’ Concussion Injury Litigation},\(^{39}\) disgruntled plaintiffs will have an opportunity to opt out.\(^{40}\) Although doing so would significantly delay resolution of the opting-out plaintiff’s claims, those that can afford to do so should strongly consider it, as they could continue litigating along with former football players not currently involved in the lawsuit but interested in pursuing individual claims. While threat of federal law compelling arbitration pursuant to the CBA is possible, preemption should not apply here. Additionally, while proving causation will be difficult, there are methods available for the plaintiffs to do so, including through a potential extension of the loss of chance doctrine.

Part II will discuss the adequacy of the proposed concussion


\(^{38}\) See generally David A. Fischer, \textit{Tort Recovery for Loss of a Chance}, 36 WAKE FOREST L. REV. 605 (2001). One concern is that the doctrine, once accepted widely, becomes difficult to limit and may “swallow” the traditional more-likely-than-not rule. \textit{Id.} at 606–07. If this happened, there are concerns that all negligent actors could wrongly become liable for injuries they did not \textit{cause} but somewhat \textit{contributed to}, an extremely uncertain determination in many contexts.


litigation settlement, specifically whether it provides sufficient sums to compensate former players and provide for their medical care, and how the uncertainties of litigation and the necessity for immediate relief incentivized the plaintiffs to accept an unfavorable settlement. Part III will discuss the threat of preemption, and why it should not be applied to this case. Part IV will discuss the issues inherent in proving causation and offer a potential solution through judicial extension of the loss of chance doctrine. Should NFL concussion plaintiffs pursue litigation, avoid preemption, and prove causation, they will be able to hold the NFL accountable for its actions, and may better assure themselves and their families of fair compensation.

II. ISSUES OF TIMING AND CERTAINTY INDUCED PLAYERS TO SETTLE

A. Adequacy and Timing

The proposed settlement agreement has various components. First, the NFL will provide $675 million over an extended period to compensate former players for their injuries, with various payments depending on the player’s individual diagnosis. For example, the settlement awards a maximum of $3 million for “moderate dementia,” $3.5 million for Alzheimer’s or Parkinson’s Disease, $4 million for death with chronic traumatic encephalopathy (CTE), a degenerative brain disease associated with multiple concussions, and $5 million for amyotrophic lateral sclerosis (ALS), or Lou Gehrig’s Disease. The NFL will also provide an additional $75 million for medical testing, $10 million for educational purposes, and $4 million for class notice costs. It also provides for over $110 million in attorney’s fees and an


42 These maximum awards are reduced if the former player played less than five “Eligible Seasons,” and/or if the player was diagnosed after the age of forty-five. Id. at 11–13.

43 ADR Press Release, supra note 2.

44 Sofia Pearson & Jef Feeley, NFL’s $914 Million Concussion Deal
additional $37.5 million contribution if the Settlement Administrator determines the Injury Fund is inadequate.\textsuperscript{45} Settlement funds are expected to last for sixty-five years.\textsuperscript{46}

Despite a proposed settlement that appears to include a large amount of funds, there is good reason to support the doubts of Judge Brody and a large number of journalists, experts, and members of the class action. Though Christopher Seeger (lead co-counsel for the plaintiffs) made public assurances that forthcoming reports from experts, economists, and actuaries would confirm that the proposed settlement will be “sufficiently funded,” some basic mathematics have brought that claim into serious question.\textsuperscript{47} Judge Brody expressed concerns that the settlement provides insufficient compensation if “even . . . only 10 percent” of retired players qualify for one of the tiers outlined above.\textsuperscript{48} Indeed, enrollment numbers in prior NFL player injury compensation programs have indicated that the number of players with serious brain injuries may be high enough to quickly empty the fund.\textsuperscript{49}

One such program is the NFL’s “88 Plan,” implemented in 2007.\textsuperscript{50} The 88 Plan is a program designed to provide nearly $100,000 in yearly aid for the medical and custodial expenses of qualified former players suffering specifically from dementia,


\textsuperscript{46} Class Action Settlement Agreement, \textit{supra} note 40, at 32.

\textsuperscript{47} Fainaru & Fainaru-Wada, \textit{supra} note 37.


\textsuperscript{50} Alan Schwartz, \textit{Before Dementia Assistance, Help With N.F.L. Application}, N.Y. TIMES (Jan. 21, 2010), http://www.nytimes.com/2010/01/22/sports/football/22eckwood.html?_r=0.
including “dementia due to head trauma.” The League designed the 88 Plan partly in response to increasing media attention and player complaints regarding the effects of concussions. Since 2007, 223 former NFL players have qualified for the program, and the League has approved over $23 million in assistance. It is likely many of these individuals would also qualify for the proposed settlement’s larger payment tiers, which includes awards of $3 million for dementia and $5 million for Alzheimer’s, since the 88 Plan was specifically designed to aid players diagnosed with dementia.

Patrick Hruby of Sports on Earth, an online sports blog, used numbers from 88 Plan enrollment to argue against the adequacy of the proposed settlement. He accounted for the 233 athletes that qualified for the 88 Plan, and added to that number, thirty-four former players who have already been diagnosed with CTE (a disease not covered by the 88 Plan, but covered under the settlement). He took the sum, 267, and multiplied it by an average award of $2 million. The amount, $534 million, accounts for a vast majority of the available funds, without adding any newly diagnosed injuries whatsoever. In addition, less than one third of retired NFL players were involved in the concussion litigation at issue, but under the proposed settlement, all retired NFL players would be eligible for this fund.

53 Id.
55 Hruby, supra note 49.
56 Id.
57 Id.
58 Id. According to Hruby, there are between 15,000 and 18,000 living
possible that concussion-related, long-term injuries will only increase as time goes on. Younger players have generally played more football than their predecessors (from youth leagues to high school and collegiate football), during a period where athletes have generated greater impacts and commonly used painkillers like Toradol, which may have exacerbated concussion harms.

In other words, not only may there already be enough retired NFL athletes to empty the settlement funds, but the number of retired players with qualifying diagnoses will likely increase with time.

The proposed settlement also has serious issues outside the amount of overall compensation. Many seriously impaired plaintiffs may not qualify for seven-figure awards, yet will need or are already receiving nursing home care, where residence costs can average $80,000 per year—and often much more. The proposed settlement also disqualifies awards for the families of former players diagnosed with “football-related brain damage” who died prior to 2006, precluding a number of wrongful death suits.

In addition, while the proposed settlement will take care of retired NFL players. The concussion litigation here had around 4,600 plaintiffs. Id.  


60 Toradol, a painkiller with blood-thinning effects, was the subject of a 2011 lawsuit where players alleged that the drug’s ability to dull pain made it more difficult for players to recognize concussion symptoms. See Ken Belson, Ex-NFL Players Suing Over Use of Painkiller, N.Y. TIMES (Dec. 5, 2011), http://www.nytimes.com/2011/12/06/sports/football/nfl-sued-by-ex-players-over-painkiller-toradol.html.

61 See Hruby, supra note 49.

62 See Caplan & Igel, supra note 18.

63 Class Action Settlement Agreement, supra note 40, at 29. The NFL hoped to bar all wrongful death claims from the settlement whose two-year statute of limitations (typical for most states) had expired. While negotiations extended the provisions to players dying after 2006, the families of those dying prior to that year were not included. See Fainaru & Fainaru-Wada, supra note 37.
some of the plaintiffs’ legal fees, many former athletes may still have to pay significant portions of any awards to their attorneys. Instead of fees being paid out of the settlement, dozens of plaintiffs’ attorneys would collect fees directly from their clients pursuant to previously negotiated agreements. This means not only will some attorneys get as much as one-third of their clients’ settlement monies directly from the players, but they may be paid twice, receiving a share of the League’s settlement fund as well.

Nonetheless, Commissioner Goodell defended the proposed settlement against such concerns about its inadequacy, attempting to dispel the notion that the NFL could have afforded a higher settlement. Goodell noted that despite the NFL grossing approximately $10 billion per year, because “there’s a difference between making (money) and revenue,” the settlement was best for the plaintiffs and a “tremendous amount of money.” This argument was lampooned by Deadspin writer Reuben Fischer-Baum, who noted the League will generate approximately $180 billion in profits by the time the entire settlement is paid out to the plaintiffs. The proposed settlement would account for only 0.425% of this projection.

Although the parties’ agreement raises serious questions, some individuals have considered it a necessary evil. Many commentators, former and current NFL athletes, and legal experts examined the settlement from the players’ perspective,

---

64 See ADR Press Release, supra note 2.
65 See Fainaru & Fainaru-Wada, supra note 37.
66 See id.
67 Begley, supra note 6 (alteration in original). Some, like Goodell, were quick to evaluate the effectiveness of the settlement solely based on the dollar amount, without reference to the staggering costs and debilitating injuries sustained by former players. For example, on the day of the settlement announcement, Sports Illustrated writer Peter King tweeted sarcastically: “I love everyone calling $765m chump change.” Peter King, TWITTER (Aug. 29, 2013, 1:30PM), https://twitter.com/SI_PeterKing/status/373135592684396544.
68 Reuben Fischer-Baum, Infographic: The NFL’s Puny Concussion Settlement, Visualized, DEADSPIN (August 29, 2013, 4:14 PM), http://deadspin.com/infographic-the-nfls-puny-concussion-settlement-visu-1222822576. This is likely a conservative estimate, as it assumes the NFL maintains, and will not exceed, its current profit levels.
and noted that while settling could cause the plaintiffs to lose billions of dollars and an admission of liability, an agreement assured the plaintiffs of both timeliness and certainty. As Brett Romberg, an initial plaintiff in 2010, stated, although the NFL “messed up in the past,” the $765 million “will be a much-needed Band-Aid, especially for those who suffered injuries 20 and 30 years ago.”

Timing was perhaps the paramount issue for the former players with the most developed injuries and diseases. Kevin Turner, a 44-year-old former running back suffering from ALS or Lou Gehrig’s disease, stated that “[f]or those who are hurting, this will bring comfort today . . . . The compensation in this settlement will lift a huge burden off the men who are suffering right now.” Indeed, the plaintiffs include former players as old as eighty-four years old, many of whom played less than three years in the League, some of whom never made it on an NFL roster, and many of whom have been rendered

---


72 ALS is a “progressive neurodegenerative disease that affects nerve cells in the brain and the spinal cord,” eventually leading to paralysis and death. What is ALS?, ALS Ass’n, http://www.alsa.org/about-als/what-is-als.html. One study showed that the risk of ALS and Alzheimer’s disease among football players is between three and four times greater than that of the general population. See Everett J. Lehman et al., Neurodegenerative Causes of Death Among Retired National Football League Players, 79 NEUROLOGY 1 (2012).

incapable of holding a job. These factors have created a large class of individuals who have serious long-term injuries but little money, placing a huge burden on these players and their families. Mary Lee Kocourek, widow of Dave Kocourek—a nine-year professional and four-time AFL All-Star—described the hardships the couple faced less than a year before Dave passed away. Doctors diagnosed Dave with dementia before his sixty-fifth birthday, and his condition deteriorated to the point that Mary Lee had no choice but to place him in a nursing home. Although she received some financial help from the NFL, the cost of nursing home care was close to $80,000 annually, while Dave’s yearly salary as a professional never exceeded $35,000.

By agreeing to settle with the NFL, the former players and their families in the most need would receive immediate help, rather than waiting until litigation is resolved, possibly years down the road. Rejecting settlement offers and proceeding with the lawsuit could have easily delayed monetary aid to the plaintiffs for at least another two years, given the complexities of the suit, and possibly resulted in even less compensation. These are people who need funds now. If Judge Brody eventually grants preliminary approval, the decision will likely be appealed, which will prevent class members from opting out of the settlement and fully pursuing their own claims until all appeals are fully exhausted. Given these harsh realities, and the fact that the terms of the current settlement require the NFL to pay

---


76 Id.

77 Id.

78 Id.

79 Rishe, supra note 69.

approximately fifty percent of the settlement amount over the next three years, it is not surprising that many concussion litigation plaintiffs support the proposed settlement.\textsuperscript{81} Though there is substantial evidence that the current agreement is not the best agreement that the plaintiffs could have achieved, it nonetheless provides some immediate help to those suffering the most.

\textbf{B. The Problem of Certainty}

Certainty of the outcome of litigation was another major issue for the players. If the plaintiffs do not receive any assistance from the NFL, many will be unable to continue paying for their medical care.\textsuperscript{82} The figures of the proposed settlement, despite its inadequacies, at least guaranteed the plaintiffs some assistance with medical bills. Paul D. Anderson, attorney and concussion litigation expert, asserted that despite the settlement’s shortcomings, “when balanced against the lives of many players and families that are on the verge of bankruptcy and death, the urgency is clear. Guaranteed money now is much better than no money after years of litigation.”\textsuperscript{83} The settlement was partly induced by fears that should the former players fail to settle, their lawsuit could end in dismissal or a judgment for the NFL.\textsuperscript{84} Paramount among these fears was the issue of preemption by the NFL-NFL Players’ Association (“NFLPA”) CBA, the challenges of obtaining class certification, and the difficulty associated with proving tort causation.

While avoiding preemption and proving causation will be difficult, the apparent inadequacies of the proposed settlement may make going to trial necessary, as litigation may be the only route to ensure fair compensation.\textsuperscript{85} Subsequent examination of

\textsuperscript{81} The balance of the settlement would be paid over the subsequent seventeen years. See ADR Press Release, supra note 2.

\textsuperscript{82} See Jenkins & Maese, supra note 52.

\textsuperscript{83} Paul D. Anderson Consulting, LLC, Report: Judge Brody Threatened to Dismiss the Heart of the Players’ Case, NFL CONCUSSION LITIG. (Sept. 3, 2013), http://nflconcussionlitigation.com/?p=1508.

\textsuperscript{84} Id.

\textsuperscript{85} Attorney Paul Anderson expressed his extreme dissatisfaction with the settlement—and no longer able to refrain from taking an active role in
the proposed terms indicate that while the settlement could lessen the burden on those injured plaintiffs in the most need, many others would not receive the security they envisioned and deserve. In addition, further pursuing a lawsuit would allow for discovery, disclose the NFL’s private information, and could force the League to admit liability. Though the road is uncertain, preemption should not affect the plaintiff’s claims, and increasing medical evidence—along with a possible extension of the loss of chance doctrine—could allow plaintiffs to succeed at trial.

III. THE THREAT OF PREEMPTION

A. Section 301

If the plaintiffs did not agree to settle, they faced the possibility that the Labor Management Relations Act of 1947 (“LMRA” or the “Taft-Harley Act”) would preempt their claims against the NFL. In its memorandum in support of their motion to dismiss dated August 30, 2012, the NFL focused on preemption and section 301 of the LMRA.86 This section has been interpreted to preempt all state law claims “the resolution of which is substantially dependent upon or inextricably intertwined with the interpretation of the terms of a collective bargaining agreement, or that arise under the collective bargaining agreement.”87 The NFL argued that the plaintiffs’ tort claims required the


interpretation of several terms in the CBA,\(^{88}\) and therefore, any adjudication must take place pursuant to the CBA’s agreed-upon grievance procedures. This would require arbitration, and thus dismissal from federal court.\(^{89}\)

Under section 301 of the LMRA, federal law governs any lawsuit concerning a violation of a contract between an employer and a labor organization (here, the NFLPA).\(^{90}\) Because it would be an excessive burden to require bargaining parties to reach an agreement that complies with the laws of all fifty states, section 301 seeks to ensure “uniform interpretation” of bargaining agreements through the use of federal law.\(^{91}\) As Justice William Douglas made clear in *Textile Workers Union of America v. Lincoln Mills of Alabama*, the purpose of section 301 was not only to give federal courts jurisdiction over labor disputes, but to evidence “a federal policy that federal courts should enforce [collective bargaining] agreements . . . and that industrial peace can best be obtained only in that way.”\(^{92}\) Should bargaining parties agree to a dispute resolution provision in their CBA, Congress intended it to be enforced: “Final adjustment by a method agreed upon by the parties is declared to be the desirable method for settlement of grievance disputes arising over the application or interpretation of an existing collective-bargaining agreement.”\(^{93}\)

Therefore, if resolution of a state law claim is “substantially dependent upon analysis of the terms” of a labor contract between the parties, it is preempted by federal law and may be dismissed.

\(^{88}\) The NFL specifically referred to a number of CBA provisions it felt required interpretation, including medical care provisions “relating to assessment, diagnosis, and treatment of player injuries,” player rights and obligations provisions including the ability to choose surgeons and obtain second opinions, rule-making and player safety provisions in order to help make the sport safer, and provisions discussing player benefits and grievance procedures. Defendants’ Motion to Dismiss, *supra* note 86, at 12–15.

\(^{89}\) *Id.* at 14.

\(^{90}\) 29 U.S.C. §185(a).


\(^{92}\) 353 U.S. 448, 455 (1957).

pursuant to the parties’ collective bargaining agreement. Hence, if plaintiffs’ dispute is “dependent on” or “intertwined with” the NFL-NFLPA CBA’s provisions, it will be adjudicated pursuant to the CBA, which compels arbitration. Resolution through arbitration gives the NFL a distinct advantage: while plaintiffs in employment disputes succeed in thirty-six percent of federal court cases, only twenty-five percent of such plaintiffs succeed through arbitration, with the average award being less than eighteen percent of what prevailing receive on average from federal courts. Arbitration also requires adjudication pursuant to contract law, rather than tort law, making punitive damages (to disincentivize the NFL from engaging in such conduct in the future) unavailable. Finally, unlike a public trial, arbitration pursuant to the NFL-NFLPA CBA must be confidential, preventing the public from learning the specifics of the proceeding.

Before agreeing to settle, the plaintiffs were justifiably concerned that their claims would be preempted. A news report released on September 1, 2013, prior to the settlement agreement, claimed presiding Judge Anita Brody “signaled” that she would accept some part of the NFL’s preemption argument, and that the “bulk” of the players’ case would be dismissed. In addition, the NFL and its teams have often successfully argued for LMRA preemption in the past. For example, in Givens v. Tennessee
Football Inc., former player David L. Givens sued his former team, the Tennessee Titans, alleging bad faith in performing contractual obligations, negligence, and outrageous conduct for withholding important medical information regarding Given’s knee.101 Ultimately, the Tennessee Titans successfully argued for preemption, since Article XLIV of the CBA required team physicians to advise a player of any conditions that could affect their health or performance.102

In addition, in Stringer v. NFL, the court found the plaintiff’s wrongful death claim was preempted after her husband, Pro Bowl lineman Korey Stringer, died of heat stroke during training camp.103 Although the court held that the plaintiff’s claim did not arise out of the CBA (which made no mention of preventing or treating heat-related illnesses),104 the court did find that resolving her claim was “substantially dependent” on the interpretation of CBA Article XLIV’s team trainer and physician regulations.105

The plaintiffs in the current NFL lawsuit, to avoid preemption, argued that the NFL owed them a duty of care completely independent from the CBA.106 First, plaintiffs argued that the NFL assumed the duty to act as a guardian of player safety since the NFL’s inception in the 1920s, decades before the first CBA.107 Additionally, the plaintiffs asserted that because the CBA provisions cited by the League make no mention of the NFL itself, “the duties they impose on teams are legally irrelevant to


102 Id. at 990.
104 Id.
105 Id. at 906, 911, 915.
106 Surreply of Plaintiffs in Response to Defendants National Football League’s and NFL Properties LLC’s Reply Memorandum of Law in Further Support of Motion to Dismiss the Amended Master Administrative Long-Form Complaint at 1, In re NFL Players’ Concussion Injury Litig., 961 F. Supp. 2d 708 (E.D. Pa. 2014) [hereinafter Plaintiffs’ Surreply].
107 Id.
the NFL’s separate duty to safeguard players from neurological injuries.” Second, the plaintiffs argued that the NFL assumed a duty of care based on its “unrivaled access to neurological-injury data,” and its voluntary creation of a committee to “opine on the risks of brain injuries in football.”

B. The Failures of the Mild Traumatic Brain Injury Committee

The “committee” the plaintiffs referred to was the Mild Traumatic Brain Injury Committee (“MTBIC”). The League formed the MTBIC in 1994 to study the effects of concussions and brain injury in football. Dr. Elliot Pellman, a former New York Jets team doctor and rheumatologist, was appointed chair of the panel despite little experience in neurology (Pellman was not a neurologist) or the type of brain injuries at issue. He remained chairman until he resigned in 2007, in large part because of increasing controversy and negative press about his tenure as chairman, including his troubling lack of expertise and support of incorrect and misleading research that he conducted and disseminated during his tenure.

The details of Dr. Pellman’s incompetence and deception border on the absurd. The New York Times reported that Pellman had “exaggerated several aspects of his medical education and professional status.” For example, Dr. Pellman maintained he received his medical degree from SUNY Stony Brook, when in reality he attended a school in Guadalajara, Mexico. Further, Pellman claimed he was an associate clinical

---

108 Id.
109 Id.
110 Barrett, supra note 21.
112 Id.
114 Id.
115 Id.
professor, but was actually a non-teaching assistant.\footnote{Id.} He also purported to be a fellow of the American College of Physicians, though he had not held the title for over six years.\footnote{Id.} In addition to questionable credentials, Dr. Pellman displayed questionable judgment. It was, in the eyes of many experts and critics, very troubling that the individual entrusted with the serious task of studying mild traumatic brain injuries in order to ensure player safety was attributed the following quote: “Concussions are part of the profession, an occupational risk. [A football player is] like a steelworker who goes up 100 stories, or a soldier.”\footnote{Id.} Dr. Pellman garnered little respect amongst his colleagues: “When neuropsychologists sit around telling jokes, we call him ‘Mr. Pellman.’”\footnote{Peter Keating, \textit{Doctor Yes}, ESPN (Nov. 6, 2006), \url{http://espn.go.com/nfl/story/_/id/9793720/elliot-pellman-says-okay-play-nfl-suffering-concussion}.} Another colleague told a reporter “I would hear him say things in speeches like, ‘I don’t know much about concussions, I learn from my players . . . .’”\footnote{Id. (internal quotations marks omitted).}

In addition, the scientific findings of MTBIC under Dr. Pellman baffled many and garnered much criticism. In evaluating numerous studies linking concussions to serious long-term harm, “Pellman’s committee . . . repeatedly questioned and disagreed with the findings of researchers who didn’t come from their own injury group.”\footnote{Id.} In compiling their own research, Dr. Pellman’s studies “didn’t include results from hundreds of NFL players.”\footnote{Id.} A troubling 2006 MTBIC study asserted:

\begin{quote}
[M]any NFL players can be safely allowed to return to play on the day of the injury after sustaining a mild [traumatic brain injury]. [T]here were no adverse effects, and the results once again are in sharp contrast to the recommendations in published guidelines and the standard of practice of
\end{quote}

\begin{itemize}
\item \footnote{Michael Farber, \textit{The Worst Case}, \textsc{Sports Illustrated} (Dec. 19, 1994), \url{http://sportsillustrated.cnn.com/vault/article/magazine/MAG1006087/2/index.htm}.}
\item \footnote{Id. (internal quotations marks omitted).}
\end{itemize}
most college and high school football team physicians.\textsuperscript{123}

In the words of an anonymous scientist who reviewed the Committee’s work,

[t]hey’re basically trying to prepare a defense for when one of these players sues . . . . They are trying to say that what’s done in the NFL is okay because in their studies, it doesn’t look like bad things are happening from concussions. But the studies are flawed beyond belief.\textsuperscript{124}

After Dr. Pellman’s resignation, the NFL recast the MTBIC as its “Head, Neck, and Spine Medical Committee” in 2010.\textsuperscript{125}

The new members of this group, now headed by two neurosurgeons, sought to distance themselves from Dr. Pellman’s research. These new members made it very clear that “they would not use any of the old committee’s data or ongoing studies on helmets and retired players’ cognitive decline—all of which had been overseen by Dr. Pellman and blasted by Congress as ‘infected’—because they didn’t want their ‘professional reputations damaged,’” given the studies’ widely reported inaccuracies.\textsuperscript{126}

\textbf{C. The NFL’s Arguments}

The MTBIC’s failures, while appalling, now provide the basis for the plaintiffs’ strongest argument against section 301 preemption. The plaintiffs’ counsel in the current NFL action argue that the duty to prevent concussions and related brain injuries is completely separate from the CBA. While the CBA regulates a number of “health-related duties” associated with NFL teams and team doctors, the plaintiffs argued that the CBA


\textsuperscript{124} Keating, \textit{supra} note 119.


\textsuperscript{126} Hruby, \textit{supra} note 111.
does not impose any such duties on the NFL itself. Instead, these duties are wholly independent of the CBA, and arose voluntarily through the League’s creation of the MTBIC, its involvement in concussion research, and its long history of providing for player safety through rule changes and equipment requirements in order to prevent injuries.

However, the NFL maintained its stance that the CBA preempted the plaintiffs’ claims, positing that CBA terms that facially constrained only individual teams, actually applied to the “League” itself as well. The NFL argued the plaintiffs could not escape preemption by trying to make an “artificial” distinction between the NFL and its member clubs, as the NFL is simply an “unincorporated association of 32 member clubs,” engaging in a “joint enterprise” to organize and promote professional football. Essentially, the NFL argued that because the teams and the larger league are essentially the same entity, CBA terms that explicitly constrain only NFL teams are still applied to the NFL as well, and are therefore not independent of the agreement.

The NFL also pointed to a number of CBA provisions it believes preempted the former players’ claims. These included several rule-making and safety provisions. For example, Article 50, section 1(a) of the 2011 CBA requires the maintenance of a “Joint Committee on Player Safety and Welfare.” This Joint Committee is tasked with discussing “player safety and welfare relating to equipment, playing surfaces, stadium facilities, playing rules, and more.” The NFL also referenced the CBA’s grievance procedures—including a broad arbitration clause

127 Plaintiffs’ Surreply, supra note 106, at 7.

128 The plaintiffs referenced, for example, the League’s making helmets mandatory in 1943, and making it illegal to strike at an opponent’s head, neck, or face in 1980. Plaintiffs’ Master Complaint, supra note 3, at 14–19.


130 Defendants’ Motion to Dismiss, supra note 86, at 7–8.

131 Id. at 9.

132 Id.
requiring mediation of “all disputes involving the ‘interpretation of, application of, or compliance with, any provision of’ the CBA’s, player contracts, or any applicable provision of the [League] Constitution.” According to the NFL, the plaintiffs’ negligence, fraud, and misrepresentation claims “bear directly on issues addressed by the CBA’s health and safety provisions,” though such provisions do not mention concussions or brain injuries explicitly. Therefore, the NFL argued that the plaintiff’s claims should be preempted by federal law and arbitrated.

The NFL referenced several key cases to support its preemption claims under section 301. In these lawsuits, courts consistently held that the CBA preempted the plaintiffs’ claims. In Duerson v. National Football League, Inc., the estate of former Chicago Bears safety, David Duerson, brought a wrongful death suit against the NFL. The plaintiff alleged that Duerson committed suicide as a result of brain damage he sustained during his playing career. In Duerson, the court held that the CBA preempted the estate’s negligence claims. The court explained that Article XLIV, section 1 of the 1993 CBA required club physicians to advise players if their condition “could be significantly aggravated by continued performance.” The court explained that resolving the plaintiff’s claim required a determination of whether the club, by allowing Duerson to return to the field, “significantly aggravated” his injuries. Therefore, the plaintiff’s claims were “substantially dependent” on the interpretation of Article XLIV, implicating LMRA section 301

133 Id. at 10. This specific provision can be found at Article 50, §1(a) of the 2011 CBA.
134 Id. at 16.
136 Id. at *1. Duerson suffered from the effects of CTE, including “intense headaches, lack of short term memory, language difficulties, vision trouble, and problems with impulse control.” Id. The plaintiff’s complaint alleged counts of negligence, fraudulent concealment, conspiracy to publish false information, and negligent failure to warn, against the NFL. Id.
137 Id. at *4.
138 Id.
139 Id. (citation omitted).
and requiring federal jurisdiction.\textsuperscript{140} The court additionally hypothesized that other CBA provisions addressing player safety may create a general “duty on the NFL’s clubs to monitor a player’s health and fitness to continue to play football,” a duty more than broad enough to include the plaintiff’s claims in Duerson.\textsuperscript{141} The court further noted “preemption is still possible even if the duty on which the claim is based arises independently of the CBA, so long as resolution of the claim requires interpretation of the CBA.”\textsuperscript{142}  

The NFL also cited to the aforementioned decision in Stringer v. National Football League.\textsuperscript{143} In Stringer, the widow of Korey Stringer, a former Minnesota Vikings offensive lineman, filed a five-count complaint against the League after Stringer died due to complications from heat stroke and exhaustion.\textsuperscript{144} The plaintiff argued that the NFL had no contractual duty to protect players from heat-related illnesses, and that while individual teams were responsible for their players’ health and safety, the NFL voluntarily assumed the duty to “provide complete, current, and competent information and directions to NFL athletic trainers, physicians, and coaches about heat-related illnesses.”\textsuperscript{145} This duty was assumed, Stringer argued, when the League issued a set of “Hot Weather Guidelines” for the protection of players.\textsuperscript{146}  

Although the court agreed that the wrongful death claim did not arise under the CBA, it accepted the NFL’s argument that the CBA preempted Stringer’s wrongful death claim because resolution of the claim was still “substantially dependent” on the CBA.\textsuperscript{147} The district court found that “the degree of care owed by

\textsuperscript{140} Id. at *6.  
\textsuperscript{141} Id. Such provisions include those requiring each team to have a board-certified orthopedic surgeon, requiring that the NFL pay for any medical care rendered by club staff, and provisions regarding certification requirements for trainers. Id.  
\textsuperscript{142} Id.  
\textsuperscript{143} Id. at *5 (citing Stringer v. Nat’l Football League, 474 F. Supp. 2d 894 (S.D. Ohio 2007)).  
\textsuperscript{144} Stringer, 474 F. Supp. at 898.  
\textsuperscript{145} Id. at 905.  
\textsuperscript{146} Id.  
\textsuperscript{147} Id. at 908–09. More specifically, the court found the plaintiff’s claim
the NFL in republishing the Hot Weather Guidelines . . . and what was reasonable under the circumstances, must be considered in light of pre-existing contractual duties imposed by the CBA on the individual NFL clubs concerning the general health and safety of the NFL players."148 In deciding that Stringer’s claims were “inextricably intertwined” with the CBA,149 the majority noted a CBA provision requiring team trainers to be “certified by the National Athletic Trainers Association.”150 Since the “degree of care” that the NFL owed by republishing the Hot Weather Guidelines was dependent on whether or not team trainers were educated on treating heat-related illnesses as part of the certification process, the court decided the CBA must be interpreted, and the plaintiff’s claims preempted.

The NFL also relied on Williams v. National Football League, in which several players, including the plaintiffs, tested positive for the banned diuretic bumetanide.151 The players, who all testified they took StarCaps diet pills in order to control their weight, also stated that they did not know the supplement contained the banned diuretic.152 Plaintiffs argued that despite warnings about supplements, a hotline that provided banned substance information, and the League’s strict liability policy on banned substances—the NFL owed a duty to the plaintiffs because the NFL and its drug policy administrator knew StarCaps contained bumetanide, yet failed to disclose it.153 Failure to advise players of this fact, the plaintiffs argued, constituted a breach of implicated CBA Art. XLIV §2, requiring the certification of training staff, including instruction on how to “to prevent, recognize, and treat heat-related illness, id. at 910., and Art. XLIV §1, requiring team physicians to inform players if their physical condition “will be ‘significantly aggravated by continued performance,’” id.

148 Id. at 910.
149 Id. at 908–09.
150 Id. at 910. The court also referenced Article XLIV, section 1, (the provision at issue in Duerson) requiring team physicians to advise athletes if a further game action would “significantly aggravate” the player’s injuries.
151 582 F.3d 863 (8th Cir. 2009).
152 Id. at 871.
153 Id.
the League’s fiduciary duty to its players. Plaintiffs also brought claims for negligence, gross negligence, and misrepresentation against the League.

However, the court held that the CBA preempted each of the players’ claims. Even though the players alleged that the duty to provide “an ingredient-specific warning for StarCaps” arose not under the CBA, but under Minnesota law, the court held that whether the NFL owed this duty to the players “[could] not be determined without examining the parties’ legal relationship and expectations as established by the CBA . . . .” Further, the court held that the CBA preempted plaintiffs’ misrepresentation claims because “the question of whether the Players [could] show that they reasonably relied on the lack of a warning that StarCaps contained bumetanide cannot be ascertained apart from the terms of the [League’s drug policy].” Finally, the court held that the CBA also preempted plaintiffs’ claims for intentional infliction of emotional distress, because determining whether the NFL engaged in “outrageous” conduct required an evaluation of the League’s drug policy, a part of the CBA.

D. The Plaintiffs’ Arguments

The plaintiffs and the NFL differed substantially in their interpretations of these key cases. In arguing their claims shouldn’t be dismissed, the plaintiffs attempted to distinguish the NFL’s precedent cases, including Duerson, Stringer, and Williams. For example, the plaintiffs pointed to a key difference between their lawsuit and Duerson. They argued that unlike their own claims, the estate in Duerson never alleged that the NFL, as a whole, assumed a duty of care independent from that of the clubs and team doctors governed by the CBA. In Duerson, the plaintiff barely referenced any duty assumed by the
League itself, only referring to a “generic duty ‘to keep [players] reasonably safe.’” The current plaintiffs also highlighted the NFL’s evasion of what the plaintiffs considered the “fundamental flaw” of Duerson: that the court merely speculated that CBA provisions might permit the League to exercise a lower standard of care, without ever identifying an “actual dispute” over a CBA term.

The plaintiffs also attempted to distinguish the present case from Stringer. First, they argued that unlike the present litigation, the plaintiff in Stringer did not allege that the NFL misled athletes, making that case inapplicable to the player’s fraud claims here. Second, the plaintiff in Stringer alleged (arguably to her detriment) that “[a]thletic trainers in the NFL serve as the first line of treatment for players. It is their initial responsibility to recognize and treat football-related injuries or conditions, including heat-related illness.” Therefore, preemption was only required in Stringer because Stringer’s claims referred specifically to a breach of duty by team trainers, implicating the CBA, which explicitly governs team medical staff. By contrast, the plaintiffs argued, their concussion litigation sought to establish a duty wholly independent from that of team medical personnel. Therefore, no interpretation of the CBA would be necessary to resolve their claims.

The concussion plaintiffs distinguished Williams based on divergent facts. Their attorneys focused on the difference between the “voluntary assumption of duty” on the part of the NFL and the assumption at issue in the concussion litigation. As previously mentioned, the current plaintiffs asserted that the NFL assumed a duty of care to protect athletes from brain trauma harm, arising from its historical assumption of duty for player

---

160 Id. at 21.
161 Id. at 22–23 (emphasis added).
162 Id. at 23.
164 Plaintiffs’ Surreply, supra note 106, at 23.
165 Id.
166 Id. at 24.
167 See supra text accompanying notes 106–09.
care and safety, and the NFL’s formation of the MTBIC in 1994.\textsuperscript{168} By contrast, in Williams, “the challenged steroid testing regime was set forth in a comprehensive written ‘Policy’ that the CBA ‘expressly incorporate[d].’”\textsuperscript{169} Because the NFL’s drug policy was therefore part of the CBA, the CBA was obviously implicated in resolving the plaintiffs’ claims, and preemption was proper. In the present case, by contrast, the plaintiffs argued that “the NFL identify[d] no written policy specifically governing head injuries, and certainly not one assigning responsibility for those injuries to the NFL.”\textsuperscript{170}

The NFL addressed many of the plaintiffs’ arguments,\textsuperscript{171} but not what attorney Paul Anderson considered “the strongest theory in the plaintiffs’ case”—that the creation of the MTBIC Committee, to “spearhead concussion research,” represented an independent assumption of duty by the NFL.\textsuperscript{172} Reviewing Third Circuit precedent, Anderson concluded that the NFL did create an independent duty through creation of the Committee, and therefore, “the case law should have foreclosed the dismissal of all negligence and fraud-based claims that relied upon the [MTBIC’s] conduct.”\textsuperscript{173} Referencing the news report that Judge Brody threatened to dismiss the plaintiffs’ claims on preemption grounds, Anderson asserted that such a decision would be “an unpredictable shocker,” and hypothesized that the rumor’s source might have been “jockeying for a settlement in an attempt to counter the public’s perception that this deal was lousy.”\textsuperscript{174}

Federal precedent supports Anderson’s position: that the plaintiffs’ claims cannot be preempted by the CBA since the NFL assumed a duty of care through the MTBIC. In Trans Penn Wax Corp. v. McCandless, the Court of Appeals for the Third Circuit held that section 301 of the LMRA did not preempt the plaintiff’s claims.\textsuperscript{175} The plaintiffs, former employees of Trans Penn, were

\textsuperscript{168} Plaintiffs’ Surreply, \textit{supra} note 106, at 6.
\textsuperscript{169} \textit{Id.} at 24.
\textsuperscript{170} \textit{Id.} at 25.
\textsuperscript{171} Defendants’ Reply Memorandum, \textit{supra} note 129, at 15–18.
\textsuperscript{172} Paul D. Anderson Consulting, \textit{supra} note 83.
\textsuperscript{173} \textit{Id.}
\textsuperscript{174} \textit{Id.}
\textsuperscript{175} 50 F.3d 217, 233 (3d Cir. 1995).
given a written “contract” (separate from the parties’ CBA) by their employer guaranteeing their jobs, but were subsequently fired less than a year later. The court held that the plaintiff’s claims were not preempted because they never alleged a violation of duties assumed specifically in the CBA. The court reached the same conclusion in *Kline v. Security Guards, Inc.*, noting that the fact that the CBA was simply related to the plaintiff’s claims was not sufficient for the court to find preemption. There, the plaintiff employees alleged their employer’s surveillance practices, including the use of microphones to record oral communications, amounted to several torts including invasion of privacy. Judge Stapleton asserted that “the mere fact that we must look at the CBA in order to determine that it is silent on any issue relevant to Appellants’ state claims does not mean that we have ‘interpreted’ the CBA” for Section 301 purposes. Noting that the CBA did not mention the terms at issue (e.g., “surveillance,” “video cameras,” or “microphones”), the court found that no “interpretation” of the CBA was necessary to adjudicate plaintiffs’ claims.

On May 14, 2014, the *In re National Football League Players’ Concussion Injury Litigation* plaintiffs received welcome news that at least one District Court judge accepted similar arguments against section 301 preemption. Judge Catherine D. Perry remanded *Green, et al. v. Arizona Cardinals Football Club, LLC*, a suit brought by three former players (and their spouses) against their former team, to state court, finding that “the merits of the plaintiffs’ claims can be evaluated without

---

176 *Id.* at 221.
177 *Id.* at 232.
178 386 F.3d 246, 250 (3d Cir. 2004).
179 *Id.* at 256.
180 *Id.* at 250. The employer alleged these claims were preempted by CBA clauses relating to “management rights” and “shop rules.” *Id.* at 257.
181 *Id.* at 256.
182 *Id.* at 259.
interpreting [the 1977 or 1982 CBAs]....” Judge Perry found the bargaining agreements did not bear on negligence claims “premised upon the common law duties to maintain a safe working environment, not to expose employees to unreasonable risks of harm, and to warn employees about the existence of dangers of which they could not reasonably be expected to be aware.” Similarly, Judge Perry noted that the players’ negligent misrepresentation and fraudulent concealment claims could be resolved without CBA interpretation, as they arose from common law duties of an employer “‘to inform himself of those matters of scientific knowledge’ that relate to the hazards of his business, and relay that knowledge to his employees.” Judge Perry’s decision may have enormous effects on the future of In re National Football League Players’ Concussion Injury Litigation. As one journalist noted, “[t]he outcome [of Green] also could result in the plaintiffs in the settled case to quit trying to persuade Judge Anita Brody to approve the settlement, opting instead to proceed with the litigation. If the players in that case secure the same victory Roy Green and others have realized in Missouri, the value of the claims would potentially skyrocket.”

Since the NFL has assumed a general duty to protect its athletes, and more specifically a duty to warn them of the risks of neurological injury (through its formation of the MTBIC), the plaintiffs’ claims should not be preempted. As in Kline, the NFL-NFLPA CBA makes no mention of the specific duty at issue. Just as terms like “surveillance” or “microphones” were not mentioned in the Kline CBA, discussion of concussions or brain injuries do not appear in the NFL-NFLPA CBA with any reference to the NFL itself, only to issues relating to team doctors. Though the plaintiff’s claims in Stringer openly arose

---

185 Id. at *10.
186 Id. at *16.
out of team duties to their athletes, the plaintiffs here look to the NFL itself. While the players’ claims may relate to the CBA terms the NFL highlighted (such as the creation of the Joint Committee on Player Safety and Welfare), such terms do not require interpretation to resolve the claims. This distinguishes the present litigation from *Williams*, where the drug policy at issue was expressly incorporated into the collective bargaining agreement. The CBA’s arbitration clause only applies to disputes involving the CBA, player contracts, and the League Constitution. And as in *Trans Penn Wax Corp.*, none of the plaintiffs’ claims refer to explicit duties in the bargaining agreement. None of the CBA clauses proffered by the NFL relate specifically to a League concussion policy. Therefore, LMRA section 301 should not apply.

IV. PROVING CAUSATION AND THE LOSS OF CHANCE DOCTRINE

A. Causation Issues: Tobacco Litigation, Team Trainers, Assumption of Risk and Contributory Negligence

Even if concussion litigation plaintiffs avoid preemption, they must still prove causation in order to successfully prove negligence. This requires plaintiffs to demonstrate that the “head [injuries] players sustained while playing in the NFL” directly caused the plaintiffs’ current health problems. It may be extremely difficult for plaintiffs to show that the game of professional football caused long-term cognitive injuries, especially where high school or collegiate-level athletics, non-football activities, genetics, and diet also play a large role in the incidence of these diseases.

---

189 Defendants’ Motion to Dismiss, *supra* note 86, at 7–9.


In order to prove causation, some scholars have drawn parallels between the concussion litigation and big-tobacco lawsuits.\(^{192}\) Both the NFL and the tobacco industry sought to discredit growing scientific data indicating a causal link to long-term illness and formed research committees to “refute the mounting evidence load that protected the vitality of their products.”\(^{193}\) Despite the attractiveness of using big-tobacco litigation as a framework for pursuing concussion lawsuits against the NFL, some do not believe it is an apt comparison. Attorney Joseph Hanna succinctly explained the problem with comparing the tobacco litigation and the former players’ concussion claims:

> [U]nlike tobacco use, the effect of individual concussions on a football player remains unclear. Further, the NFL retains medical personnel who are employed specifically to detect and prevent player injuries, whereas smoker plaintiffs were given no such attention. Lastly, because NFL players could have sustained permanent mental injuries at any point in their career (high school, college, etc.), proving the causal chain—i.e., that the NFL’s failure to warn resulted in injury—is difficult at best.\(^{194}\)

Although statistical evidence linking concussions to long-term disease such as CTE is becoming increasingly overwhelming,\(^{195}\)

---

\(^{192}\) This idea gained traction after a 2009 congressional hearing was conducted to evaluate the League’s concussion policy. There, Representative Linda Sanchez of California “analogized the denial of a causal link between NFL concussions and cognitive decline to the tobacco industry’s denial of the link between cigarette consumption and ill health effects.” Joseph Hanna & Daniel Kain, *The NFL’s Shaky Concussion Policy Exposes the League to Potential Liability Headaches*, 21 ENT., ARTS & SPORTS L.J., no. 3, Fall/Winter 2010, at 33, 34.


\(^{195}\) For example, Dr. Ann McKee has studied the brains of at least forty-
proving causation would require the NFL to answer questions avoided by the proposed settlement; including exactly what the NFL knew about the long-term effects of football-related brain injuries, when they knew it, and whether they deliberately spread misinformation or remained willfully blind to the problem. Without this information, the NFL cannot be held accountable. Fortunately for the plaintiffs, there is already ample evidence that the NFL ignored or dismissed mounting evidence linking concussions to neurological damage. In 1994, NFL Commissioner Paul Tagliabue responded to concerns over concussions by stating “the number [of concussions] is relatively small . . . . [T]he problem is a journalist issue.” Further, the Pellman-lead MTBIC asserted that “[r]eturn to play does not involve a significant risk of a second injury either in the same game or during the season,” and argued that individuals “prone to delayed or poor recovery after MTBI” are actually “selected out” of organized football, and never reach the NFL. The NFL also rejected the American Academy of Neurology’s 1997 return-to-play guidelines, including the suggestion that concussed players not return to the field until being symptom-free for at least a week. In addition, MTBIC co-chair Ira Casson’s famous

six former NFL players has found CTE in forty-five of them. Perhaps unsurprisingly, when asked to speak about her research in front of the MTBIC in 2009, Dr. McKee was allegedly confronted with aggressive questioning and mocking interruptions, especially from committee co-chair Ira Casson. Transcript, League of Denial: The NFL’s Concussion Crisis, FRONTLINE, http://www.pbs.org/wgbh/pages/frontline/sports/league-of-denial/transcript-50/ (last visited Mar. 28, 2014); see also Mark Fainaru-Wada & Steve Fainaru, League of Denial, SPORTS ILLUSTRATED (October 7, 2013), http://sportsillustrated.asia/vault/article/magazine/MAG1208801/index.htm.

196 See Campbell, supra note 13.


198 Id.

199 Id.

200 MTBIC doctors criticized the guidelines as not being supported by ample research, stating, “[W]e see people all the time that get knocked out briefly and have no symptoms.” James C. McKinley Jr., Invisible Injury: A Special Report; A Perplexing Foe Takes an Awful Toll, N.Y. TIMES (May 12,
2007 “no, no, no” denial when asked about any link between football and depression, dementia, Alzheimer’s, or other long-term problems, further evidences that the League at least turned a blind eye to the problem. It may be nearly impossible to show that the NFL alone caused the plaintiffs’ injuries, but mounting evidence suggests that it did spread misinformation, failed to disclosure unpopular data, and relied on poor science. This, when coupled with the NFL’s strong denial of any causal links between playing football and long-term brain injuries, indicates that the NFL prevented players from making informed decisions about their health and contributed to the prevalence of such harm.

Importantly, the fact that NFL teams retain physicians does not mean that the League is not responsible for plaintiff’s injuries. While NFL teams do retain medical personnel “to detect and prevent player injury,” the physicians’ efforts do not preclude a finding that the League caused the litigation plaintiffs’ injuries, due to the doctors’ inherent conflicts of interest. This conflict of interest exists because both trainers and doctors are paid by team management and thus they face pressure to return athletes to the field as soon as possible, hoping to keep their employer happy and retain their title as an “official” team medical provider or physician group. Although the NFL added independent neurological consultants to the sidelines in 2013, this does not solve all the problems associated with concussion diagnoses, and obviously does little to help the retired players comprising the plaintiffs in the current lawsuit. Some players do not show immediate symptoms, making an on-scene neurologist


202 Kain, supra note 188, at 728–29.

203 Id. at 708–09.

204 Curtis Crabtree, NFL Will Have Independent Neurological Consultants on Sidelines Next Season, NBC PROFOOTBALLTALK (Jan. 31, 2013, 4:10 PM), http://profootballtalk.nbcSports.com/2013/01/31/nfl-will-have-independent-neurological-consultants-on-sidelines-next-season/.
ineffective.\textsuperscript{205} Also, typical sideline chaos can cause a breakdown in protocol, and players tend to refuse to leave the game.\textsuperscript{206} Evidence, therefore, suggests that team trainers and physicians were not particularly effective at handling and treating injuries arising from concessions.

Two of the most potent defenses that the NFL would likely raise are assumption of risk and contributory negligence.\textsuperscript{207} Such defenses arise out of the belief that many football players, despite knowing that they risk injury or further injury, still “tough it out” on the field and fail to be honest with their team’s trainers and physicians. However, it is not clear that the NFL would be successful. In order to raise assumption of risk as a defense, the plaintiffs must, “knowing the risk and appreciating its quality, voluntarily [choose] to confront it.”\textsuperscript{208} If plaintiffs voluntarily place themselves in harm’s way, despite the risks, they cannot later claim negligence if they are injured. Given the deliberate misinformation provided by the MTBIC, and the potential existence of data and information allegedly withheld by the NFL, proving athletes had “actual knowledge” of the risks that arose from concussions would be difficult to prove.\textsuperscript{209} In other words, while it is reasonable to argue that football players assume the risk of being concussed, it will be challenging for the NFL to argue players actually knew how these concussions would ultimately affect them, even if a substantial number of athletes may have tried to play through their injury regardless.

Contributory negligence may provide a better defense for the League.\textsuperscript{210} In 2007, the NFL distributed a pamphlet to players giving players the burden of notifying team doctors and trainers of possible concussion symptoms, and advising that players should not return until they are entirely free of symptoms.\textsuperscript{211}

\begin{footnotes}
\item[206] Id.
\item[207] See Wong, supra note 20.
\item[208] Dan B. Dobbs et al., \textit{THE LAW OF TORTS} §235 (2d ed. 2000).
\item[209] See Hanna & Kain, supra note 187, at 11.
\item[210] See id. at 11–12.
\item[211] See National Football League, \textit{NFL Outlines for Players Steps Taken
However, it appears that many players have ignored this advice, likely contributing to their risk of long-term injury. The NFL has a long history of lauding “toughness” and the ability to play through injury, where if a player left a game with a “slight concussion,” they weren’t “giving it all” for their team. For example, quarterback Peyton Manning admitted to intentionally underperforming on baseline concussion tests in order to lower his return-to-play standards. By lowering his baseline, Manning hoped to return from concussions earlier than recommended, as he might later be able to meet his baseline after an injury, even if he were still suffering the concussion’s effects. Though some players may not have acted in the best interests of their health, this behavior cannot be viewed in a vacuum and should not affect the outcome of concussion litigation. It would be difficult for the NFL to prove that many of the plaintiffs hid concussive injuries, and it is likely that many players concealing brain injuries would not have done so absent the League’s deliberately cultivated “tough-it-out” culture and frequent minimization of concussion risks.

In addition, there is considerable incentive for NFL players to play down their own injuries. NFL contracts are not guaranteed beyond the season in which an injury occurs if the player cannot pass his team physical before the subsequent season, and football players can be terminated at-will if the team decides another player would increase team performance. In order to keep their jobs then, many players do not reveal if they are suffering from any concussion symptoms. Furthermore, even if players report

---


213 *Id.*


215 Kain, supra note 187, at 710–11.

216 For example, linebacker Dan Morgan, who had endured a number of
their symptoms, those players are still under significant pressure to return before becoming completely asymptomatic in order to stay in their club and coach’s good graces, and retain their roster position.217

Faced with these systemic issues, and the lack of information and misinformation provided by the League to its athletes, it is unlikely that the NFL could prove contributory negligence. For players entering the NFL today, however, both assumption of risk and contributory negligence arguments might bar future lawsuits, since much more information is becoming known and available to current and future professional football players. However, even if the plaintiffs here survive these defenses, proving the NFL caused the plaintiffs’ injuries will require some creativity, as a wide variety of factors outside of professional football contribute to the long-term illnesses at issue.

B. The Loss of Chance Doctrine as a Basis for Proving Causation

Outside of the NFL concussion litigation context, proving that any one actor caused an illness is extremely difficult. The process of evaluating disease causation is “typically multifactorial,” as “a large constellation of factors and variables coalesce to produce a particular person’s unique set of illness experiences.”218 It is hard to find liability where elements such as genetics, upbringing, personal habits, and environment all play an indeterminate role in causation, in addition to any tortious activity. This issue is even more complex for the plaintiffs here, who not only have to prove the NFL caused neurological injury, but must separate its negative contributions to players’ health from those of other levels of football (youth leagues, high school, college, etc.), genetic predisposition, previous head trauma from accidents

conussions and missed significant playing time, restructured his contract bonus in order to remain on the Carolina Panthers through a calculation based on number of games played. Hanna & Kain, supra note 186, at 12.

217 See Kain, supra note 187, at 711–12.

unrelated to sports, and abuse of drugs or alcohol.\textsuperscript{219}

In order to prove causation, the plaintiffs’ lawyers had intended to use “alternative causation (or ‘multiple-causation’) theory.”\textsuperscript{220} Alternative causation theory extends liability when multiple actors are negligent (for example, the NFL, the NCAA, and Riddell— the company that manufactures the League’s helmets), but only one actually caused the harm, and it is impossible to discern which.\textsuperscript{221} In the seminal case \textit{Summers v. Tice}, two hunters negligently fired their shotguns while hunting quail, injuring the plaintiff third hunter.\textsuperscript{222} Though the court could not determine which defendant actually shot the plaintiff, it found that both should be found jointly liable.\textsuperscript{223} Therefore, once the negligence of the multiple tortfeasors is established, the burden shifts to each defendant to show they did not cause the plaintiff’s harm.\textsuperscript{224} If the defendants cannot meet this burden, both will become liable under alternative causation theory, even though one negligent actor may have caused no damage at all.\textsuperscript{225} In justifying its decision, the court in \textit{Summers} noted that defendants typically have better access to evidence of the actual cause than plaintiffs, and that placing the burden of proof on plaintiffs would leave many without remedy.\textsuperscript{226} Because the “innocent” negligent actor made it difficult (or impossible) for the plaintiff to prove causation, “the defendants, rather than the innocent plaintiff,


\textsuperscript{221} Dobbs et al., \textit{supra} note 202, §193.

\textsuperscript{222} Summers v. Tice, 199 P.2d 1 (Cal. 1948). For another application of multiple causation theory, see Landers v. E. Tex. Salt Water Disposal Co., 248 S.W.2d 731 (Tex. 1952).

\textsuperscript{223} \textit{Summers}, 199 P.2d at 5.

\textsuperscript{224} Dobbs et al., \textit{supra} note 202, §193.

\textsuperscript{225} \textit{Id.}

\textsuperscript{226} \textit{Summers}, 199 P.2d at 4.
should bear the loss.”

As attorney Paul Anderson hypothesized, plaintiffs using alternative causation will argue that while other actors contributed to the plaintiffs’ harm, “[a]s the industry leader, it is appropriate for the NFL to be held jointly and severally liable for substantially contributing to the players’ injuries.”

Unfortunately, alternative causation remains a tenuous strategy with no guarantee of success, and even Anderson admitted that it was a “legal stretch.” This is in part because successfully finding alternative causation liability requires the presence of more than one negligent actor. In this case, “proof that one of the two actors is negligent simply does not aid the plaintiff at all.”

In order for concussion plaintiffs to use this theory successfully, they must be able to show that actors other than the NFL were also negligent, and in some jurisdictions, must have all tortfeasors joined as defendants, or show that each defendant created “qualitatively similar risks of harm.”

In the alternative, plaintiffs’ counsel may want to argue for an extension of the “loss of chance” (or “lost chance”) doctrine. Used almost exclusively in medical malpractice suits, loss of chance permits plaintiffs to recover for tortious actions substantially reducing their chance of survival, even if that chance was less than fifty percent. In *Herskovits v. Group Health*, the court found the plaintiff successfully proved causation by showing the defendant’s failure to diagnose the plaintiff’s lung cancer substantially reduced the plaintiff’s chance of survival, even though Herskovits had less than a fifty-percent chance of living regardless of when the diagnosis was made.

Use of the lost chance doctrine would mitigate the harshness

---

227 Dobbs et al., supra note 202, §193.
228 Berman, supra note 214 (emphasis in original).
229 Id.
230 Dobbs et al., supra note 202, §193.
231 Id.
of the usual “all or nothing” causation standard, and allow concussion plaintiffs to recover damages even if a plaintiff endured additional brain trauma outside the scope of their NFL employment, such as through college football or a car accident. Though loss of chance doctrine has been applied where malpractice reduces the plaintiff’s chance of survival (such as a missed or late diagnosis), courts generally prefer to use this doctrine when the individual suffered serious harm and partially contributed to the tortious activity, but the harm may have occurred absent the malpractice.

In *Wendland v. Sparks*, the Supreme Court of Iowa held that “loss of chance of less than 50% is compensable,” where the defendant doctor failed to perform CPR to resuscitate a patient. There, even though the patient had entered cardiorespiratory arrest and had drawn (what would be) her last breath prior to the defendant’s negligence, the court found the doctor liable for his failure to attempt resuscitation. Noting that the patient had been successfully resuscitated multiple times prior, the fact that the patient never made a “no code” request, and that the doctor acted against the known wishes of the patient’s husband, the court found that even though “the chances of successful resuscitation were questionable, and any recovery for wrongful death would be severely limited . . . even a small chance of survival is worth something.” The loss of chance has been likened to the loss of a lottery ticket—although the ticket “represents a less than even chance of recovery,” the ticket nonetheless has “clear market value.” In the concussion litigation context, the “ticket” represents someone who have may have suffered from a concussion outside of the NFL but nonetheless might have been healthy—or at least healthier—had the NFL not withheld

---

235 The traditional standard permitted a recovery only upon a preponderance of evidence—a fifty-one percent certainty of causation. See Kaufman, supra note 35.

236 Phillips, supra note 228, at 85.

237 574 N.W.2d 327, 333 (Iowa 1998).

238 Id. at 328.

239 Essentially, this is a “do not resuscitate” request.

240 Id. at 328, 332 (emphasis in original).

241 Kaufman, supra note 35.
information and opposed reform, leading to even more concussions and neurological injuries.

With the loss of chance doctrine gaining traction in a number of state courts, the same policy reasons behind its acceptance support an expansion of the doctrine to other claims, and specifically, to concussion lawsuits. Loss of chance is not confined to suits dealing with negligent diagnoses and has been accepted for claims outside the malpractice context, albeit in narrow circumstances. A number of scholars have advocated for a limited extension of the doctrine to tort cases more broadly, arguing for the increased use of “probabilistic causation,” where tortfeasors are liable in proportion to the harm contributed, especially in mass tort contexts. Professor Glen O. Robinson noted the “lagged effects” of harm in toxic tort cases (an issue applicable to concussion litigation, where the long-term effects of traumatic brain injuries often arise years later), and posited that the search for “deterministic causes” (such as “substantial factor” or “but for” causation) was “both artificial and misleading,” arguing that “the basic objectives of tort law are better served if liability is based on risk of injury, than if it is based on the actual occurrence of a harm.” Even a “narrow” formulation of loss of chance, limiting the doctrine to “failure[s] to protect a person from a preexisting condition,” may permit recovery for concussion plaintiffs’ failure to warn claims, as failures to diagnose—like failures to warn—deal with “protection

---


243 Wendland v. Sparks, 574 N.W.2d 327, 332 (Iowa 1998). See, e.g., Gardner v. Nat’l Bulk Carriers, Inc., 310 F.2d 284 (4th Cir. 1962) (where defendants were liable for a reduction in a deceased seaman’s chance of survival after failing to search for him); Hake v. Manchester Twp., 486 A.2d 836 (N.J. 1985) (where defendants were liable for reduction in the chance of survival of deceased juvenile arrestee who had attempted suicide, by failing to perform CPR).

244 Fischer, supra note 38, at 606.


246 Robinson, supra note 239, at 780–81, 783.
against an external risk.”

One prevailing policy justification for extending loss of chance doctrine is fairness. This argument has particular force in the concussion litigation context, where a traditional causation rule might bar concussion plaintiffs from recovery. Due to the harshness of the traditional “more-likely-than-not” causation standard, some advocate for loss of chance under fairness principles, where wrongful conduct not only increased the incidence of a future illness, but prevented a determination of whether the illness would have occurred “but for” the wrongdoing. This rationale applies to the concussion litigation context, where the alleged misrepresentations and omissions of the NFL and MTBIC may have increased the likelihood of long-term neurological injury by failing to properly address concussion concerns, and spreading misinformation, causing players to misjudge the risks involved. In addition, the possibility that the League intentionally concealed evidence about the seriousness of concussion injuries makes causation difficult to prove: how do we know if the plaintiffs’ long-term injuries would have occurred absent the NFL’s and the MTBIC’s alleged deception?

Deterrence is another popular policy justification for the extension of loss of chance to concussion litigation. The use of loss of chance prevents tortfeasors who have caused less than fifty percent of the plaintiff’s harm from escaping liability. By contrast, a more traditional rule incentivizes potential defendants who might substantially contribute to an injury, but not necessarily “more-likely-than-not” have caused it, to avoid additional precautions. For example, under a traditional causation rule, a player who received four concussions in college and two in the NFL could not argue the League caused more than

---

247 Fischer, supra note 38, at 606, 610.
248 Id. at 626–27.
249 Loss of chance should also be limited to cases where the duty owed by the defendant arose from a “special relationship,” a standard the NFL litigation likely satisfies. Joseph H. King, Jr., “Reduction of Likelihood” Reformulation and Other Retrofitting of the Loss-of-a-Chance Doctrine, 28 U. MEM. L. REV. 491, 535 (1998).
250 Fischer, supra note 38, at 627–35.
251 Id. at 605–06, 632.
fifty percent of his concussion-related injuries. Loss of chance allows him to argue that the NFL should still be liable for a third of the injury because it reduced his chances of survival or of living a healthy life.

Though opponents of this rationale argue that the traditional more-likely-than-not rule actually incentivizes actors to take additional steps because they may be liable for more harm than they actually caused,252 this argument does not have as much force in the concussion litigation context, where it is more difficult to prove the NFL caused a majority of the alleged injuries. Still, a substantial number of states decline to use loss of chance at all, let alone an expansive application.253 The primary concern is that an extension of loss of chance will be highly difficult to limit. Because an extension of loss of chance “can apply to all cases where a tortfeasor creates a risk of harm and it is uncertain whether the harm has already occurred or will occur in the future,” there is fear that the loss of chance would “swallow the [traditional all-or-nothing] rule,” rather than remain the exception.254 Such jurisdictions fear that permitting loss of chance recovery allows the compensation of speculative or uncertain injuries. Therefore, a significant number of jurisdictions refuse to consider loss of chance when the victim had less than a fifty percent chance of survival,255 which would be problematic for concussion plaintiffs.

The Supreme Court of Florida rejected loss of chance in Gooding v. University Hospital Building, Inc., concerned that “[r]elaxing the causation requirement might ... create an injustice. Health care providers could find themselves defending cases simply because a patient fails to improve or where serious disease processes are not arrested because another course of action could possibly bring a better result.”256 This argument is problematic, however, as loss of chance does not create a

252 Id. at 627–28.
254 Fischer, supra note 38, at 606–07.
255 Férot, supra note 248, at 611–12.
“heightened duty” for all actors, but instead seeks to hold negligent tortfeasors liable for the reduction in a victim’s chance of recover. Nonetheless, many state courts remain concerned that innocent parties could ultimately become liable for injuries they didn’t clearly cause. In its Restatement (Third) of Torts, the American Law Institute stated it would not take a stance on loss of chance, noting that it would be a “drastic” expansion of traditional doctrine and left the issue to state courts.

While loss of chance has been traditionally used only in individual suits, this doctrine should be applied here should future NFL concussion plaintiffs be certified as a class. There are compelling justifications for applying the loss of chance doctrine to class actions. In addition, loss of chance can actually facilitate class certification for former NFL class action plaintiffs, as it may increase the chance of satisfying the certification “commonality” requirement under Federal Rule of Civil Procedure 23(b). Loss of chance has the capacity to “smooth over many of the differences” between individual class members by comparing them all to a “baseline in order to determine damages” (for example, the incidence of brain injury in non-NFL playing football players), and then applying those damages pro-rata to all members of the class. Without the use of the relevant baseline in loss of chance determinations, courts would have to examine class members on an individual basis to determine the likelihood of injury absent the tortious behavior, a process that would lead to a wide variety of results that would likely “destroy the commonality holding a potential class together.”

Loss of chance, then, seems like a potentially viable alternate

---

257 Férot, supra note 247, at 615.
259 See generally Speaker, supra note 236, at 349. Professor Speaker argues that class action lawsuits are not only appropriate for the application of loss of chance, but are actually a “better fit” for the doctrine than individual claims. Id. at 353–54.
260 Id. at 366–68.
261 See Scheuerman, supra note 33, at 104.
262 See Speaker, supra note 236, at 365–68.
263 Id. at 367.
basis to alternative causation theory and will allow former NFL players to plead sufficient causation, despite the existence of other contributing factors. Though any extension of the loss of chance doctrine should be crafted with care, loss of chance is supported by principles of fairness and deterrence, and it provides an opportunity for former players to recover despite the role of other potential causes. Finally, loss of chance may aid in the class certification context by providing a baseline by which courts can apportion damages with less reference to the individualized concerns of specific plaintiffs.

V. CONCLUSION

The proposed settlement between the NFL and over 4,500 former football players is currently an insufficient remedy to the extensive damage caused by the League’s inaction, failure to warn, and spread of misinformation, with regard to concussions and repeated brain trauma. While the potential settlement could provide some immediate aid to former players suffering from the most serious effects of repeated head trauma, the amount itself will likely be inadequate compensation for most players. Further, the proposed settlement allows the NFL to avoid the discovery process and any admission of wrongdoing, allowing the NFL to escape a good deal of bad publicity and public pressure to reform. It is due to these terms that the proposed settlement is inadequate and therefore should be denied.

Given the proposed settlement’s insufficiencies, the plaintiffs may need to opt out and pursue litigation to receive adequate damages, unless the settlement is significantly reworked. In order for future plaintiffs to be successful they will need to be able to argue that section 301 of the LRMA did not preempt their claims. It is likely plaintiffs would avoid preemption, since the League—

---

264 The concern is that an extension of loss of chance will be highly difficult to limit. Because an extension of loss of chance “can apply to all cases where a tortfeasor creates a risk of harm and it is uncertain whether the harm has already occurred or will occur in the future,” there is fear that the loss of chance would “swallow the rule,” rather than remain the exception, “with probabilistic causation completely supplanting the traditional” all-or-nothing rule. Fischer, supra note 38, at 606–07.
through its creation of the Mild Traumatic Brain Injury Committee—assumed a duty of care that is independent of the NFL-NFLPA’s collective bargaining agreement. Unlike precedent cases such as Duerson and Williams, the plaintiffs’ claims in In re National Football League Players’ Concussion Injury Litigation arise under a duty of care completely independent of the NFL-NFLPA collective bargaining agreement. Nor can the NFL rely on Stringer, as the plaintiffs here do not reference or implicate the duties of team trainers. While the plaintiffs’ claims may be “related” to the CBA, resolving them would not be “dependent on” or “intertwined with” the bargaining agreement’s interpretation, as Judge Perry recently found in Green v. Arizona Cardinals.

Proving causation will be more difficult. Current and future plaintiffs should be able to overcome potential defenses of assumption of risk and contributory negligence. While athletes knew they risked concussions, it was nearly impossible for them to fully appreciate this risk, given the NFL’s obstinacy in failing to take such injuries seriously and fighting the increasing weight of science in order to preserve its own image. While some players may have hid (and still hide) concussions from their teams, the desire to do so was borne largely from the tough-it-out culture that the NFL deliberately crafted. By successfully arguing for an expansion of the loss of chance doctrine, NFL concussion litigation plaintiffs may find a pathway to adequate recovery for the harm allegedly caused by the NFL’s negligence and seemingly active spread of misinformation, despite multiple factors potentially contributing to long-term cognitive illness and injury. As Pro Football Hall of Famer Frank Gifford stated in 1960, “Pro football is like nuclear warfare. There are no winners, only survivors.”

Unfortunately, for many of those survivors suffering from devastating illnesses, such as CTE, ALS, Alzheimer’s, dementia, and Parkinson’s Disease, the war continues.

---

INTERSEX CHILDREN IN FOSTER CARE: CAN THE GOVERNMENT ELECT SEX ASSIGNMENT SURGERY?

Ashley Huddleston*

The surgeon told [me] he was going to “fix [me].” But I didn’t know I was broken . . . . [I] knew I was different but not that I was broken.¹

INTRODUCTION

In December 2004, M.C. was born in a South Carolina hospital and pronounced a male.² However, doctors soon discovered that he had “ambiguous genitals and both male and female internal reproductive structures.”³ M.C. had been born “intersex.”⁴ For the first few years of M.C.’s life,⁵ the doctors

* J.D. Candidate, Brooklyn Law School, 2015; B.A., University of Rochester, 2010. Thank you to my friends and family for their endless and unwavering encouragement, patience, and support. Thank you also to the editors and staff of the Journal of Law and Policy for their time and invaluable feedback.

¹ Martha Coventry, Making the Cut, Ms., Oct. 2000, at 52, 57.
³ M.C. Complaint, supra note 2, ¶ 3.
⁴ As discussed later, there is not one universally accepted definition for exactly what constitutes an “intersex” condition, but this Note adopts the definition used by the Intersex Society of North America, which they say is “a
were unsure of his sex\textsuperscript{6} and repeatedly indicated that he could be raised as either a male or a female.\textsuperscript{7} Though the doctors had no way of knowing which gender M.C. would identify with as an adult,\textsuperscript{8} they “decided to remove M.C.’s healthy genital tissue and radically restructure his reproductive organs in order to make his body appear to be female.”\textsuperscript{9} But it was not M.C.’s parents who elected surgery for him; it was the doctors and the South Carolina Department of Social Services, whose custody he was in at all relevant times.\textsuperscript{10} In 2006, a couple learned about M.C.’s

\textsuperscript{5} &ldquo;Intersex&rdquo; is a general term used for a variety of conditions in which a person is born with a reproductive or sexual anatomy that doesn’t seem to fit the typical definitions of female or male.” Intersex Society of North America, \textit{What is Intersex?}, http://www.isna.org/faq/what_is_intersex (last visited Apr. 2, 2014). M.C. was in fact born with the condition of ovotesticular DSD, which was formerly referred to as “true hermaphroditism.” M.C. Complaint, \textit{supra} note 2, ¶ 40.

\textsuperscript{6} Theorists argue that traditionally “sex” was considered biologically determined at birth, and “gender” was generally understood as the “sociocultural manifestation of one’s sex.” However, some theorists now believe that the two are more intertwined and reliant on one another for their proper meaning. They now theorize that sex is not objective, but rather has a cultural component, and gender is created by factors other than just “being an outgrowth of sex.” See Laura Hermer, \textit{Paradigms Revised: Intersex Children, Bioethics and the Law}, 11 ANNALS HEALTH L. 195, 200–01 (2002). The medical profession, through the theories and treatments now associated with intersex children, has moved away from the historically accepted principle of “true sex” based strictly on gonadal tissue. Instead, it has focused on the gender that they believe should be assigned to the child and match the sex to that gender. Alice Domurat Dreger, “Ambiguous Sex”— or Ambivalent Medicine? Ethical Issues in the Treatment of Intersexuality, HASTINGS CTR. REP., May–Jun. 1998, at 24, 26–27. Case management of intersex infants reflects that the physicians are concerned with “perpetuating the notion that good medical decisions are based on interpretations of the infant’s real ‘sex’ rather than on cultural understandings of gender.” Suzanne J. Kessler, \textit{The Medical Construction of Gender: Case Management of Intersexed Infants}, 16 SIGNS 3, 10 (1990). For the purposes of clarity, this Note will use the terms in their traditional sense: “sex” as the biological determination and “gender” as the socialization.

\textsuperscript{7} M.C. Complaint, \textit{supra} note 2, ¶¶ 3, 42

\textsuperscript{8} Id. ¶ 4.

\textsuperscript{9} Id.

\textsuperscript{10} Id. ¶¶ 34–39. Though M.C. lived with two foster families before his
condition from the South Carolina adoption website and contacted the Department in an attempt to prevent the sex assignment surgery, but they were too late; they adopted M.C. after the surgery was already completed. Now eight years old, M.C. has psychologically rejected his female gender assignment and is living as a boy. But nothing can replace the permanent changes made to his body.

Remarkably, such sex-change surgery is not a new procedure, and hundreds of children have been subjected to the same fate as M.C., usually at the mercy of their parents. Controversy over the surgery itself has been rampant for years as scholars and physicians question the foundational theory that an intersex child will always identify with whatever gender they are surgically assigned to, and whether parents can and should elect sex assignment surgery at all. This Note will explore what has happened, and what should happen, when an intersex child is in the custody of the state, just as M.C. was.

The law is silent on whether or not government officials have the ability to consent to sex assignment surgery; but if they do not, then the question becomes what those officials should do when confronted with that situation. The answer is that social workers and government officials must choose to do nothing—

adoption, South Carolina Department of Social Services “retained legal custody of M.C. while he was in foster care.” Id. ¶ 39.

11 Id. ¶ 64.
12 Id. ¶¶ 2, 7, 8.
13 Id. ¶¶ 8, 9, 11.
they should let the child develop as he or she naturally would—and if the parents or the child choose at a later time to undergo surgery, that is their own choice to make.\(^\text{17}\) It is not appropriate, however, for a case worker to make such a major life decision for a child who is only temporarily in the government’s custody, especially when that child will have to live with the results of that surgery for the rest of his or her life.

Part I discusses the background of intersexuality\(^\text{18}\) and the theories that have led to the current clinical treatment of intersex children in the United States.\(^\text{19}\) In particular, it will explore Dr. John Money’s development of the fundamental theory\(^\text{20}\) that any intersex child can successfully be assigned to either gender as long as his or her external genitals are surgically made to match that assignment.\(^\text{21}\) In addition, Part I articulates the procedures that physicians follow in assessing a child that may have an intersex condition, and addresses the factors that physicians take into account when deciding how to assign a child to a particular sex.\(^\text{22}\)

Part II discusses the lawsuits recently filed in both federal court and South Carolina state court by M.C.’s adoptive parents

\(^{17}\) I do not suggest that adoptive parents either do or do not have the ability to seek surgery for their children if they adopt them at a young age, as there is already wide debate on whether or not biological parents should have that right. I simply suggest that surgery may be appropriate or desirable at a later stage in the child’s life.

\(^{18}\) See AM. PSYCHOLOGICAL ASS’N, ANSWERS TO YOUR QUESTIONS ABOUT INDIVIDUALS WITH INTERSEX CONDITIONS (2006), available at http://www.apa.org/topics/sexuality/intersex.pdf (“A variety of conditions that lead to atypical development of physical sex characteristics are collectively referred to as intersex conditions. These conditions can involve abnormalities of the external genitals, internal reproductive organs, sex chromosomes, or sex-related hormones.”); sources cited supra note 6.

\(^{19}\) See infra Part I.


\(^{21}\) See generally MONEY & EHRRHARDT, supra note 20; JOHN MONEY, SEX ERRORS OF THE BODY: DILEMNAS, EDUCATION, COUNSELING 45 (1968). See also PREVES, supra note 14, at 3; Hermer, supra note 6, at 196–98.

\(^{22}\) See infra Part I.
on his behalf against the individual doctors and the employees of the South Carolina Department of Social Services who approved his sex assignment surgery. Part III discusses the state laws governing health care consent in the foster care system and how those laws impact the government’s ability to elect this kind of surgery for a child in their care. Finally, Part IV discusses new policies that other countries have adopted to deal with the rising concerns over the surgical practice on intersex children, and what measures the United States can take when an intersex child is born into the care of a state Department of Social Services.

I. MEDICINE’S RESPONSE TO INTERSEXUALITY

The strict division between female and male bodies and behavior is our most cherished and comforting truth. Mess with that bedrock belief, and the ground beneath our feet starts to tremble.

A. Background on Intersexuality

It is estimated that “[b]etween 1.7% and 4% of the world population is born with intersex conditions.” This means that the condition “occurs about as often as the well-known conditions of cystic fibrosis and Down syndrome.” The number of people born with an intersex condition is much higher than the public may be aware of, since it has only started receiving more attention as people come forward with their stories.

“[I]n medical terms the definition of intersex genitalia is

---

23 See infra Part II.
24 See infra Part III.
25 See infra Part IV.
26 Coventry, supra note 1, at 55.
27 Kate Haas, Who Will Make Room for the Intersexed?, 30 AM. J.L. & MED. 41, 41 (2004); Melanie Blackless et. al., How Sexually Dimorphic Are We? Review and Synthesis, 12 AM. J. HUM. BIOLOGY 151, 159 (2000) (estimating that 1.728% of live births result in individuals that do not categorically fit into male or female).
28 PREVES, supra note 14, at 3.
29 Id.
somewhat arbitrary” because there is not one medically standard measurement or criterion that determines the sex of a child. In the broadest sense, “intersexuality constitutes a range of anatomical conditions in which an individual’s anatomy mixes key masculine anatomy with key feminine anatomy.” The “[i]ntersex conditions are myriad in number and type; virtually all develop in utero” when “the fetus is exposed to an inappropriate amount of hormones. . . .” While the American Academy of Pediatrics (“AAP”) outlines seven clinical findings that raise the possibility of intersexuality, most cases of ambiguous genitalia are not medically considered cases of “true intersex” conditions. However, for the purposes of this Note,

30 Ursula Kuhnle & Wolfgang Krahl, The Impact of Culture on Sex Assignment and Gender Development in Intersex Patients, 45 PERSP. IN BIOLOGY & MED. 85, 87 (2002).
31 Physicians have different measurements for when the penis is considered inadequate or the clitoris too large. Compare Committee on Genetics, Section on Endocrinology and Section on Urology, Evaluation of the Newborn with Developmental Anomalies of the External Genitalia, 106 PEDIATRICS 138, 139 (2000) [hereinafter Evaluation of the Newborn] (“In full-term newborns the stretched penile length should measure at least 2 [centimeters].”) with SUZANNE KESSLER, LESSONS FROM THE INTERSEXED 43 (1998) (measuring a “medically acceptable” clitoris as up to one centimeter, and a “medically acceptable” penis as between 2.5 and 4.5 centimeters); Alice D. Dreger, supra note 6, at 28 (noting that a clitoris is considered too big if it is larger than one centimeter, and a penis is too small if it is less than 2.5 centimeters).
32 Dreger, supra note 6, at 26.
33 Hermer, supra note 6, at 204.
35 The clinical findings that raise the possibility of intersexuality are as follows: Apparent male: “Bilateral nonpalpable testes in a full-term infant; hypospadias associated with separation of the scrotal sacs; undescended testis with hypospadias.” Ambiguous genitalia signify an indeterminate sex. Apparent female: “Clitoral hypertrophy; Foreshortened vulva with single opening; Inguinal hernia containing a gonad.” Evaluation of the Newborn, supra note 31, at 139.
36 “True intersexed” conditions refer to children who have both ovarian and testicular tissue in one or both of their gonads, but these cases represent less than five percent of those with ambiguous genitalia. Modern literature
all such conditions will be referred to as “intersex” \(^{37}\) and will use the definition provided by the Intersex Society of North America, which defines intersexuality as a condition where “a person is born with a reproductive or sexual anatomy that [does not] seem to fit the typical definitions of female or male.” \(^{38}\)

The idea of a “normal” male or female body is in fact a man-made concept created by categories that society has arbitrarily and superficially defined. \(^{39}\) Physical gender identification does not seem to exist as a strict male/female dichotomy, but rather as more of a continuum. \(^{40}\) Modern societal standards have created two categories that bodies must fit into, \(^{41}\) but “biologically speaking, there are many gradations running from female to male.” \(^{42}\) In fact, the development of surgical methods to “normalize” sexual variation evidences “the regularity with which sexual variation occurs.” \(^{43}\) However, even doctors with ample experience in the field of pediatric intersexuality “hold an incorrigible belief in and insistence upon female and male as the only ‘natural’ options.” \(^{44}\)

refers to those that have either testes or ovaries as “intersex” as well. Kessler, supra note 6, at 5. See also Dreger, supra note 6, at 30 (“[W]hile unusual genitalia may signal a present or potential threat to health, in themselves they just look different.” (emphasis in original)).

\(^{37}\) Children with ambiguous genitalia are candidates for sex assignment surgery, whether or not their condition fits that of technically being intersex. The number of children subjected to these surgeries would be grossly underestimated if those with ambiguous genitalia were excluded. Sara R. Benson, Hacking the Gender Binary Myth: Recognizing Fundamental Rights for the Intersexed, 12 CARDOZO J.L. & GENDER 31, 33 (2005).


\(^{39}\) PREVES, supra note 14, at 3

\(^{40}\) See Blackless et. al., supra note 27, at 162–63 (arguing that sexual distribution is properly represented as an overlapping bell curve instead of two totally separate genders that fail to ever overlap).

\(^{41}\) See ANNE FAUSTO-STERLING, SEXING THE BODY: GENDER POLITICS AND THE CONSTRUCTION OF SEXUALITY 108–09 (2000) (discussing other cultures that have recognized a third gender).


\(^{43}\) PREVES, supra note 14, at 3.

\(^{44}\) Kessler, supra note 6, at 4.
Presupposing the existence of only those two gender options, doctors are supposed to inspect any child born with potentially ambiguous genitalia before definitively pronouncing a sex.\textsuperscript{45} Though the AAP characterizes intersexuality as a “social emergency”\textsuperscript{46} and not a true medical emergency,\textsuperscript{47} such a characterization belies the response from physicians, who often proceed as quickly as possible to definitively assign a sex.\textsuperscript{48} Although some of the conditions that cause intersexuality can be life-threatening, being intersex is, by itself, not life threatening.\textsuperscript{49} “‘Ambiguous’ genitalia do not constitute a disease. They simply constitute a failure to fit a particular (and, at present, a particularly demanding) definition of normality.”\textsuperscript{50} Regardless, doctors and parents sometimes rush into surgery, as was the case for M.C., as if intersexuality is a medical emergency that must be remedied.

\textbf{B. Doctor John Money’s Theories and the John/Joan Case}

Most of the contemporary theory that guides the treatment of intersex children arose from the work of sexologist Dr. John Money in the 1950s.\textsuperscript{51} When Dr. Money published his allegedly

\textsuperscript{45} \textit{Evaluation of the Newborn}, supra note 31, at 138.

\textsuperscript{46} \textit{Id.}

\textsuperscript{47} There do not appear to be any other conditions that the AAP classifies as a “social emergency.”

\textsuperscript{48} \textit{See} Dreger, supra note 6, at 27 (“In the United States today . . . typically upon the identification of an ‘ambiguous’ or intersexed baby teams of specialists . . . are immediately assembled, and these teams of doctors decide to which sex/gender a given child will be assigned.”); Kessler, \textit{supra} note 6, at 8 (“The doctors interviewed concur with the argument that gender be assigned immediately, decisively, and irreversibly . . . ”).

\textsuperscript{49} For example, those children “whose condition is caused by androgen insensitivity are in danger of malignant degeneration of the tests unless they are removed.” Kessler, \textit{supra} note 6, at 5 n.6.

\textsuperscript{50} Dreger, \textit{supra} note 6, at 30.

\textsuperscript{51} \textit{See} KESSLER, \textit{supra} note 31, at 6 (“Virtually all academic writing on sex and gender refers to a case first described by sexologist John Money in 1972.”); PREVES, \textit{supra} note 14, at 53 (“John Money’s theory of gender identity development and suggested standards of care are at the center of late-twentieth-century debates on how to best respond to intersex.”); Alice
successful case study of John, a boy with a damaged penis who was surgically made into and raised as a girl, “the treatment’s purported success spread rapidly and [was] frequently recounted in the professional literature.”52 The view that children are born “psychosexually neutral and would accept their gender of rearing . . . offered a relatively simple solution to what was seen as a difficult situation.”53 One single case, widely reported as a success, “became the justification for surgical treatment of intersex infants.”54

Dr. Money developed his theory primarily from the case study that has come to be known as the John/Joan case.55 John66 was born an XY male77 with an identical twin brother. He

---

52 Beh & Diamond, supra note 15, at 17.
53 Id. at 18.
56 John is a pseudonym that has been used in all literature written about this case, as the person did not want to reveal his real name. See Colapinto, supra note 55, at 54–97; Diamond & Sigmundson, supra note 55, at 298–304. He revealed his true identity as David Reimer after a book published by John Colapinto in 2000 gave him a chance to tell his story. See David Reimer, 38, Subject of the John/Joan Case, Dies, N.Y. TIMES (May 12, 2004), http://www.nytimes.com/2004/05/12/us/david-reimer-38-subject-of-the-john-joan-case.html. For the purposes of this Note, I will use John/Joan.
57 Genetic males are typically born with XY sex chromosomes, while females are born with XX sex chromosomes. Y CHROMOSOME, GENETICS HOME REFERENCE, http://ghr.nlm.nih.gov/chromosome/Y (last updated
suffered a burn to his penis during a circumcision when he was eight months old,\textsuperscript{58} causing it to essentially break away into pieces until nothing remained.\textsuperscript{59} The doctors told his mother that reconstructive surgery would fail to give him a normal-looking penis,\textsuperscript{60} and the urologist wrote in the medical record that “restoration of the penis as a functional organ is out of the question.”\textsuperscript{61} A psychologist’s opinion of the situation was even less encouraging for John’s parents, saying that “[John] will be unable to consummate marriage or have normal heterosexual relations; he will have to recognize that he is incomplete, physically defective, and that he must live apart.”\textsuperscript{62}

While searching for what to do next, John’s parents saw Dr. John Money on television talking about gender transformation and a clinic at Johns Hopkins University that was performing sex changes.\textsuperscript{63} Shortly thereafter, John’s parents took him to see Dr. Money at Johns Hopkins, where Dr. Money explained the advantages of sex reassignment for John and told them that he saw no reason why it would not work.\textsuperscript{64} Dr. Money later wrote about this first meeting with the family and recalled that he used non-medical terms and photos to explain everything to them.\textsuperscript{65} But in discussing the meeting years later, John’s parents said that they were caught up in the confidence that Dr. Money exuded and could not appreciate until much later the fact that this procedure had only been performed on “hermaphrodites”;\textsuperscript{66} not

---

\textsuperscript{58} \textsc{Colapinto, supra} note 20, at 12–15; \textsc{Colapinto, supra} note 55.
\textsuperscript{59} \textsc{Colapinto, supra} note 20, at 15 (“Over the next few days, baby [John’s] penis dried and broke away in pieces. It was not very long before all vestiges of the organ were gone completely.”).
\textsuperscript{60} \textit{Id.}
\textsuperscript{61} \textsc{Colapinto, supra} note 55, at 58 (internal quotation marks omitted).
\textsuperscript{62} \textit{Id.} (emphasis added).
\textsuperscript{63} \textsc{Colapinto, supra} note 20, at 17–23.
\textsuperscript{64} \textit{Id.} at 49; \textsc{Colapinto, supra} note 55.
\textsuperscript{65} \textsc{Colapinto, supra} note 20, at 50.
\textsuperscript{66} \textit{See Is a Person Who is Intersex a Hermaphrodite?, INTERSEX SOC’Y OF NORTH AM., http://www.isna.org/faq/hermaphrodite (last visited Mar. 23, 2014) (explaining that the term “hermaphrodite” is the older terminology used for intersex conditions).}
on someone who had been born with normal genitalia.\textsuperscript{67}

John’s parents struggled with what to do, but worried about
the embarrassment that John would face without an adequate
penis, they decided to raise him as a girl.\textsuperscript{68} In 1967, John
underwent surgical castration. The surgeons removed his testicles
and constructed an exterior vagina.\textsuperscript{69} His parents were instructed
to call him by his new female name, Joan, and to treat him as an
ordinary girl without telling her\textsuperscript{70} of the surgery.\textsuperscript{71} The family
continued to return to Johns Hopkins after the surgery to meet
with Dr. Money for follow-up treatment and monitoring.\textsuperscript{72}

Meanwhile, medical literature published during this time
portrayed the sex reassignment procedure as a success, and such
literature “had a significant impact on the standard of care that
developed for certain intersex conditions . . . ”\textsuperscript{73} One account
read: “This dramatic case . . . provides strong support . . . that
conventional patterns of masculine and feminine behavior can be
altered. It also casts doubt on the theory that major sex
differences, psychological as well as anatomical, are immutably
set by the genes at conception.”\textsuperscript{74} Medical texts and social science
writings well into the 1990s continued to reflect the impact of the
purportedly successful case study.\textsuperscript{75}

\textsuperscript{67} \textit{Colapinto}, supra note 20, at 50.

\textsuperscript{68} \textit{Id.} at 52 (discussing the potential embarrassment, his father said, “You
know how little boys are . . . who can pee the furthest? Whip out the wiener
and whiz against the fence. Bruce wouldn’t be able to do that, and the other
kids would wonder why.” (internal quotation marks omitted)).

\textsuperscript{69} \textit{Id.} at 53–54.

\textsuperscript{70} “Her”, “she”, or “Joan” will be used to describe the time period that
John was treated as a girl by his parents and by the clinicians. “Him” or “he”
will be used to discuss John at all other times.

\textsuperscript{71} \textit{Colapinto}, supra note 55 at 64.

\textsuperscript{72} \textit{Id.} at 68.

\textsuperscript{73} \textit{Beh \\& Diamond}, supra note 15, at 4.

\textsuperscript{74} \textit{Diamond \\& Sigmundson}, supra note 55, at 299 (citation omitted).

\textsuperscript{75} \textit{Id.} (“The following quote is typical: ‘The choice of gender should be
based on the infant’s anatomy, not the chromosomal karyotype. Often it is
wiser to rear a genetic male as a female. It is relatively easy to create a vagina
if one is absent, but it is not possible to create a really satisfactory penis if the
phallus is absent or rudimentary. Only those males with a phallus of adequate
size that will respond to testosterone at adolescence should be considered for
Based on the John/Joan case study, Dr. Money developed his theory regarding the malleability of sex and gender identification. “Money’s theory holds that (1) all children, intersexed and non-intersexed, are psychosexually neutral at birth, and (2) you can therefore make virtually any child either gender as long as you make the sexual anatomy reasonably believable.” He believed that “children differentiate a gender role and identity by way of complementation to members of the opposite sex, and identification with members of the same sex.” In addition, he thought it was crucial for a child to define the difference between males and females primarily by one’s sex organs and for that child to have confidence in how to identify his or her own sex organs. Thus, “the boundaries of the masculine and feminine gender roles” needed to be clearly defined. He also believed that a child’s “gender identity” could be molded and changed until the age of eighteen months, the time at which children are cognizant enough to differentiate between sexes and may have trouble adopting a newly assigned sex.

Dr. Money claimed that children born intersex should definitely be assigned to one gender:

[T]he experts must insure that the parents have no doubt about whether their child is male or female [that is, the parents must fully believe that the child will identify with the gender assigned]; the genitals must be made to match the assigned gender as soon as possible; gender-appropriate hormones must be assigned at puberty; and intersexed children must be kept informed about their situation with age-appropriate explanations.
He claimed that unambiguous genitalia is necessary for an intersex child to identify with a gender; that an intersex child who does not have surgery and develops either an ambiguous gender identity, or rejects the one assigned, does so in response to the ambiguous sex organs.\footnote{See \textit{Money} \& \textit{Ehrhardt}, supra note 20, at 19.} Dr. Money believed that the sex organ is not incidental to how a person internally identifies himself/herself, but is at the very definition of one’s gender.

What nobody reported until 1997, years after Dr. Money had developed his theories, was that Joan had in fact rejected her female assignment very early on and had been living as a male since 1979, when he was 14 years old.\footnote{See generally \textit{Colapinto}, supra note 20, at 111–80; \textit{Diamond} \& \textit{Sigmundson}, supra note 55, at 298–304.} Two doctors conducted a follow-up interview with John in 1994 and 1995, and John exposed the psychological anguish that he experienced as a child, his feelings of knowing that he did not feel like a girl, and his ultimate decision to forego living as a female.\footnote{See generally \textit{Diamond} \& \textit{Sigmundson}, supra note 55.} The follow-up paper not only fueled debate among medical professionals about the convention of performing sex assignment surgery, but also “raised troubling questions about why the case was reported in the first place, why it took almost 20 year for a follow-up to reveal the actual outcome and why that follow-up was conducted not by Dr. Money but by outside researchers.”\footnote{Colapinto, supra note 55. At 56}

In relaying the story, John’s mother recalls that as an infant, Joan did not want to stay in her female clothing, that she always preferred boy’s games and toys,\footnote{\textit{Diamond} \& \textit{Sigmundson}, supra note 55, at 299.} and that she was ridiculed by students in school as early as kindergarten because she was a tomboy.\footnote{Colapinto, supra note 55. At 66.} As per Dr. Money’s instruction, the entire family returned to Johns Hopkins for follow-up visits,\footnote{\textit{Id.} at 68.} but Joan began refusing to go because of her “discomfort and embarrassment with forced exposure of her genitals and [the] constant attempts, particularly after the age of 8 years, to convince her to behave
more like a girl and accept further vaginal repair.” She dreaded going to visit Dr. Money, and she continually refused to undergo any vaginal surgery or take the hormones that he insisted she take to further her change into a female. Finally, “[a]fter age 14 years, Joan adamantly refused to return to the hospital [and] came fully under the care of local clinicians.” Joan did not fully realize she was not a girl until between age 9 and 11. John relates:

There were little things from early on. I began to see how different I felt I was, from what I was supposed to be. But I didn’t know what it meant. I thought I was a freak or something . . . I figured I was a guy but I didn’t want to admit it. I figured I didn’t want to wind up opening a can of worms. She regularly saw psychologists and physicians, but at the age of 14, she had had enough, and told her doctors that she did not want to be a girl. It was only then that her father told her the truth of what happened. Joan recalls: “All of a sudden everything clicked. For the first time things made sense and I understood who and what I was.” Joan affirmatively decided to live as a male starting at age 14, had surgeries to construct a phallus at age 15 and 16, and underwent hormone therapy to facilitate living as a male. Joan then became John.

Unfortunately, John tragically committed suicide in 2004 at the age of 38. John Colapinto, who chronicled John/Joan’s story, revealed that there were many things that contributed to his suicide: financial and marital problems, unemployment, genetics,

---

91 Diamond & Sigmundson, supra note 55, at 300.
92 Colapinto, supra note 55 at 70.
93 Diamond & Sigmundson, supra note 55, at 300.
94 Id. at 299–300 (internal quotation marks omitted).
95 COLAPINTO, supra note 20, at 178–79.
96 Diamond & Sigmundson, supra note 55, at 300 (internal quotation marks omitted).
97 Id.
98 David Reimer, 38, Subject of the John/Joan Case, Dies, supra note 56. His mother reported that she thinks “he felt he had no options. It just kept building up and building up.” Id.
and his childhood.\textsuperscript{99} Colapinto recounts that John told him “that he could never forget his nightmare childhood, and he sometimes hinted that he was living on borrowed time.”\textsuperscript{100} However, John did not blame his parents, instead saying, “[they] feel very guilty, as if the whole thing was their fault . . . . But it wasn’t like that. They did what they did out of \textit{kindness} and love and desperation. When you’re desperate, you don’t necessarily do all the right things.”\textsuperscript{101}

The truth about the results of John’s purported successful surgery did not enter professional medical literature until more than twenty years after the procedure, leaving nothing to “counter the positive reports of the case nor impact the standard of care as it had developed since the 1960s . . . .”\textsuperscript{102} Dr. Money did not report in his literature some of the signs that Joan was possibly rejecting her female assignment, and when Joan finally refused to return to Johns Hopkins, he wrote that she was “lost to follow-up.”\textsuperscript{103} The reality was that John was unwilling to continue on Dr. Money’s course of treatment, but that is not what he relayed to the public.\textsuperscript{104} It is impossible to know whether the course of treatment for intersex infants would have developed differently had the truth been reported earlier, but it is unquestionable that Dr. Money’s theories were critical in shaping the current medical practice.\textsuperscript{105}

\textbf{C. The Current Course of Intersex Treatment}

Without knowing that Dr. Money’s experiment was in reality a failure, other doctors and researchers developed medical practices in accordance with his allegedly successful theory. The result being that “[o]ver the past four decades, early surgical

\begin{itemize}
\item \textsuperscript{100} Id.
\item \textsuperscript{101} COLAPINTO, \textit{supra} note 20, at xvii (emphasis in original).
\item \textsuperscript{102} Beh & Diamond, \textit{supra} note 15, at 12.
\item \textsuperscript{103} Id. at 8–9 (citation omitted).
\item \textsuperscript{104} Id. at 9.
\item \textsuperscript{105} Id. at 19 n.81 (citation omitted).
\end{itemize}
intervention for infants who are born with ambiguous genitalia or who suffer traumatic genital injury often has been recommended as standard procedure.”106 The widespread adoption of Dr. Money’s theory has resulted in few subsequent studies “evaluating the sexual and psychological success or failure of sex assignment surgeries, even though such surgeries have been performed long enough for a substantial cohort to have reached adulthood.”107 The few studies that have been done suggest that the procedures “cause[] substantial and unreasonable harm to infant subjects.”108 Two of the larger published studies in fact provide strong evidence that many intersex individuals fail to identify with their surgically assigned gender.109 These two studies are not enough, though, given how widely circulated and deeply engrained Dr. Money’s theories are. In addition, people who underwent surgery as children have provided anecdotal evidence demonstrating widespread physical and mental dissatisfaction.110 There is nothing to indicate that intersexed adults who did not undergo surgery fare any worse than those with did have sex assignment surgery.111 It is clear that there is a pressing need for additional follow-up studies and data collection of those intersex individuals who were subjected to surgery at an early age.

Even with increased debate and discussion about sex assignment surgery for intersex infants,112 surgery still remains the standard practice:113 While the AAP notes that some have

106 Id. at 3 (citation omitted).
107 Hermer, supra note 6, at 212.
108 Ford, supra note 51, at 474.
109 See Sarah M. Creighton et al., Objective Cosmetic and Anatomical Outcomes at Adolescence of Feminising Surgery for Ambiguous Genitalia Done in Childhood, 358 LANCET 124, 125 (2001) (documenting the outcome of “a retrospective study of cosmetic and anatomical outcomes in 44 adolescent patients who had ambiguous genitalia in childhood and underwent feminising genital surgery”); Hermer, supra note 6, at 212–13 (noting that many of the individuals in one of the largest studies experienced psychological and identity problems).
110 See Navarro, supra note 16.
111 Haas, supra note 27, at 48.
112 Beh & Diamond, supra note 15, at 3.
113 Id. at 18.
suggested “the current early surgical treatment be abandoned in favor of allowing the affected person to participate in gender assignment at a later time[.]” The AAP itself still urges “a definitive diagnosis be determined as quickly as possible.” The AAP acknowledges that some people might “have conflicts between their psychosexual orientation and their genital appearance and function, [but that] the principles [it] outlined [for deciding how to determine which sex assignment is proper] . . . should minimize these problems when conducted by an appropriately constituted intersex team.” Effectively, the AAP has admitted that it is possible that a child could reject his or her assigned sex, but it still advocates for such surgical procedures.

The AAP guidelines, produced in 2000 for physicians treating intersex children, outline a number of factors that doctors need to consider when deciding which sex the child should appropriately be assigned to. The factors include fertility potential, capacity for normal sexual function, endocrine function, malignant change, testosterone imprinting, and timing of surgery. However, “[the AAP] still appear[s] to cling to the idea that the size of a boy’s penis should be the deciding factor regarding how to raise a child.” Doctors may delay any surgery for an XY male with an underdeveloped penis until there is opportunity to see if the child’s body responds to hormone treatment, but “[i]f at the end of the treatment period the phallic tissue has not responded, what has been a potential penis . . . is now considered an enlarged clitoris . . . and reconstructive surgery is planned as for the genetic female.”

Doctors seem to be most concerned with the size and outer physical appearance of the penis rather than its

114 Evaluation of the Newborn, supra note 31, at 141.
115 Id. at 138.
116 Id. at 141.
117 Id. The exact medical effect that each factor has on the treatment and outcome of intersex children is beyond the scope of this Note. For more extensive discussion on this topic, see generally Gender and Genetics: Genetic Components of Sex and Gender, WHO, http://www.who.int/genomics/gender/en/index1.html (last visited Mar. 23, 2014).
118 Benson, supra note 37, at 35.
119 Kessler, supra note 6, at 11–12.
120 Id. at 13.
functionality, reflecting adherence to Dr. Money’s approach that “chromosomes are less relevant in determining gender than penis size, and that, by implication, ‘male’ is defined not by the genetic condition of having one Y and one X chromosome . . . but by the aesthetic condition of having an appropriately sized penis.”

Under this analysis, if a genetic male’s penis is “determined to be ‘inadequate’ for successful adjustment as [a male, he is] assigned the female gender and reconstructed to look female.”

“Meanwhile, genetic females (that is, babies lacking a Y chromosome) born with ambiguous genitalia are declared girls—no matter how masculine their genitalia look.” Additionally, “surgeons seem to demand far more for a penis to count as ‘successful’ than for a vagina to count as such[.]” so the odds of having “successful” surgery seem to increase if the surgeons construct a vagina than if they take on the task of constructing a penis. The effect has been that “more intersex infants are being assigned to the female sex” than the male sex; however, the number of individuals who actually accept their assigned gender is unclear because there have been very few long-term follow-up studies of intersex children.

121 Id. at 12.
122 Dreger, supra note 6, at 28.
123 Id.
124 Id. at 29 (“For a penis to count as acceptable functional’—it must be or have the potential to be big enough to be readily recognizable as a ‘real’ penis. In addition, the ‘functional’ penis is generally expected to have the capability to become erect and flaccid at appropriate times, and to act as the conduit through which urine and semen are expelled, also at appropriate times . . . . [T]ypically, surgeons also hope to see penises that are ‘believably’ shaped and colored. Meanwhile, very little is needed for a surgically constructed vagina to count among surgeons as ‘functional.’ For a constructed vagina to be considered acceptable by surgeons specializing in intersexuality, it basically just has to be a hole big enough to fit a typical-sized penis. It is not required to be self-lubricating or even to be at all sensitive . . . . [A]ll that is required is a receptive hole.”).
125 Kuhnle & Krahl, supra note 30, at 89. See also Kessler, supra note 6, at 13 (“[A]s long as the decision rests largely on the criterion of genital appearance, and make is defined as having a “good-sized” penis, more infants will be assigned to the female gender than to the male.”).
126 Hermer, supra note 6, at 212.
As soon as doctors suspect that a child may be intersex, they work rapidly to make a definitive diagnosis and proceed to surgery. The AAP guidelines indicate that laboratory and imaging studies should be done when a child has ambiguous genitalia in order to confirm a diagnosis, but doctors indicate that they “feel an urgent need to provide an immediate assignment and genitals that look and function appropriately.” One physician specializing in the area of intersex conditions said, “We can’t do [the diagnosis] early enough . . . . Very frequently a decision is made before all this information is available, simply because it takes so long to make the correct diagnosis . . . . There’s a lot of pressure on parents [for a decision] and the parents transmit that pressure onto physicians.”

Another endocrinologist acknowledged that a family who was waiting to see if the infant’s phallus would grow with hormone treatment was so impatient that they “could only wait a month” before making a definitive decision about the child’s gender. Physicians’ language regarding intersex children in their care suggests that though they try to speak neutrally about the child, it is difficult to think of and speak of a child as one whose gender has yet to be determined. The hurry and impatience stem from the parents’ desire to have a “normal” child that they can unequivocally call a boy or girl, which only fuels the speed with which doctors proceed to surgery.

In these situations, physicians have the dual responsibility of evaluating the child’s condition and “also [managing] the parents’

---

127 Evaluation of the Newborn, supra note 31, at 139–40.
128 Kessler, supra note 6, at 13.
129 Id.
131 Kessler, supra note 6, at 13.
132 See KESSLER, supra note 31, at 19 (“[Doctors’] language suggests that it is difficult for them to take a completely neutral position and to think and speak only of a phallic tissue that belongs to an infant whose gender has not yet been determined or decided.”).
133 Kessler, supra note 6, at 13.
uncertainty about a genderless child.” Dr. Money claimed that “the best procedure of sex education and counseling is one of not creating emotional indigestion by saying too much, too soon, and also of not allowing emotional malnutrition by saying too little, too late.” Doctors encourage parents not to feel compelled to disclose their child’s sex to other people, but instead, to tell others that a problem is being resolved and that they would prefer not to get into the details of it. However, a physician interviewed indicated that he does not believe it is really possible for parents to think of their child as gender neutral. Physicians “respond to the parents’ pressure for a resolution of psychological discomfort” by using technology to make a child fit in to one of the two genders society defines as normal.

Doctors engage in a normalizing process with the child’s parents to educate them about their child’s intersex condition. “First, physicians teach parents about usual fetal development and explain that all fetuses have the potential to be male or female.” This description can be done with diagrams or pictures that show how a fetus develops and the point at which all fetuses start to differentiate into either male or female. Second, the doctors stress the other features of the child that are normal. Doctors aim to redirect the parents’ attention away from the problem and toward the “good things” about their child. Third, physicians use language to imply that it is the child’s genitals, not the child’s gender, that are ambiguous. Doctors use medical terminology instead of words like “hermaphrodite” to show that it is the child’s physiology that is unusual and “not that [the intersex

134 KESSLER, supra note 31, at 21.
135 MONEY, supra note 21, at 45.
136 KESSLER, supra note 31, at 21.
137 Id.
138 Kessler, supra note 6, at 25.
139 KESSLER, supra note 31, at 22.
140 Id.
141 Id.
142 Id.
143 Kessler, supra note 6, at 15–16.
144 KESSLER, supra note 31, at 22.
child] constitute[s] a category other than male or female.” 145 This language places emphasis on “the premise of the child’s having been born sexually unfinished.” 146 The situation is illustrated by using terms implying that “the trouble lies in the doctor’s ability to determine the gender, not in the baby’s gender per se.” 147 The doctors portray their work as a task of uncovering the child’s “true sex,” 148 instead of changing the child’s condition to conform to one of the two established sexes. Finally, the doctors stress the social factors that shape a child’s gender development; they detract attention from the biological factors even as they search for the biological cause of the intersexuality. 149 The doctors stress that “gender is fluid” and “not a biological given,” 150 while noting that much of the child’s gender identification will “depend, ultimately, on how everybody treats [the] child and how [the] child is looking as a person.” 151 Thus, Dr. Money’s principles regarding social gender construction are still very much present and pervasive among physicians. 152

The fact that the physicians go through this “normalizing process” 153 with parents illustrates how society, and even the medical community, continues to cling to the idea that there are only two genders, despite “incontrovertible physical evidence that

---

145 FAUSTO-Sterling, supra note 41, at 50–51 (emphasis in original).
146 MONEY, supra note 21, at 46. In fact, Dr. Money wrote that “the concept of being sexually unfinished is invaluable” to the parents. Id. at 62.
147 KESSLER, supra note 31, at 23.
148 PREVES, supra note 14, at 55.
149 KESSLER, supra note 31, at 23. “[D]octors make decisions about gender on the basis of shared cultural values that are unstated, perhaps even unconscious, and therefore considered objective rather than subjective.” Kessler, supra note 6, at 18. The almost definitive role that penis size has in the determination of sex assignment shows doctors’ focus on only one physical characteristic, “one that is distinctly imbued with cultural meaning.” Id.
150 Kessler, supra note 6, at 17.
151 Id.
152 MONEY, supra note 21, at 48 (“First and foremost, [parents] need to know that gender identity and role are not preordained by genetic and intrauterine events alone, but that psychosexual differentiation is largely a postnatal process and highly responsive to social stimulation and experience.”).
153 Kessler, supra note 6, at 15.
[it] is not mandated by biology.”154 Sex assignment surgery is thought to offer children a more “normal” way of life, but it fails to account for the fact that “the child might one day have a different concept of ‘normal’ and want to choose a different course of treatment, or none at all.”155 The process that doctors use to diagnose and treat intersex children reflects their adherence to the two-gender system.

Once a doctor has suspicion that a child may be intersex, they proceed with a course of treatment plan. Generally, one physician has chief responsibility for an intersexed child’s case and acts as a liaison between the doctors and the parents,156 while a team of specialists work collaboratively to make a treatment plan for the child.157 The specialists involved usually include a “pediatrician, pediatric urologist,158 pediatric psychiatrist, [and] pediatric endocrinologist159 . . . .”160 It is also recommended that the parents consult with a child psychiatrist who can aid parents not only at the time of diagnosis, but also as the child grows and may have questions about his or her condition.161 It is important that

154 Id. at 25.
155 Beh & Diamond, supra note 15, at 57.
156 Kessler, supra note 31, at 27. Remarkably, the specialist in charge of the case can sometimes have an impact on the sex that the child is assigned to. Some doctors acknowledge that when there is a decision to be made, pediatric endocrinologists tend to choose making the child into a female, while urologists gravitate toward making the child into a male. Id. at 27–28.
158 “A urologist is a physician who is trained to evaluate the genitourinary tract, which includes the kidneys, urinary bladder and genital structures in men and women, and the prostate and testicles in men.” AM. UROLOGICAL ASSOC. FOUND., WHAT IS A UROLOGIST?, available at http://www.urologyhealth.org/_media/_pdf/whatisaurologist.pdf.
160 Reiner, supra note 157, at 1045.
161 Id.
“the team . . . form a bond with the parents, assisting them even if some members of the team disagree with the parents’ decision.”

Though doctors are the experts and guide the parents’ decision, it is ultimately the child’s parents who have the final word on how their child’s treatment should be carried out.

The parents of an intersex child are positioned uniquely in that they have the ability and the burden of choosing their child’s gender. Once the parents make a choice, the “physicians merely provide the right genitals to go along with that socialization.”

But as one scholar notes, “at normal births, when the infant’s genitals are unambiguous, the parents are not told that the child’s gender is ultimately up to socialization. In those cases, doctors do treat gender as a biological given.” Thus, as the AAP concedes, an intersex child presents a “social emergency,” not a biological one. In fact, most of the discussion surrounding surgical intervention is not about what is in the best social interest of the child, but what is in the best social interest for the family unit.

Doctors acknowledge the importance of family socialization by stating that “the family’s perceptions, expectations, and desires should be assessed and included in the decision regarding the sex of rearing.” The perceived societal pressures to be either a male or female, and the family’s desire to have a son or daughter, push both the parents and the doctors to make the decision to fit an intersex child into a specified category with surgery.

Both doctors and parents share the responsibility and blame in deciding to surgically alter an intersexed infant’s body. The doctors act in accordance with the developed medical course of treatment, which may be faulty in and of itself because it is premised on faulty and outdated principles of sex and gender, to correct something that may not necessarily need correcting.

---

162 Id.
163 See id.
164 Kessler, supra note 6, at 17.
165 Id.
166 See Evaluation of the Newborn, supra note 31, at 138.
167 PREVES, supra note 14, at 53.
168 Id. at 55 (citation omitted).
169 In general, intersex conditions are not medically harmful to the child’s
The family, on the other hand, acts in response to social pressure, which requires that a child fit into the standard “male” or “female” category. Parents seem to be generally concerned about the psychological well-being of their child and the future of both the individual child and the family unit. Though parents seem to make the wrong decision in electing sex assignment surgery at all for their child, they presumably also act out of love and a desire to create the best situation for the child. When a government agency has custody of a child, though, as they did in M.C.’s case, that family interest, sense of love and attachment, and deeply rooted concern for the future are absent.

II. THE CASE OF M.C. V. AARONSON

The case of M.C. v. Aaronson presents the question of whether or not the government has the ability to elect sex assignment surgery for an intersex child in his or her control. M.C., through his adoptive parents Pamela Crawford and John Mark Crawford, filed two lawsuits against all parties allegedly responsible for the sex assignment operation performed on him when he was an infant, including the South Carolina Department of Social Services (“SCDSS”), its employees, and individual physicians who treated him. The first complaint, filed in federal court, asserts two causes of action against the SCDSS and the physicians. First, that in approving and performing the surgery their actions violated M.C.’s “substantive due process rights to bodily integrity, privacy, procreation, and liberty.” This violation occurred because the removal of M.C.’s phallus and potential sterilization was not medically necessary, caused significant physical pain, imposed unreasonable risks of future physical and mental pain and suffering, and deprived M.C. of the opportunity to
make his own deeply intimate decisions about whether to undergo genital surgery, if any, when he reached maturity.\footnote{Id. \¶ 71.}

Second, the SCDSS’ and the physicians violated M.C.’s fourteenth amendment right to procedural due process because defendants chose to perform surgery on M.C. without requesting or initiating any hearing on the procedure.\footnote{Id. \¶¶ 80, 82, 84, 86–87.} The second complaint, filed in state court in the Court of Common Pleas in Richmond County, South Carolina, also asserts two causes of action: medical malpractice and gross negligence.\footnote{Crawford Complaint, supra note 171, \¶¶ 40, 49.} The facts presented in the federal complaint, which were adopted by the District Court in a subsequent order,\footnote{Order, supra note 2. As these facts have been accepted as true by the Court and there have been subsequent findings and orders on the federal complaint, all facts in this Note are drawn from the federal complaint. It should be noted, though, that both the federal and state complaints essentially allege the same facts.} demonstrate that because M.C. was in the custody and care of SCDSS at the time his surgery took place,\footnote{M.C. Complaint, supra note 2, \¶ 39.} “SCDSS officials made decisions whether to authorize medical treatment [of M.C.], including the sex assignment surgery . . . .”\footnote{Id. \¶ 71.}

M.C. was born to his biological parents in South Carolina in 2004\footnote{Id. \¶ 16.} with a type of intersex condition that is “characterized by the presence of both ovarian and testicular tissue.”\footnote{Id. \¶ 40. The specific disorder is called ovotesticular difference/disorder of sex development.} Due to other medical complications, M.C. remained in the hospital for two and a half months after his birth,\footnote{Id. \¶ 35. M.C. had complications from being born prematurely and acid reflux.} during which time “SCDSS began an investigation into possible neglect by M.C.’s biological parents.”\footnote{Id. \¶ 36.} He was released from the hospital into the care of his biological parents but a week later, his biological parents
“notified SCDSS that they wanted to relinquish their parental rights.”183 Pursuant to a court order, M.C. was placed in SCDSS’s custody on February 16, 2005 and “[t]he court terminated M.C.’s biological parents’ parental rights on September 9, 2006.”184 He was placed with two foster families prior to his adoption by the Crawfords in 2006,185 but “SCDSS retained legal custody of M.C. while he was in foster care,”186 during which time he received the sex assignment surgery.

M.C. was initially identified as a male at birth; his medical records “noted that his ‘phallus was rather large’... [and] [r]outine blood tests indicated that [his] testosterone levels were ‘extremely elevated.’”187 But he also had “male and female internal reproductive structures,” as well as a “small vaginal opening below a ‘significant’ phallus . . . .”188 In fact, M.C.’s medical records indicate that, for the first few months of his life, the doctors were unsure what sex he was, as he was sometimes referred to as a male and sometimes as a female.189 During a surgery to correct M.C.’s acid reflux condition, the doctor “performed exploratory surgery to inspect M.C.’s sex organs” and reported that he had “ambiguous” genitalia.190 Over the next year or so, the doctors working on M.C.’s case agreed “that there was no compelling biological reason to raise M.C. as either a male or female.”191 However, they also repeatedly emphasized that they had the ability to make M.C. into either sex with surgery and that, based on whichever surgery was performed, M.C. could be raised as either a male or a female.192 For example, one doctor stated, “Due to the nature of M.C.’s external genital anatomy, either sex of rearing is possible with
appropriate surgery."\textsuperscript{193} Later, this same doctor stated that "[c]urrently [M.C.] could be potentially raised, surgically reconstructed, and treated to be male or female."\textsuperscript{194} Though one doctor did consider the possibility that there would be problems with assigning M.C. to the female gender,\textsuperscript{195} the team ultimately urged the SCDSS officials to allow them to perform sex reassignment surgery on M.C. in order to make his body look like that of a female.\textsuperscript{196} It is important to note that M.C.'s condition at this point did not present any negative physical side effects, and the surgery was not necessary for this physical health.\textsuperscript{197} However, when M.C. was sixteen-months old, doctors performed sex reassignment surgery to make M.C.'s body look like a female.

Defendant employees, in their capacity as M.C.'s guardians, were charged with making all medical decisions for M.C. from the time he was removed from his biological parents until the day he was adopted.\textsuperscript{198} As such, they were instrumental not only for the purposes of ultimate legal consent for the surgery, but also in the treatment plan itself and its implementation.\textsuperscript{199} Among other things, SCDSS "coordinated [all] the logistical steps needed to implement [M.C.'s sex assignment,]"\textsuperscript{200} told M.C.'s foster parents when to bring M.C. to the hospital for the surgery,\textsuperscript{201} and

\begin{footnotes}
\item[193] Id.
\item[194] Id. \textsuperscript{¶} 46c.
\item[195] See id. \textsuperscript{¶} 46d ("My bias at the moment is towards female, although I have raised the possibility, because of the substantial virilization of the external genitalia, that there may have been sufficient testosterone imprinting to question ultimate gender identity.").
\item[196] Id. \textsuperscript{¶} 49.
\item[197] Id. \textsuperscript{¶} 52. Defendants, in their motions following filing of the complaint, have accepted the facts presented by Plaintiff in the Complaint as true. However, they do not concede to all of the facts and stated that "[t]he fact that these Defendants reference or incorporate certain of the Plaintiff's factual assertions in this motion does not represent a waiver of these Defendants to later challenge any of those factual assertions." Motion to Dismiss, \emph{supra} note 2, at 2 n.2.
\item[198] Id. \textsuperscript{¶} 55.
\item[199] Id. \textsuperscript{¶¶} 55–63.
\item[200] Id. \textsuperscript{¶} 57.
\item[201] Id. \textsuperscript{¶} 58.
\end{footnotes}
authorized the surgery both over the telephone\textsuperscript{202} and in writing.\textsuperscript{203}

The Crawfords contacted SCDSS in June 2006 after they saw M.C.’s profile on the State’s adoption website and were interested in adopting him.\textsuperscript{204} They learned of his condition and “called the agency and clearly expressed the family’s desire not to subject M.C. to unnecessary sex assignment surgery.”\textsuperscript{205} Unfortunately, the Crawfords were too late, as the surgery was already completed on April 18, 2006.\textsuperscript{206} M.C. might not have been subjected to this sex assignment surgery if the Crawfords learned of his condition sooner, but the damage was already done. The Crawfords “gained custody of M.C. in August, 2006, and legally adopted him on December 11, 2006.”\textsuperscript{207} They “initially raised M.C. as a female in accordance with the gender” assigned through surgery,\textsuperscript{208} but “[h]is interests, manner and play, and refusal to be identified as a girl indicate that M.C.’s gender has developed as a male.”\textsuperscript{209} “M.C. is currently eight years old”\textsuperscript{210} and “is living as a boy with the support of his family, friends, school, religious leaders, and pediatrician.”\textsuperscript{211} That does not change the fact, though, that the government gave their legal consent to a medically unnecessary and invasive surgery.\textsuperscript{212} Defendants permanently altered M.C.’s body, and though he has the ability to live as a boy in his community, there is no way to

\begin{flushright}
\textsuperscript{202} Id. ¶ 59. \\
\textsuperscript{203} Id. ¶ 61. The defendants even signed a form called the “check list of necessary information” that SCDSS requires when a child in their care undergoes “any ‘major surgery’ requiring in-patient hospitalization.” \textit{Id.} \textsuperscript{204} Id. ¶ 64. \\
\textsuperscript{205} Id. Mrs. Crawford was knowledgeable about the surgery from the negative experience of a childhood friend and did not want M.C. to likewise undergo unnecessary surgery. \\
\textsuperscript{206} Id. ¶ 51. \\
\textsuperscript{207} Id. ¶ 64. \\
\textsuperscript{208} Id. ¶ 65. \\
\textsuperscript{209} Id. (“His interests, manner and play, and refusal to be identified as a girl indicate that M.C.’s gender has developed as male.”). \\
\textsuperscript{210} Id. ¶ 2. \\
\textsuperscript{211} Id. ¶ 65. \\
\textsuperscript{212} Id. ¶ 8.
\end{flushright}
regain what the government took from M.C.—his autonomy.  

III. HEALTH CARE CONSENT IN FOSTER CARE

State foster care systems have specific procedures in place to determine whether or not the government has the ability to consent to medical treatment for a child in their care, and the analysis shows that they do not have the legal ability to consent to sex assignment surgery. As established by federal law, every child in foster care is entitled to a minimum set of health services. Federal law serves as the base guidelines for care, but since states are afforded some discretion, particularly with those programs funded by Medicaid, there is often some difference between state and federal laws. Each state has its own agency, such as an Office of Health and Human Services or Department of Social Services (“DSS”), which ensures proper care and safety for children and families. The goal of the foster care

213 Id. ¶¶ 8–11.
214 “Foster care means 24-hour substitute care for children placed away from their parents or guardians and for whom the State agency has placement and care responsibility.” 45 C.F.R. § 1355.20 (2012). As M.C.’s case illustrates, the State Department of Social Services retains ultimate decision-making control for children in the state foster care system. See M.C. Complaint, supra note 2, ¶ 39.
216 Schweitzer & Larsen, supra note 215, at 10.
217 The exact name of the agency varies by state but Department of Social Services will be used for this Note to encompass all alike agencies.
system is to return the child to his or her parents when possible, or to place the child in a different permanent home. This is the case whether children are placed into the foster care system voluntarily or removed by a court proceeding. Though each state has its own set of regulations, the laws often closely resemble each other. The Massachusetts and New York laws will be used as the prototypical systems for purposes of this Note. While there are some variations between state foster care laws, no officials in any state should be able to legally consent to surgery for intersex infants in their custody and under their care.

In deciding “who can consent to medical care, the first determination is whether an emergency exists[,]” since consent is not required for emergency circumstances. As discussed earlier, physicians and the AAP acknowledge that intersex conditions do not present immediate health risks that would constitute a medical emergency. A medical emergency is generally a condition that is life-threatening and, as such, necessitates immediate attention in order to prevent death or hinderance to the individual’s mental or physical well-being. Even M.C.’s physicians said that he could live with the condition and that it was not necessary to immediately perform surgery.

When there is no emergency circumstance, the next “question is whether the treatment is routine . . . or extraordinary” as defined by the state statutes. Routine medical care includes treatments such as dental care, developmental assessments,

---

219 See, e.g., MASS. GEN. LAWS ch. 110, § 1.03 (2011).
220 See, e.g., id. ch. 119, § 23 (outlining the rules and regulations of the department for providing foster care).
222 Massachusetts and New York present the most comprehensive set of state laws regarding medical consent and the foster care system.
223 110 MASS. CODE REGS. § 11.01.
224 Id.; see also N.Y. PUB. HEALTH LAW § 2504.4 (McKinney 2012).
225 Evaluation of the Newborn, supra note 31, at 138.
226 110 MASS. CODE REGS. § 11.03.
227 M.C. Complaint, supra note 2, ¶ 46.
228 110 MASS. CODE REGS. § 11.01.
immunizations, preventative health services, and vision tests. If the treatment is not routine, there are two ways to proceed: either the DSS must obtain a court order, or the parent or patient may provide informed consent. There are certain treatments that are explicitly deemed extraordinary, such as sterilization, life-prolonging medical treatment, and administration of antipsychotic drugs. Since treatment for intersex children is not one specifically outlined in any state statutes, it becomes necessary to weigh certain factors outlined in the state statutes to determine if the care is considered extraordinary. “If it is not extraordinary, it is routine. There is no other possibility.” While the Department may consent to routine health care, if it is found that extraordinary or non-routine health care is necessary, there is a higher burden on the DSS.

229 Id. § 11.04. The examples provided are not an exhaustive list, as other treatments listed in the statute are considered routine as well.

230 See id. § 11.17(2) (“The Department shall not give its consent to extraordinary medical treatment for any child in the care or custody of the Department. For all such children, the Department shall seek prior judicial approval for any extraordinary medical treatment (unless parental consent is obtained . . . .”); 55 PA. CODE § 3130.91(1)(iii) (“If the child is placed under a voluntary placement agreement [the county agency] shall obtain an order of the court authorizing routine or nonroutine treatment if the child’s parent refuses, or cannot be located to provide consent.”).

231 See NEW YORK STATE OFFICE OF CHILDREN & FAMILY SERVS., WORKING TOGETHER: HEALTH SERVICES FOR CHILDREN IN FOSTER CARE 6-9 (2009), available at http://ocfs.ny.gov/main/sppd/health_services/manual/Chapter%206%20Consent.pdf [hereinafter WORKING TOGETHER] (stating that the worker’s actions for consent depends on the legal authority with which the child was placed in foster care., as a parent must still provide consent for non-routine care if they voluntarily placed their child in foster care).

232 110 MASS. CODE REGS. §11.11.

233 Id. §11.12.

234 Id. §11.14.

235 Id. §11.01.

236 Id.

237 Id. § 11.04(2); N.Y. COMP. CODES R. & REGS. tit. 18, § 441.22 (2013).

238 See 110 MASS. CODE REGS. § 11.17(2) (“The Department shall not give its consent to extraordinary medical treatment for any child in the care or custody of the Department. For all such children, the Department shall seek
Physicians are obligated to seek consent to treatment either from the patient himself or the parent, unless there is an emergency circumstance that makes consent impossible.\textsuperscript{239} When a child is under the control of the DSS, the DSS takes the legal place of a child’s parent, which means that it has the power to give consent as a parent ordinarily would.\textsuperscript{240} The Restatement (Second) of Torts provides that an emergency situation is one in which the person needs immediate care “in order to prevent harm to [him]” and where “the actor has no reason to believe that the other, if he had the opportunity to consent, would decline.”\textsuperscript{241} A child born with an intersex condition does not require immediate care or surgery to prevent harm to his or her well-being, as the condition is not considered a medical emergency.\textsuperscript{242} Further, with the new evidence to suggest that many children would have chosen not to undergo the surgery had they had the option at the time,\textsuperscript{243} the second requirement that the actor believe the person, if possible, would choose the same thing, cannot be satisfied.

\textbf{A. Court Order Necessary for Extraordinary Care}

Sex assignment surgery for intersex children is clearly an extraordinary procedure when the factors outlined in state statutes are evaluated. If the DSS officials deem a procedure extraordinary, they must obtain a court order to proceed with the treatment.\textsuperscript{244} Once there is an indication that a treatment prior judicial approval for any extraordinary medical treatment (unless parental consent is obtained . . . .)); WORKING TOGETHER, supra note 231, at 6–9.

\textsuperscript{239} RESTATEMENT (SECOND) OF TORTS § 892D (1979).

\textsuperscript{240} See Schweitzer & Larsen, supra note 215, at 2–6.

\textsuperscript{241} RESTATEMENT (SECOND) OF TORTS, supra note 239, § 892D.

\textsuperscript{242} See Evaluation of the Newborn, supra note 31, at 138. One scholar writing about the issue of informed consent noted that “it is the parents and doctors of intersexed infants who are experiencing a medical emergency, not the intersexed infant. Intersexed genitalia make almost everyone—doctors, parents, and society as a whole—uncomfortable.” Ford, supra note 51, at 477.

\textsuperscript{243} See Coventry, supra note 1, at 56; Hendricks, supra note 16; Tamar-Mattis, supra note 54, at 68–72, 76, 78.

\textsuperscript{244} 110 MASS. CODE REGS. § 11.17(2) (2013). See also 55 PA. CODE § 3130.91(2) (2012) (stating that the Department must obtain either parental consent or a court order for non-routine treatment).
necessitates a court order, the question, which in the area of sex assignment surgery an American court has never had the opportunity to decide, is whether or not the court should allow the procedure to go forward. Massachusetts provides an outline of factors to consider when determining whether medical treatment is extraordinary. They include: “complexity, risk and novelty of the proposed treatment,” “possible side effects,” “intrusiveness of proposed treatment,” “prognosis with and without treatment,” “clarity of professional opinion,” “presence or absence of an emergency,” “prior judicial involvement,” and “conflicting interests.” Sex assignment surgery on infants must be deemed extraordinary, especially when the factors of intrusiveness and prognosis with and without treatment are analyzed.

The guidelines outlined by state statutes provide that “[t]he more intrusive the treatment the greater the need to determine that the treatment is extraordinary, and obtain parental consent or to seek judicial approval prior to authorizing treatment.” The Massachusetts statute cites case law to illustrate the kind of treatment that has been deemed intrusive and necessary for judicial approval. Because DSS, not a court, decides whether the procedure should be deemed extraordinary, the court

245 See Tamar-Mattis, supra note 54, at 81 (“[N]o one in the United States has questioned in court the parental authority to make this decision.”). Likewise, M.C.’s case is the first challenge to the government’s ability to make this decision.
246 110 MASS. CODE REGS. § 11.17(1).
247 Id. § 11.17(1)(a).
248 Id. § 11.17(1)(b).
249 Id. § 11.17(1)(c).
250 Id. § 11.17(1)(d).
251 Id. § 11.17(1)(e).
252 Id. § 11.17(1)(f).
253 Id. § 11.17(1)(g).
254 Id. § 11.17(1)(h).
255 Id. § 11.17(1).
256 Id. § 11.17(1)(c).
257 Id.
258 Id. § 11.17.
merely concludes whether or not the treatment sought is appropriate. Examples of treatments deemed extraordinary by both the DSS and the courts, as cited by the Massachusetts legislature, include life sustaining procedures, but there is also a case where the court evaluated a petition for an order to sterilize a mentally handicapped woman. Sex assignment surgery should be considered extraordinary. It is not only physically invasive, but it also deeply affects the child’s ability to make decisions regarding his or her identity and reproductive future. Those individuals who undergo surgery may be denied the right to reproduce in the future. The invasive nature of the procedure renders it intrusive enough to be submitted to the court.

Once the Department uses the outlined factors to determine that a procedure is extraordinary, it must “seek prior judicial approval for [the] extraordinary medical treatment.” The courts then use a “substituted judgment” view for evaluating whether or not the procedure should be authorized, and ask whether, given all surrounding circumstances, the individual patient would want the treatment, regardless of what the ordinary or prudent person might want. The standard seems to be one that is reasonable and takes into account the individual’s circumstances, but there is reason to think that the standard of judicial review of a proposed sex assignment surgery should be even higher. There is no way to accurately predict whether or not the child will accept the sex that the doctors assign, and given that many of the intersex individuals that have undergone sex assignment surgery have spoken out against it, it seems unlikely that a court would be able to

---

260 See In re Mary Moe, 432 N.E.2d 712 (Mass. 1982).
261 See Haas, supra note 27, at 42–43 (“Genital reconstruction surgery may result in scarred genitals, an inability to achieve orgasm, or an inability to reproduce naturally or through artificial insemination.”).
262 Id. at 48.
263 110 MASS. CODE REGS. § 11.17(2).
265 Tamar-Mattis, supra note 54, at 68–72.
conclude that a patient would want to undergo surgery at such a young age.

Another factor that weighs in favor of deeming sex assignment surgery as an extraordinary procedure is that of the “prognosis with and without treatment.” The less clear the benefit from the proposed treatment the greater the need for parental consent or prior judicial approval. Many children born with intersex conditions have the ability to live healthy lives without any surgical intervention and instead can be assigned a gender without surgery. The alleged benefit of this surgery is said to be one of “normality,” but there is no indication either that children who have the surgery feel “normal” or that they are any better off later in life. In fact, “rather than alleviating feelings of freakishness, in practice the way intersexuality is typically handled may actually produce or contribute to many intersexuels’ feelings of freakishness.” There has been an inadequate amount of follow-up with large groups of intersex individuals who have undergone surgery to determine the exact results. However, anecdotal evidence and those studies that

\[266\] 110 MASS. CODE REGS. § 11.17(d).
\[267\] Id.
\[268\] Evaluation of the Newborn, supra note 31, at 138.
\[270\] See Dreger, supra note 6, at 30–31 (noting that much of the treatment is fueled by desire to fit within the definitions of normality); Tamar-Mattis, supra note 54, at 67 (noting that parents want their children to have a normal childhood, which includes having a “normal” gender identity).
\[271\] See Haas, supra note 27, at 48 (indicating that there is no proof to support the idea that people who have sex assignment surgery are any better off than those that do not have it).
\[272\] Dreger, supra note 6, at 31.
\[273\] See FAUSTO-STERLING, supra note 41, at 85 (“[L]ong-term studies of genital surgery are as scarce as hen’s teeth.”); KESSLER, supra note 31, at 53 (“Surprisingly, in spite of the thousands of genital operations performed every
have been conducted lend themselves to the idea that more people would have preferred not to have the surgery and been given the freedom to decide at a later point whether or not they wanted it.274 Given the lack of adequate scientific knowledge, it is difficult for physicians to argue that there are any real substantial benefits from the surgery.

It seems clear that sex assignment surgeries would undoubtedly qualify as extraordinary procedures that necessitate judicial approval, because it is so intrusive into the individual’s well-being and future, and the benefits are so unclear. Sex assignment surgery denies an intersex child the opportunity to make a decision at a later point about whether or not to undergo a medically unnecessary procedure that will forever affect his or her life. If DSS sought court approval for one like M.C.’s, the court could prevent the surgery from happening in the first place, instead of assessing the consequences of a surgery only after it happens.

B. The Doctrine of Informed Consent

Instead of requiring a court order for extraordinary medical treatment, some states allow for consent to non-routine medical treatment as long as that consent is informed.275 There are three prongs that must be satisfied in order for a medical decision to be considered legally informed: (1) the decision must be informed; (2) it must be voluntary; and (3) the patient must “have an appreciation of the nature, extent, and probable consequence of

---

274 See Coventry, supra note 1; Hendricks, supra note 16; Tamar-Mattis, supra note 54, at 68–72, 76, 78. Cf. Kessler, supra note 31, at 94–96 (recounting stories of parents who were happy that they chose not to have their child undergo sex assignment surgery).

275 See, e.g., Working Together, supra note 231.
the conduct consented to.” The informed consent doctrine protects “the right of every individual to the possession and control of his own person, free from all restraint or interference of others, unless by clear and unquestionable authority of law.” Because minors are unable to legally give consent to medical treatment, parents must give “proxy consent” on their behalf. The AAP recognizes that such consent “poses serious problems for pediatric healthcare providers” because they must balance doing what is in the “best interest of the child” with the wishes of the parents. It is difficult for physicians to define what is best for a certain child given the differences in “religious, social, cultural, and philosophic positions on what constitutes acceptable child rearing and child welfare.” “Courts have upheld parental consent on the basis that parents, as natural guardians of their children, are best situated and best able to make important decisions on their behalf.”

However, some scholars contend that parents make decisions about what sex to assign their child without full disclosure from doctors about the ramifications for the future. Doctors use language that is thought to comfort the parents, but it often

276 Ford, supra note 51, at 475 (internal quotation marks omitted).
278 Ford, supra note 51, at 477 (“Infants’ underdeveloped communication and comprehension abilities preclude appreciation of the nature, extent, and probable consequences of a proposed treatment. Nor can they weigh its alternatives. Therefore, infants are literally unable to give legal informed consent for their own medical treatment.”).
280 Id.
281 Id.
282 Claudia Wiesemann et. al., Ethical Principles and Recommendations for the Medical Management of Differences of Sex Development (DSD)/Intersex in Children and Adolescents, 169 EUR. J. PEDIATRICS 671, 674 (2010).
283 Committee on Bioethics, supra note 279, at 315.
284 Ford, supra note 51, at 478.
286 See Kessler, supra note 31, at 22–24.
“leaves parents ill-equipped to make thoughtful decisions.”

Many authors argue that because of this failure to disclose all necessary facts, parents cannot truly provide informed consent for their children to undergo surgery.\(^{287}\) If parents cannot provide informed consent for sex assignment surgery on their child, it seems even less likely that a governmental department has the ability to give consent. Such a department lacks the same relationship and set of interests that any biological parent has with the child and therefore cannot really provide informed consent.

Thus, the central issues that scholars tend to focus on when discussing consent for intersex surgery are two-fold. First, that they do not have all of the information necessary to make a truly informed decision,\(^{289}\) and second, that the parents and the child have potentially conflicting interests that prevent the parent from acting in the true best interest of the child.\(^{290}\) As discussed earlier, doctors often relay information to parents with the aim of catering to the parents’ emotional needs,\(^{291}\) but that also means many of the facts central to the decision-making process are not conveyed.\(^{292}\) Further, even if parents do have all of the necessary information, it is possible that they could consent to surgery not because they believe it is in the best interest of the child, but for any other reason.\(^{293}\)

Parents are also often driven by their own values and worries,\(^{294}\) and by their guilt that they are responsible for their

\(^{287}\) Beh & Diamond, supra note 15, at 48.

\(^{288}\) See Beh & Diamond, supra note 15, at 1; Dreger, supra note 6, at 32–33; Ford, supra note 51; Hermer, supra note 6, 222–25.

\(^{289}\) See id. at 1.

\(^{290}\) Tamar-Mattis, supra note 54, at 88.


\(^{292}\) Haas, supra note 27, at 62.

\(^{293}\) Tamar-Mattis, supra note 54, at 88 (asserting that parents could “authorize the surgery for any reason—parental discomfort, embarrassment over raising a son with a small penis or a daughter with a noticeable clitoris, desire for a child of one gender or the other—as long as they were fully informed of the risks.”)

\(^{294}\) Navarro, supra note 16.
“imperfect” child.295 One parent, concerned about what she thought her child would experience in the future if she did not have the surgery said, “Growing up a teenage girl is hard enough. I never want her to feel different. I never want her to have extra issues to deal with.”296 While parents do the best they can in this situation, “the decision to perform surgery may be centered more around the needs of the caregivers than the needs of the child.”297 That is not to suggest that parents do not have genuine intentions when they have to make the decision to have surgery; our cultural norms force people into one of two genders, and many parents do not want to “risk what they believe to be the well-being of their child in order to protest a cultural norm.”298 “Parents and families are . . . accorded a certain sphere of privacy to pursue their personal aims and find out what is their best in child care.”299 However, “intersex babies are not having difficulty with sexual identity or self-image. The parents are, and parental anxiety about the appearance of a child’s genitals should be treated with counseling, not with surgery to the child.”300 If there is this much debate and speculation among scholars about whether or not parents have the ability to provide informed consent for their own children, then it bears asking the question of whether or not the government, in the form of a state Department of Social Services, should have the ability to provide such consent.

The government does not aim to keep children in its care for the entirety of the children’s life, but instead aims to have only

295 Ford, supra note 51, at 487.
296 Navarro, supra note 16 (internal quotation marks omitted).
297 Tamar-Mattis, supra note 54, at 89.
298 CARL ELLIOTT, BIOETHICS, CULTURE AND IDENTITY: A PHILOSOPHICAL DISEASE 40 (1999) (“I suspect parents are often terrified at the prospect of their children being outcasts, of being seen as freaks of nature, of being desperately unhappy, of being completely bewildered about their place in the world, of never being able to attract a sexual partner, of being forced to live a life of secrecy and shame, of being tortured and bullied and ridiculed by other children while they are growing up. And who is to say that these fears are not justified? It would be a mistake to overlook the consequences of damaging and stigmatizing cultural pressures an intersexed child may face.”).
299 Wiesemann et. al., supra note 282, at 674.
300 Weil, supra note 273.
temporary care until the child can be placed in a permanent home.\textsuperscript{301} The government’s intent to keep a child for the shortest amount of time possible runs counter to the idea that they should be able to make such a life changing and permanent decision as the sex of a child. When there are multiple social workers working on one child’s case, as was the situation for M.C.,\textsuperscript{302} there seems to be a greater possibility that not all information is being funneled through one person who is legally charged with giving consent.\textsuperscript{303} Thus, the chance that full comprehension of the condition and appreciation for the ramifications of surgery will be lost only increases. A team of social workers was responsible for making and coordinating all of M.C.’s medical decisions,\textsuperscript{304} but there is no indication that one person was primarily responsible for collecting all medical information from the doctors and communicating with other members of the team.\textsuperscript{305} Unlike parents who are able to hear all information from a physician and weigh the options, case workers may be forced to compile piecemeal information from different sources and present it to the person able to give legal consent.\textsuperscript{306} Even if one case worker spoke with the doctor at all times, his or her ability to provide proxy consent for such an invasive surgery is severely hindered because they lack the same continued interest that parents have.

One may argue that the lack of parental relationship makes the government a better candidate for making this decision

\begin{footnotesize}
\begin{enumerate}
\item[301] See, e.g., 110 MASS. CODE REGS. § 1.03 (2013).
\item[302] M.C. Complaint, supra note 2, ¶¶ 21–28.
\item[303] See id. ¶¶ 55–63 (stating that there were multiple case workers that spoke with medical officials and a supervisor who signed the consent form for the procedure).
\item[304] Id. ¶¶ 55–61.
\item[305] Five of the defendants named in the complaint received communications from the defendant doctors. Id. ¶ 57.
\item[306] In M.C.’s case, there was “sporadic attendance of multiple SCDSS case workers at conferences regarding M.C.’s medical treatment within different time frames . . . .” Crawford Complaint, supra note 171, ¶ 38. The Director of the South Carolina Department of Social Services was required to sign a checklist of information before the surgery could be performed, which she did sign. M.C. Complaint, supra note 2, ¶ 61. Verbal authorization was given via telephone to the surgical nurse by one of the social workers. Id. ¶ 59.
\end{enumerate}
\end{footnotesize}
because it does not have the same emotional guilt and set of fears that a parent would have. But the government employees making the actual decisions have the same feelings of discomfort that any adult may experience with ambiguous genitalia. Arguing that the government is in a better position to make a decision about surgery because of its lack of emotional connection fails to appreciate that government officials are still driven by society’s pervasive idea of “normality” and a strictly defined gender system. Further, it also means that the government may not truly know what is in the best interest of the child since there is no way to predict what kind of environment the child will subsequently be raised in. Finally, a government agency is not committed to seeing the sex assignment process through to completion. One surgical procedure “does not alter the chromosomal, genetic or hormonal determinants of sex and so does not change an intersex child . . . into an infant of the assigned sex.” There is most often the need for more than one surgery and there are complex psychological ramifications that both parent and child have to address as the child matures. The social worker, by virtue of his relationship to the child, does not have the ability to make sure that the child gets the follow-up care and psychological counseling that is essential for his or her development.

Thus, the problems with government officials providing true informed consent for a child to undergo gender assignment surgery are multi-fold. The person who is charged with providing

307 Ford, supra note 51, at 487.
308 See Tamar-Mattis, supra note 54, at 89.
309 See Beh & Diamond, supra note 15, at 49.
310 Fausto-Sterling, supra note 41, at 86–87 (“From 30 to 80 percent of children receiving genital surgery undergo more than one operation. It is not uncommon for a child to endure from three to five such procedures.”).
311 See generally Preves, supra note 14, at 60–86 (recounting the stories of people who underwent surgery and worked to accept their identity); Kessler, supra note 6, at 22 (noting that “at adolescence, the child may be referred to a physician for counseling.”).
312 Kessler, supra note 31, at 27–29. There is also most often the need for hormonal therapy as the child develops to ensure that he or she continues to develop into the sex assigned by the doctors. Elliott, supra note 298, at 37.
legal consent might not have all of the necessary and relevant information, making consent is impossible; and even if he or she does have all of the information, they lack the ability to know what is truly in the best interest of the child.

While it is the case that children in foster care are entitled to receive certain basic health care services from their own state’s Department of Social Services, as the government has legal responsibility over the child, that does not give the government the unfettered discretion to consent to any procedure. The Department may authorize treatment for children in any emergency circumstance, but intersex conditions do not qualify as such; they are merely a deviation from society’s conception of normality. Even the doctor’s treating M.C. recognized that there was no urgent reason to surgically make M.C. into either a definitive male or female. The government may provide consent for routine medical care, but when care is deemed extraordinary, the officials must either obtain a court order or provide true informed consent to proceed with the treatment. Sex assignment surgery on intersex children is undoubtedly an extraordinary procedure that requires an elevated level of consent either from the court or from the DSS. The DSS officials handling M.C.’s case did not obtain consent of either type and overstepped their legal authority to elect procedures for a child in their care.

---

313 See, e.g., M.C. Complaint, supra note 2, ¶¶ 55–61.
314 Schweitzer & Larsen, supra note 215, at 10.
317 Dreger, supra note 6, at 30.
318 M.C. Complaint, supra note 2, ¶ 46.
321 See Working Together, supra note 231.
322 See generally M.C. Complaint, supra note 2.
IV. INTERNATIONAL RESPONSE AND SOLUTION

Several countries outside of the United States have started to address surgical procedures on intersex children and have adopted resolutions and laws to decrease both the ability and pressure to elect sex assignment surgery. Just a few months ago, the Parliamentary Assembly of the Council of Europe “addressed the issue of bodily integrity of intersex children” when they adopted a novel resolution, the “children’s right to physical integrity” resolution. It calls for “Council of Europe Member States to ‘undertake further research to increase knowledge about the specific situation of intersex people, ensure that no-one is subject to unnecessary medical or surgical treatment that is cosmetic rather than vital for health during infancy or childhood, guarantee bodily integrity, autonomy and self-determination to persons concerned, and provide families with intersex children with adequate counselling and support . . . .’” The resolution looks at the issue of intersex children from the “human rights perspective, rather than a medical approach.” The resolution categorizes “medical interventions in the case of intersex children” as “procedures that tend to present as beneficial to the children themselves despite clear evidence to the contrary[,]” perhaps giving credence to all of the first-hand statements from intersex individuals who were subjected to surgery.

There is also hope for a decrease in the societal pressure on

324 Id.
325 PARLIAMENTARY ASSEMBLY OF THE COUNCIL OF EUR., RESOLUTION 1952: CHILDREN’S RIGHT TO PHYSICAL INTEGRITY (2013) [hereinafter RESOLUTION 1952]. The resolution also addresses other issues regarding cosmetic medical procedure, such as tattooing, piercing, and female genital mutilation.
326 Id.
328 RESOLUTION 1952, supra note 325.
parents to quickly choose a gender for an intersex child. Just a few months ago, Germany became the first European country to allow parents to register their child’s birth certificate with neither the male nor female designation.\textsuperscript{329} Prior to the implementation of this new law, German parents experienced pressure to decide which sex to assign their intersex child and often made rushed decisions, as they had only a week to register their child at the registry office.\textsuperscript{330} This new law allowing parents to decide their child’s sex later “is an effort to create legal recognition for intersex individuals,”\textsuperscript{331} and “to give parents and children more time before making life-changing sex reassignment decisions.”\textsuperscript{332} According to the interior ministry, passports will soon have an “X” designation, in addition to the already-present male and female.\textsuperscript{333} Other countries, including Australia, New Zealand, and Bangladesh, already have similar laws that allow individuals to select “X” or “other” on their passport application.\textsuperscript{334} In contrast, birth certificates in the United States must be submitted within a period of time set by the state\textsuperscript{335} and must have a female or male designation; there is no other option.\textsuperscript{336}

\textsuperscript{329} Germany Allows “Indeterminate” Gender at Birth, BBC NEWS (Nov. 1, 2013), http://www.bbc.co.uk/news/world-europe-24767225.

\textsuperscript{330} Jacinta Nandi, Germany Got it Right by Offering a Third Gender Option on Birth Certificates, THE GUARDIAN (Nov. 10, 2013), http://www.theguardian.com/commentisfree/2013/nov/10/germany-third-gender-birth-certificate.

\textsuperscript{331} James Nichols, Germany to Allow Parents to Choose No Gender for Babies on Birth Certificates, HUFFINGTON POST (Oct. 31, 2013), http://www.huffingtonpost.com/2013/10/31/germany-intersex_n_4181449.html.

\textsuperscript{332} Michelle Castillo, Germany to Allow Third Gender Designation on Birth Certificates, CBS NEWS (Nov. 1, 2013), http://www.cbsnews.com/news/germany-to-allow-third-gender-designation-on-birth-certificates/.

\textsuperscript{333} Germany Allows “Indeterminate” Gender at Birth, supra note 329.

\textsuperscript{334} Id.

\textsuperscript{335} Kessler, supra note 6, at 14 (“New York State requires that a birth certificate be filled out within forty-eight hours of delivery, but the certificate need not be filed with the state for thirty days.”).

\textsuperscript{336} States have different laws that allow people to change their designated sex on their birth certificates, generally referencing transgender individuals. Some state laws require that the individual have surgery to change his or her sex before they will be eligible to change their birth certificate sex designation,
Some argue that Germany’s law does not go far enough: without an outright ban on the surgeries, they will continue because “we live in a world where having a baby classified as ‘other’ is still considered undesirable.”  However, the mere fact that parents have the option of avoiding the narrow male or female designation on the birth certificate is a step forward that the United States can emulate. The societal change to accepting someone designated as a gender other than male or female may take much longer than seems desirable, but there are more immediate options to ensure that foster children born with an intersex condition are not subject to sex assignment surgery. The legal change for birth certificate designation may alleviate the pressure that a child’s caregiver feels to make an immediate assignment, but until then, case workers must be educated about intersex conditions and care for the child just as they would any other child.

Resolutions like those in Europe that call attention to the impact of sex assignment surgery on intersex children can help the United States approach the problem from a new perspective. Allowing parents and government officials to register a child as something other than just “male” or “female” would allow them to consider all options without the pressure of having such a short time constraint. M.C. and children like him might be in a very different, and presumably better, situation if social workers allow them to make decisions about surgery themselves at a later point in life.

CONCLUSION

Modern culture has created a sharply divided system of gender normality that deeply affects the way in which someone
that does not fit neatly into one of those two categories is perceived.\textsuperscript{340} The debate over whether parents can elect sex assignment surgery will most likely continue, but there is no room or basis of authorization for a foster care system to elect that same surgery. Social Services officials have temporary care\textsuperscript{341} of a child in their custody and are unable to foresee the environment a child will be adopted into and raised in. Their interests are significantly different from those of a child’s biological parents and as such, they are unable to provide truly informed consent.

Adults may experience discomfort with the appearance of ambiguous genitalia and the uncertainty of a child’s sex,\textsuperscript{342} but the solution is not to subject the child to surgery that will have permanent and lasting effects on his or her body. M.C.’s life has forever been changed, no matter what the courts decide in his case. The policy and legal changes made recently around the world\textsuperscript{343} can serve as a template and guide for policies that the United States may adopt to aid parents in making this kind of difficult decision and allowing them more time to do so.\textsuperscript{344} But whether or not those changes come to the United States, state governments must allow intersex children to grow as they ordinarily would without surgery and provide the support that they need.

There is also little indication at this point that children who had gender assignment surgery are any better off than those who did not.\textsuperscript{345} In fact, the few studies that have been conducted indicate that adults who underwent surgery wish they had not had

\begin{itemize}
\item \textsuperscript{340} Beh & Diamond, \textit{supra} note 291, at 15 (“Fear of the monster still continues to dominate the decisional process.”)
\item \textsuperscript{341} See, e.g., 110 MASS. CODE REGS. § 1.03 (2013).
\item \textsuperscript{342} Tamar-Mattis, \textit{supra} note 54, at 89.
\item \textsuperscript{343} See \textit{supra} Part IV.
\item \textsuperscript{344} Castillo, \textit{supra} note 332 (writing that a goal of the German law was to allow more time for parents to decide on their child’s sex).
\item \textsuperscript{345} Haas, \textit{supra} note 27, at 48; Beh & Diamond, \textit{supra} note 15, at 24 (“[C]ritics point to evidence that persons born with genitalia that fall outside our normal expectations can achieve a satisfying psychosexual adjustment without surgical intervention and argue that the imperative to create typical genitalia is of overrated significance.”).
\end{itemize}
them, or have transitioned from the gender that was surgically assigned to it. Some suggest that surgery is the best fix because it helps to “normalize” people who would otherwise have a hard time because of their differences. But as one scholar aptly analogized: “[W]e still live in a nation where dark-skinned people have a harder time than light-skinned people do. But would he suggest we work on technologies to ‘fix’ dark skin?” Surely not.

346 Beh & Diamond, supra note 291, at 24–25 (recounting testimony of people who wish they had the ability to choose for themselves whether or not they wanted surgery); Kuhnle & Krahl, supra note 30, at 96.

347 See Dreifus, supra note 34 (“The most important [of findings] is that about 60 percent of the genetic male children raised as female have retransitioned into males.”).


349 Id.
DRUGS: YOU USE, YOU GAIN? WHY COURTS SHOULD UPHOLD LONG-TERM DISABILITY BENEFITS FOR RECOVERING ADDICTS

Gregory M. Juell*

I. INTRODUCTION

In July 2004, a member of a Massachusetts hospital’s nursing staff found Dr. Julie Colby, an anesthesiologist, unconscious on a hospital table.\(^1\) Dr. Colby had served as a partner in a Merrimack anesthesiology practice for sixteen years when she became addicted to Fentanyl, an opioid commonly used in the practice.\(^2\) She took a leave of absence to enter an inpatient substance treatment facility, where she was diagnosed with an opioid dependence, depression, and obsessive-compulsive personality traits.\(^3\) Pursuant to her employer’s group employee benefit plan, her insurer provided long-term disability (LTD) benefits during inpatient treatment.\(^4\) She remained at the treatment facility until November 2004 when she left to begin outpatient treatment, during which she was under regular medical supervision and did not resume her use of Fentanyl.\(^5\) Nevertheless, the Massachusetts Board of Registration in Medicine revoked her license and her insurer refused to provide benefits for any of her outpatient

* J.D. Candidate, Brooklyn Law School, 2015; A.B., Dartmouth College, 2007. I would like to thank the Journal staff, for their helpful comments and edits; my sister, for inspiring my interest in the public health; and my parents, for their unending encouragement and support.

1 Brief of Plaintiff–Appellee at 3, Colby v. Union Sec. Ins. Co., 705 F.3d 58 (1st Cir. 2013) (No. 11-2270).
2 Colby v. Union Sec. Ins. Co., 705 F.3d 58, 60 (1st Cir. 2013).
3 Id.
4 Id. See infra Part II for more background on LTD benefits.
5 Colby, 705 F.3d at 60.
treatment because it did not consider her risk of relapse a “current disability” under her employee benefit plan.\(^6\)

Unfortunately, Dr. Colby’s experience is not particularly rare among anesthesiologists. A 2005 study surveying anesthesiology residency programs from 1991 to 2001 determined that eighty percent of programs reported opioid abuse among residents and nineteen percent reported pretreatment fatalities from opioid abuse.\(^7\) While most residents attempted to reenter anesthesiology after treatment, only forty-six percent who attempted reentry had completed an anesthesiology residency at the time of the survey. The substance-related death rate for those who remained in anesthesiology was nine percent.\(^8\) Forty percent of those who were treated and returned to medicine ultimately entered another specialty.\(^9\) Long-term follow-up for treated residents indicated that fifty-six percent were successful in medicine, though often in a different specialty.\(^10\)

Possible factors contributing to high rates of drug abuse among anesthesiologists include: ease of access to highly addictive drugs, the ease of diverting small quantities for personal use, a high-stress work environment, and the increased sensitivity of the brain’s reward pathways\(^11\) resulting from workplace

---


\(^{7}\) Ethan O. Bryson & Jeffrey H. Silverstein, Addiction and Substance Abuse in Anesthesiology, 109 Anesthesiology 905, 905 (2008).

\(^{8}\) Gregory B. Collins et al., Chemical Dependency Treatment Outcomes of Residents in Anesthesiology: Results of a Survey, 101 Anesthesia & Analgesia 1457, 1457 (2005).

\(^{9}\) Id. at 1459.

\(^{10}\) Id. at 1460. Long-term follow-up data was available for ninety-three percent (185/199) of the study residents. Id.

\(^{11}\) Reward pathways are the parts of the brain that are “responsible for driving our feelings of motivation, reward and behavior.” See The Reward
exposure to the drug. Unsurprisingly then, anesthesiologists and others in the field are common plaintiffs in Employee Retirement Income Security Act (ERISA) actions against insurers that deny LTD benefits to individuals recovering from substance abuse disorders.

Because of the prevalence of anesthesiologist plaintiffs in cases determining whether recovering addicts should be entitled to LTD benefits, the cases discussed in this Note focus on the anesthesiology context. However, whether insurers should be required to provide LTD benefits to recovering addicts is an important question in any field, particularly those in which the public health and safety are at risk. Additionally, this Note focuses on ERISA-governed LTD plans. As the following discussion will demonstrate, it is in this context that the case for treating the risk of relapse as a “current disability” is particularly strong. However, many of the arguments that follow will be equally applicable outside of the ERISA context.

Courts are divided as to whether the risk of relapse into drug addiction constitutes a “current disability.” Under LTD benefit plans, a disability is generally defined for the first year or two as a condition that prevents one from engaging in his regular occupation. Afterward, the definition changes: it requires that


12 Bryson & Silverstein, supra note 7, at 905.


15 Compare Colby, 705 F.3d 58 (upholding reversal of LTD benefit denial), with Stanford v. Continental Cas. Co., 514 F.3d 354 (4th Cir. 2008) (upholding LTD benefit denial). The disability plan at issue in Stanford defined disability as “injury or Sickness [that] causes physical or mental impairment to such a degree of severity that You are . . . continuously unable to perform the Material and Substantial Duties of Your Regular Occupation.” Brief of Appellant at 10, Stanford, 514 F.3d 354 (No. 06-2006). For a claimant to be entitled to benefits, his disability must therefore be “current.” Id.

the individual be unable to perform any gainful occupation.\textsuperscript{17}

If a court determines that the risk of relapse is a “current disability,” then it will require the insurer to provide benefits under standard LTD benefit plans.\textsuperscript{18} The First and Fourth Circuits have come to opposite conclusions on this issue.\textsuperscript{19} The First Circuit in \textit{Colby v. Union Security Insurance Co.}\textsuperscript{20} determined that the risk of relapse into drug abuse is akin to the risk of relapse into cardiac distress or orthopedic complications, and can therefore be so severe as to constitute a current disability for which LTD benefits must be provided. The court explained that a current disability could exist even when an individual is physically capable of performing his job.\textsuperscript{21}

By contrast, the Fourth Circuit in \textit{Stanford v. Continental Casualty Co.}\textsuperscript{22} came to a different conclusion. In \textit{Stanford}, the insurer had determined that the “potential risk of relapse” is not a current disability for which LTD benefits must be provided, and the court held that the insurer did not abuse its discretion in making this determination.\textsuperscript{23} The court ruled that while the risk of relapse into cardiac arrest is a likely result of a stressful work environment, the risk of relapse into substance abuse is a choice.\textsuperscript{24} Also in contrast to the First Circuit, the Fourth Circuit

\begin{flushleft}
\textsuperscript{17} \textit{Id.}
\textsuperscript{18} For more traditional types of conditions for which courts have upheld LTD benefits, see generally \textit{Rothman v. Office Env’ts of New England Health & Welfare Benefit Plan}, 794 F. Supp. 2d 276 (D.Ma. 2011) (awarding LTD benefits to a salesperson who suffered from post-concussion syndrome); \textit{Adams v. Hartford Life & Acc. Ins. Co.}, 694 F. Supp. 2d 1342 (M.D. Ga. 2010) (holding that a plan participant who experienced cognitive problems following a stroke was entitled to LTD benefits); \textit{Alexander v. Winthrop, Simpson, Putnam & Roberts Long Term Disability Coverage}, 497 F. Supp. 2d 429 (E.D.N.Y. 2007) (upholding LTD benefits to a legal secretary who suffered from persistent and severe lower back pain).
\textsuperscript{19} \textit{Compare Colby}, 705 F.3d 58 (upholding reversal of LTD benefit denial), \textit{with Stanford}, 514 F.3d 354 (upholding LTD benefit denial).
\textsuperscript{20} \textit{Colby}, 705 F.3d at 59–60.
\textsuperscript{21} \textit{Id.} at 66.
\textsuperscript{22} \textit{Stanford}, 514 F.3d at 358–59.
\textsuperscript{23} \textit{Id.}
\textsuperscript{24} \textit{Id.} at 358.
\end{flushleft}
distinguished heart conditions from drug addiction on the ground that one who is heart attack-prone has a current physical impairment, while one who risks relapse into substance dependence does not. The court agreed with the insurer that the mere risk of relapse is not a current disability for which the insurer must provide LTD benefits.

This Note examines whether the Fourth Circuit’s decision in Stanford was justified and asserts that Stanford contravenes both Supreme Court precedent and the congressional intent that motivated ERISA’s passage. Furthermore, Stanford is at odds with current psychology literature, which views addiction as a disease rather than a choice, and there is no compelling reason why ERISA plan administrators should treat the risk of relapse differently from other chronic medical conditions. Finally, the Fourth Circuit failed to properly take into account the potentially disastrous public policy consequences of Stanford. The First Circuit’s decision in Colby is more firmly grounded in law and psychology, and it makes for better public policy. Colby therefore provides better guidance for future courts confronted with the issue of whether to construe the risk of relapse as a disability for which LTD benefits should be provided.

Part II provides a brief historical background of ERISA and LTD benefits. Part III details the differences between Colby and Stanford, and discusses related decisions by other courts. Part IV examines recent psychology literature on addiction and how scholars in the field have come to regard addiction as a disease rather than a choice. Finally, Part V examines why Stanford is flawed and argues that courts should therefore follow the First

---

25 Id. at 359.
26 Id. at 361.
27 John Utz questions whether Colby and Stanford truly created a circuit split because the two courts were interpreting different plans. John L. Utz, Addict’s Risk of Relapse as Disability, 21 ERISA LITIG. REP., no. 2, 2013, at 6. However, a true split is apparent given the courts’ completely divergent attitudes regarding the nature of addiction. Utz’s skepticism also ignores Judge Wilkinson’s dissenting opinion in Stanford, which was echoed in Colby, and demonstrates how judges’ differing attitudes toward addiction can result in sharply different interpretations of a benefit plan. See Stanford, 514 F.3d at 361-65 (Wilkinson, J., dissenting).
Circuit in treating the risk of relapse into substance abuse as a “current disability.”

II. HISTORICAL BACKGROUND OF ERISA AND LTD BENEFITS

During the Second World War, several economic factors contributed to an older workforce in the years that followed. One factor was wartime inflation, which discouraged retirement by reducing the value of Social Security Old-Age and Survivors Insurance. Another was the policy of many firms to directly discourage retirement. Due to the resulting older workforce, Congress of Industrial Organizations (CIO) unions began to prioritize the interests of older workers by emphasizing retirement benefits in their collective bargaining agreements. However, increased retirement benefits for older workers typically came at the expense of liberal vesting requirements and other policies that would have benefited younger workers. Additionally, CIO unions often bargained for systems requiring employers to lay off workers in reverse order of seniority.

Events at Studebaker-Packard highlighted the vulnerability of younger workers under these systems. During the 1950s, adverse economic events, such as the loss of wartime defense contracts and a recession, made it more difficult for independent

29 Id.
30 Id.
31 Id.
32 Id. at 688. When funds are “vested,” an employee has an absolute right to them. Employers cannot reclaim vested funds.
33 Id.
35 Wooten, supra note 28, at 684.
automobile manufacturers to compete with larger firms.\textsuperscript{36} As a result of these events, in December 1963 Studebaker-Packard closed its plant in South Bend, Indiana.\textsuperscript{37} To make matters worse for the employees, Studebaker-Packard’s pension plan lacked adequate funds and the company defaulted on its obligations to workers under sixty, with some workers receiving nothing at all.\textsuperscript{38} This was the result of a 1961 collective bargaining agreement, which favored older workers by prioritizing retirees and retirement-eligible employees over younger workers.\textsuperscript{39} The plant’s shutdown gained national attention when advocates of pension reform repeatedly invoked the default as a symbol of the need for regulation and reform.\textsuperscript{40} While the closing of Studebaker-Packard became a rallying cry for pension reform advocates, pension reform remained controversial and it took more than a decade for substantial reform to occur.\textsuperscript{41}

The reform effort culminated on Labor Day in 1974, when President Gerald Ford signed into law the Employee Retirement Income Security Act (ERISA).\textsuperscript{42} Congress enacted ERISA in order to ensure that employees actually receive promised benefits in accordance with a benefit plan’s terms.\textsuperscript{43} To this end, ERISA imposes minimum standards for private industry pension plan administrators and creates causes of action for plan participants and their beneficiaries.\textsuperscript{44} ERISA-imposed duties are derived from the common law of trusts.\textsuperscript{45} Fiduciaries are therefore required to

\begin{itemize}
\item \textsuperscript{36} Id. at 693.
\item \textsuperscript{37} Id. at 683–84.
\item \textsuperscript{38} Id. at 684.
\item \textsuperscript{39} Id. at 731.
\item \textsuperscript{40} Id. at 684.
\item \textsuperscript{41} Id. at 739.
\item \textsuperscript{43} LEE T. POLK, 1 ERISA PRACTICE AND LITIGATION § 1:1 (2013). See also 29 U.S.C. § 1001 (2012).
\item \textsuperscript{44} ERISA, LEGAL INFO. INST., http://www.law.cornell.edu/wex/erisa (last visited Mar. 5, 2014). ERISA also regulates the impact of federal income taxes on the management of benefit plans. Id.
\item \textsuperscript{45} POLK, supra note 43, § 1:6. Trust law establishes principles by which one holds property for another’s benefit.
\end{itemize}
discharge their duties with the prudence of a reasonable man under like circumstances. They must also “act solely in the interests of participants and beneficiaries and for the exclusive purposes of providing benefits and defraying reasonable expenses of the plan.” The statute creates a private cause of action against plan administrators who fail to meet their obligations. While ERISA sets a benefit floor, employers can choose to provide greater benefits. Courts may therefore enforce a contractual obligation to provide benefits beyond what the statute requires.

ERISA’s “standards of fiduciary responsibility” govern both “pension plans” and “welfare plans.” “Pension plans” include an array of deferred compensation plans, while “welfare plans” include a variety of benefits, such as disability insurance. LTD insurance is designed to provide income to employees who are unable to work for extended periods due to prolonged disability. The income amount is generally a predetermined percentage of the employee’s pre-disability earnings. Employees typically must have worked for an employer for a period of five months to a year before becoming eligible for LTD benefits. In addition, employees must be disabled for an extended period, usually three

50 Id.
51 Id., supra note 43, § 1:3.
52 Id.
53 ROUTH, supra note 49, § 2:27 (“It is not uncommon for a plan to provide that disability means the inability to perform one’s regular duties for two years. After that, the definition often changes requiring the person to demonstrate an inability to perform any occupation for which the employee is reasonably qualified.”). This ongoing inability is what renders a disability “current.”
54 Id.
55 Id.
to five months, before LTD benefits begin. Thus, the LTD benefit period typically begins when the short-term disability period ends. LTD benefit payments are generally payable until recovery, a specific age, or retirement. Additionally, LTD payments may be reduced if an employee is only partially disabled, meaning the employee can either perform some duties of his original occupation or can perform another occupation in which his earnings are reduced.

In an action for benefits, the court’s standard of review will depend on whether the plan gives the administrator discretion to construe the plan’s terms. If the administrator is given no such discretion, the court will review the denial of benefits de novo. If the administrator is given such discretion, the court will apply the arbitrary and capricious standard of review, a deferential standard in which reversal is only appropriate if the lower court has failed to exercise sound and reasonable judgment. However, if an administrator with discretion is operating under a conflict of interest, the reviewing court will consider this as a factor in determining whether the administrator has abused its discretion.

III. RISK OF RELAPSE AS A CURRENT DISABILITY

Courts are conflicted as to whether the risk of relapse into substance abuse constitutes a “current” disability under LTD plans. As noted earlier, the First and Fourth Circuits disagree

---

56 Id.
57 Id.
59 Id. at 17.
60 RONALD J. COOKE, 3 ERISA PRACTICE AND LITIGATION § 8:46 (2013).
61 Id.
62 Id.
64 COOKE, supra note 60, § 8:46.
65 Compare, e.g., Colby v. Union Sec. Ins. Co., 705 F.3d 58 (1st Cir. 2013) (requiring LTD benefits), with Stanford v. Continental Cas. Co., 514
on this issue. The Fourth Circuit contends that the risk of relapse involves a choice component and is not a continuous disability, while the First Circuit asserts that the risk may be so severe as to render an individual “currently” disabled. However, such disagreement is not confined to the First and Fourth Circuits. Below is an overview of cases addressing this important question.

A. Cases Upholding the Denial of Benefits


Robert Stanford worked as a nurse anesthetist at Beaufort Memorial Hospital in South Carolina when he became addicted to Fentanyl, an anesthetic used in his practice. After completing a twenty-eight-day inpatient treatment program, he returned to work only to relapse two months later. He then began a ninety-day inpatient treatment program and filed for LTD benefits pursuant to his employer’s disability plan, administered by insurer Continental Casualty Company.

The insurer initially granted Mr. Stanford’s request for LTD benefits. However, after Mr. Stanford spent several months in recovery, the insurer terminated his benefits, citing a lack of medical evidence that he was functionally unable to perform “the material and substantial duties of [his] regular occupation.”

F.3d 354 (4th Cir. 2008) (upholding LTD benefit denial).

66 Stanford, 514 F.3d at 358 (“Whether [an addict] succumbs to that temptation remains his choice; the heart-attack prone doctor has no such choice.”).

67 Colby, 705 F.3d at 60 (“[A] risk of relapse into substance dependence . . . can swell to so significant a level so as to constitute a current disability.”).

68 However, the First and Fourth Circuits are the only federal appeals courts to have addressed this question.


70 Stanford, 455 F. Supp. 2d at 439.

71 Id. at 440; Stanford, 514 F.3d at 364.

72 Stanford, 455 F. Supp. 2d at 440.

73 Id.

74 Id.
Stanford appealed the insurer’s decision to terminate his benefits. Along with the appeal, he submitted a letter from his treating physician, which stated that he could not return to work as a nurse anesthetist because he should not be subjected to controlled substances and because the effects of his treatment medication could put patients at risk. After the insurer denied the appeal, Mr. Stanford filed suit in the United States District Court for the Eastern District of North Carolina alleging wrongful termination of benefits. The court granted summary judgment in favor of the insurer, stating that the risk of relapse did not render Mr. Stanford disabled because he was not “continuously unable” to perform his duties.

The Fourth Circuit affirmed the district court. It took a narrow view of what constituted a “mental impairment” under the insurer’s ERISA-governed plan and stated that while Mr. Stanford could not return to his old job, he was nevertheless “physically and mentally capable of performing that job and countless other jobs.” It further argued that addiction is a choice: “[w]hether [an addict] succumbs to that temptation remains his choice; the heart-attack prone doctor has no such choice.” The court therefore upheld the insurer’s determination that “Stanford no longer suffered from physical or mental impairments as a result of his drug use or his recovery, [and] the fact that he remained an addict did not [prevent him from performing] the material and substantial duties of [his] regular occupation.”

However, the Stanford court was sharply divided. Judge J. Harvie Wilkinson wrote an impassioned dissent, describing the majority’s conclusion as “an uncommonly harsh result.” He argued that the majority’s holding rested on two “abstractions”
not grounded in law. The majority’s first “abstraction” was that a disability plan was not required to cover “potential risk of relapse.” According to Judge Wilkinson, the majority's rejection of “potential risk of relapse” as a current impairment appeared to exclude all serious medical conditions that could make performing one’s job “unreasonably dangerous” because an individual could technically perform a job function at grave medical risk. According to Judge Wilkinson, such an exclusion contravened “a basic tenet of insurance law that an insured is disabled when the activity in question would aggravate a serious condition affecting the insured’s health.” The second “abstraction” was the majority’s assertion “that for disability purposes, a physical condition such as a heart attack... is different from the risk of relapse into drug use.” Judge Wilkinson stated that the majority’s attempt to distinguish these conditions was insufficient, as it rested on “moral considerations” that were not the court’s to make.

Judge Wilkinson also argued that the majority’s position was unsupported by the plan’s plain language. He noted that the plan covered “mental impairments” severe enough that one is “unable to perform the material and substantial duties of [his] regular occupation.” He defined “mental impairments” more broadly than the majority and pointed out that the plan defined “mental impairments” according to the American Psychiatric Association’s diagnostic manual, which devotes an entire section to substance-related disorders, including addiction.

Judge Wilkinson also made strong public policy arguments...
against the majority’s holding. 93 He pointed out that the insurer’s requirement that Mr. Stanford relapse in order to obtain disability benefits would not only create, as the majority acknowledged, a “perverse-incentive structure” by only paying benefits upon relapse, but would “thwart the very purpose for which disability plans exist: to help people overcome medical adversity if possible, and otherwise to cope with it.” 94 Because he did not believe that the risk of relapse could be categorically excluded from coverage, he argued that the proper inquiry as to whether a condition constitutes a current disability is “fact-intensive” and should focus on the likelihood and severity of the risk. 95 Judge Wilkinson concluded that Mr. Stanford’s prior relapses and the extensive medical evidence indicating that his risk of relapse was severe rendered him “currently” disabled. 96

2. Allen v. Minnesota Life Insurance Co. 97

Allen v. Minnesota Life Insurance Co, while not based on an ERISA claim, involves facts similar to Stanford. 98 In Allen, Dr. Robert Lee Allen, an anesthesiologist who practiced in Virginia, brought a breach of contract claim against his disability insurer claiming wrongful termination of his benefits. 99 Dr. Allen had been employed at Anesthesia Associates of Hampton for only a month when he began abusing Fentanyl. 100 However, less than three months after entering an inpatient treatment program, he was discharged with a favorable prognosis for recovery provided that he adhere to a prescribed treatment plan. 101 Nevertheless, the Virginia Board of Medicine suspended Dr.

93 Id. at 362–63.
94 Id. at 362.
95 Id. at 364.
96 Id. at 364–65.
98 Id. at 1378–81. Allen’s plan was not ERISA-governed because he purchased it individually, not through his employer.
99 Id. at 1378.
100 Id. at 1379.
101 Id.
Allen’s license.\textsuperscript{102} However, it stayed the suspension provided that Dr. Allen confine his medical practice to a Board-approved residency or fellowship.\textsuperscript{103} Dr. Allen eventually completed a residency in internal medicine and the Board reinstated his license to practice medicine on unrestricted status.\textsuperscript{104} Two months into his subsequent employment as an internist at Fayette Medical, his insurer notified him that it would discontinue his benefit payments.\textsuperscript{105} Although Dr. Allen was successfully reemployed, he was not engaged in his “regular occupation,” so he argued that he was entitled to continued benefits.\textsuperscript{106} However, the court ruled that he was not “\textit{unable} to engage in [his] regular occupation.”\textsuperscript{107} The court therefore upheld the insurer’s denial because Dr. Allen “suffer[ed] from no physical or mechanical limitations on his ability to practice anesthesiology.”\textsuperscript{108} It further determined that even though Dr. Allen’s treating physician testified that he should not return to practicing anesthesiology, the physician based his opinion on “future potentialities” only, not on any present disability.\textsuperscript{109}

The court also based its decision on its determination that Dr.

\textsuperscript{102} Id. at 1380.

\textsuperscript{103} Id. Dr. Allen commenced a residency in internal medicine shortly thereafter, although he returned to inpatient treatment due to concerns that he was violating Board-imposed restrictions on his practice. However, undisputed evidence indicated that he had not abused Fentanyl or alcohol since his initial treatment. Id.

\textsuperscript{104} Id. at 1380–81.

\textsuperscript{105} Id. at 1381.

\textsuperscript{106} Id. at 1383.

\textsuperscript{107} Id. (emphasis added). The plan defined “disability” as follows: “You have a disability if, because of continuing sickness or injury, you (1) are under the regular, reasonable, and customary care of a physician; and (2) are unable to engage in your regular occupation,” provided that “you are not earning more than 30\% of your prior average earned income from your regular occupation.” Id. at 1378–79.

\textsuperscript{108} Id. at 1383. While the \textit{Colby} court determined that the anesthesiologist’s “current occupation” was that of a physician, the \textit{Allen} court defined “current occupation” more specifically to mean anesthesiologist. \textit{See generally} Utz, supra note 27 (discussing a possible circuit split over the definition of “own occupation” under ERISA-governed LTD benefit plans).

\textsuperscript{109} Allen, 216 F. Supp. 2d at 1383.
Allen’s disability was not “uninterrupted,” as the plan required.\textsuperscript{110} This conclusion was based partially on the testimony of Dr. Allen’s treating physician, who provided testimony about his anesthesiologist patients generally, and stated that in most cases, he recommends a return to the field.\textsuperscript{111} The physician also gave testimony specific to Dr. Allen, and opined that Dr. Allen should avoid returning to anesthesiology because of the likelihood of relapse.\textsuperscript{112} The court evidently gave more weight to the general testimony than the testimony specific to Dr. Allen.\textsuperscript{113} It interpreted the treating physician’s claim—that he recommends most of his patients return to anesthesiology—as an indication “that drug addiction does not itself disable someone from practicing in that field.”\textsuperscript{114} It also noted Dr. Allen’s sobriety period and a lack of evidence “that he would inevitably regress.”\textsuperscript{115} The court determined that Dr. Allen had no “existing impediment” to his ability to practice anesthesiology and upheld the denial of Dr. Allen’s benefit payments.\textsuperscript{116}

3. Price v. Disability RMS\textsuperscript{117}

While anesthesiologists are more likely than other physicians to abuse drugs, a stressful work environment and easy access to potent drugs contribute to addiction among other physicians as well. Dr. Howard Price had worked as a urologist and surgeon at Milford-Whitinsville Regional Hospital in Massachusetts when he was forced to stop work because he began abusing opioids.\textsuperscript{118} However, Dr. Price’s insurer denied his claim for LTD benefits because he had not used opioids during the policy’s two-year

\textsuperscript{110} Id. at 1384. The plan required a “continuing disability,” which the court interpreted to mean an “uninterrupted” disability. \textit{Id.} at 1383.

\textsuperscript{111} \textit{Id.}

\textsuperscript{112} \textit{Id.} at 1383–84.

\textsuperscript{113} \textit{Id.}

\textsuperscript{114} \textit{Id.} at 1383.

\textsuperscript{115} \textit{Id.} at 1384.

\textsuperscript{116} \textit{Id.}


\textsuperscript{118} \textit{Id.} at *1.
coverage period, which began when Dr. Price stopped working.\footnote{119 Id. at *17–18.} After the insurer denied his two subsequent appeals, Dr. Price brought an ERISA action, claiming his depression, anxiety, and risk of relapse prevented him from performing all of the material duties of his occupation.\footnote{120 Id. at *1.}

In upholding the insurer’s denial of LTD benefits, the court placed significant emphasis on a lack of individualized evidence and Dr. Price’s continued “functional capacity.”\footnote{121 Id. at *18. As used in the opinion, “functional capacity” is simply the ability to practice medicine. \textit{See id. at *6}.} The court noted that the letters written by Dr. Price’s substance abuse counselor spoke only in general terms and did not make specific references to Dr. Price’s ability to function.\footnote{122 Id. at *19 (“[The doctor’s letter] did not relate the described symptoms in any persuasive way to Dr. Price’s functional capacity. How poor [was his concentration]? Did his poor concentration prevent him from performing his duties? [W]hat must be shown is that the illness caused a loss in functional capacity, and that is what was missing.”).} The court further noted that the counselor’s delineation of the disability period included several weeks during which Dr. Price was still practicing, which further illustrated the generality of the counselor’s testimony and its failure to illustrate a “functional impairment.”\footnote{123 Id.}

\textit{Price} is readily distinguishable from \textit{Stanford} and \textit{Allen}. Since the court found that Dr. Price did not have a severe risk of relapse, it never reached the question of whether a risk of relapse could be so severe as to constitute a current disability.\footnote{124 Id. at *21–22.} Also, unlike the \textit{Allen} court, the \textit{Price} court considered testimony specific to Dr. Price in making its determination that Dr. Price’s risk of relapse did not constitute a current disability.\footnote{Id.} Because of these differences, it is unclear whether the court’s conclusion would have been different had Dr. Price’s risk of relapse been more severe.

As the above cases reveal, courts upholding denial of benefits
generally place significant emphasis on an addict’s lack of “functional impairment.” Because the risk of relapse does not necessarily cause a continuous physical inability to perform one’s occupation, these courts do not view the risk of relapse as a current disability.

B. Cases Enforcing Continued Benefits

1. Colby v. Union Security Insurance Co.\textsuperscript{126}

The First Circuit’s \textit{Colby} decision rested on the court’s determination that a risk of relapse, while not necessarily a functional impairment, could be so serious as to constitute a “current disability” under an ERISA plan.\textsuperscript{127} The opinion cited medical testimony on behalf of Dr. Colby and determined that she faced a very significant risk of relapse following her departure from inpatient treatment.\textsuperscript{128} The court noted that the insurer could have possibly “limit[ed] the period of disability by arguing that this risk progressively diminished over the 36-month period,”\textsuperscript{129} but instead “took a categorical approach, steadfastly maintaining that risk of relapse, whatever the degree, could not constitute a current disability under the plan.”\textsuperscript{130} The award of a full three years of benefits therefore “flowed naturally from [the insurer’s] all-or-nothing defense of the case.”\textsuperscript{131}

The First Circuit also relied on a number of policy grounds in reaching its conclusion.\textsuperscript{132} For example, the court noted that denying benefits to those in recovery while providing them to those actively using the drug would create “a perverse incentive.”\textsuperscript{133} In addition, denying benefits to those in recovery would encourage claimants to return to work immediately upon leaving inpatient treatment, which could put their lives and their

\textsuperscript{126} Colby v. Union Sec. Ins. Co., 705 F.3d 58 (1st Cir. 2013).
\textsuperscript{127} \textit{Id.} at 60.
\textsuperscript{128} \textit{Id.} at 64.
\textsuperscript{129} \textit{Id.}
\textsuperscript{130} \textit{Id.} For an overview of the facts of this case, see \textit{supra} Part I.
\textsuperscript{131} \textit{Id.} at 68.
\textsuperscript{132} \textit{Id.} at 66–67.
\textsuperscript{133} \textit{Id.} at 66.
patients’ lives at risk. Finally, such a policy would defeat the very purpose of a disability plan, which is to help people overcome or otherwise cope with medical issues.

However, the court also emphasized the narrowness of its holding. As noted above, though the court held that Dr. Colby was entitled to LTD benefits, the insurer’s all-or-nothing approach helped the court reach that conclusion. The court suggested that the insurer might have had more success had it argued for a gradual benefit decrease over the 36-month period. Furthermore, the court pointed out that the insurer could have written into the policy an exclusion for risk of relapse, but it chose not to. Therefore, it ruled that the insurer acted arbitrarily and capriciously in denying LTD benefits to Dr. Colby.


Dr. Ronald Kufner, an anesthesiologist who suffered from alcohol and opioid dependence, brought an ERISA claim against his insurer similar to the claim brought in Colby. Dr. Kufner’s substance abuse issues forced him to stop work and undergo detoxification and other treatments, which included a week in the hospital followed by several months in a residential treatment program. Dr. Kufner received short-term disability benefits for thirteen weeks, and his insurer granted his subsequent request for

\[134 Id.\]
\[135 Id. at 67.\]
\[136 Id.\]
\[137 Id.\]
\[138 Id. Judge Wilkinson expressed doubt that such an exclusion would be permissible: “Since I do not think risk of addictive relapse and other medical risk can categorically be excluded from coverage, . . . .” Stanford v. Continental Cas. Co., 514 F.3d 354, 364 (4th Cir. 2008) (Wilkinson, J., dissenting).\]
\[139 Colby, 705 F.3d at 67.\]
\[141 Id. at 787.\]
\[142 Id. at 788–89.\]
After several months, however, the insurer terminated Dr. Kufner’s benefits since he had increased his work hours and had not experienced a relapse.\textsuperscript{144} Dr. Kufner nevertheless maintained that he remained disabled. While anesthesiologists typically work 70 to 80 hours per week, Dr. Kufner’s hours were limited to 40 to 50 per week by orders from his treating physician, who determined that his hours had to be reduced because a stressful work environment was a major factor contributing to his substance abuse problems.\textsuperscript{145} The treating physician further restricted him from handling or dispensing opioid analgesics.\textsuperscript{146} Dr. Kufner contended that because of these restrictions, his benefit payments should have continued.\textsuperscript{147}

The court determined, largely on public policy grounds, that the insurer abused its discretion in discontinuing Dr. Kufner’s LTD benefits.\textsuperscript{148} The court criticized the insurer’s decision to cut benefits in spite of “overwhelming medical evidence supporting a contrary decision.”\textsuperscript{149} It further pointed out the perversity of the insurer’s policy, which would force Dr. Kufner to work to the point of relapse at which point he would again be eligible for benefits.\textsuperscript{150} The court described this policy as one of “benefits Russian roulette” that put Dr. Kufner’s “career and his patients’ lives at risk.”\textsuperscript{151} The court explained that because anesthesiology is incredibly complex and a crucial part of surgery, the insurer’s

\textsuperscript{143} Id. at 789.
\textsuperscript{144} Id.
\textsuperscript{145} Id. at 794.
\textsuperscript{146} Id.
\textsuperscript{147} Id. at 787.
\textsuperscript{148} Id. at 796.
\textsuperscript{149} Id. This evidence included a letter from his treating physician stating that he should avoid on-call duty and not work more than 40 hours per week, another letter from the treating physician saying he could work up to 50 hours per week but that he remained at risk of relapse, and a letter from his treating psychiatrist stating that Dr. Kufner could not return to his previous level of employment. Id. at 793–94.
\textsuperscript{150} Id. at 796.
\textsuperscript{151} Id.
denial of benefits constituted a “breach of the public trust.” 152

The court also based its holding on the insurer’s ERISA-imposed obligation to discharge its duties “solely in the interests of the participants and beneficiaries.” 153 Those obligations hold insurers to “higher-than-marketplace quality standards” and require that “administrators provide a full and fair review of claim denials.” 154 According to the court, the insurer relied on “conclusory ‘peer review’ opinions” by doctors it retained rather than the extensive medical evidence and treatment records indicating that Dr. Kufner was unable to return to his previous level of employment. 155 The court concluded that the insurer’s determination was thus based on financial self-interest and pointed to Dr. Kufner’s entitlement to the plan’s maximum allowable benefits as further support for this conclusion. 156

The above case law makes clear that whether a court will uphold LTD benefit payments to a recovering addict will depend largely upon how broadly or narrowly the court construes “current disability.” Courts denying benefits commonly interpret the phrase in a strictly literal sense, at least with regard to recovering addicts. 157 In that vein, they are more likely to view an addict’s relapse into drug abuse as the choice of an otherwise healthy person. 158 On the other hand, courts ruling that benefits must be provided generally view addiction as a current disability—essentially, an ongoing condition. 159

152 Id.
153 Id. at 796–97.
154 Id. at 797 (quoting Metro. Life Ins. Co. v. Glenn, 552 U.S. 105, 115 (2008)) (internal quotation marks omitted).
155 Id.
156 Id.
157 See Stanford v. Continental Cas. Co., 514 F.3d 354, 363 (4th Cir. 2008) (Wilkinson, J., dissenting) (describing the majority’s inconsistency in denying benefits to Mr. Stanford when it would likely provide them to an individual capable of lifting heavy objects but only at risk of a serious injury).
158 See supra Part III.A.1.
159 Colby v. Union Sec. Ins. Co., 705 F.3d 58 (1st Cir. 2013) (“In our view, a risk of relapse into substance abuse—like risk of relapse into cardiac distress or risk of relapse into orthopedic complications—can swell to so significant a level as to constitute a current disability.”).
IV. THE PSYCHOLOGY OF ADDICTION

The preceding section demonstrated how a court’s understanding of addiction can affect the result of a case. If a court views an addict as one who is not functionally impaired yet chooses to use drugs, it will likely deny benefits. On the other hand, if a court views an addict as one who suffers from an ongoing, current disability, it will likely require that benefits be paid.

The following section places these differing views of addiction in the context of recent psychology literature.

A. Basics of Addiction and Environmental Factors that Precipitate Relapse

Recent psychology literature is at odds with the Fourth Circuit’s contention that addiction is a choice. The New England Journal of Medicine describes drug addiction as a “chronic, relapsing disorder in which compulsive drug-seeking and drug-taking behavior persists despite serious negative consequences.” While outdated but long-held views often see addiction as a moral failure or lack of willpower, recent neurobiological research indicates that drug addiction is in fact a brain disease. Drug addiction also has a strong genetic component: one study estimated that genetic factors are

---


162 Friedman, supra note 160, at 35 (citing Stephen J. Morse, Medicine and Morals, Craving and Compulsion, 39 SUBSTANCE ABUSE & MISUSE 437, 438–39 (2004) (arguing that addicts should, to some degree, be “held responsible for addiction-related behavior, such as seeking and using drugs”)).

163 Friedman, supra note 160, at 35.
responsible for approximately half of addiction vulnerability.\textsuperscript{164} Other research compared drug addiction to atherosclerosis, type 2 diabetes, and hypertension by noting common characteristics such as incurability, the importance of genetic risk factors, the influence of lifestyle choices, and the frequency of relapse.\textsuperscript{165} Finally, one study described drug addiction as a “chronic relapsing disorder” due to similar rates of treatment adherence and relapse when compared to type 2 diabetes, hypertension, and asthma.\textsuperscript{166} As with these other chronic illnesses, the majority of recovering addicts experience relapse, often after periods of significant improvement.\textsuperscript{167}

Despite the similarities between addiction and other chronic ailments, insurance companies place much greater limitations upon benefits for recovering addicts.\textsuperscript{168} Researchers have attributed this discrepancy to deeply-held biases and a lack of positive humanitarian feelings toward addicted individuals,\textsuperscript{169} which often lead to stigmatization and incarceration rather than proper treatment.\textsuperscript{170} The reasons for this lack of “positive humanitarian attitudes” toward addicts are undoubtedly complex, but likely explanations include the history of drug illegalization and the illegal drug trade, as well as a lack of understanding of addiction science.

Drug addiction has biological effects on the human body that are not easily overcome. Addiction triggers learning mechanisms\textsuperscript{171} and induces chemical and anatomical changes in

\begin{itemize}
\item \textsuperscript{164} Chuan-Yun Li et al., Genes and (Common) Pathways Underlying Drug Addiction, 4 PLoS COMPUTATIONAL BIOLOGY 28, 28 (2008).
\item \textsuperscript{165} Friedman, \textit{supra} note 160, at 36.
\item \textsuperscript{166} Sellman, \textit{supra} note 160, at 8.
\item \textit{See id.}
\item \textsuperscript{168} Friedman, \textit{supra} note 160, at 36.
\item \textsuperscript{169} Id.; Sellman, \textit{supra} note 160, at 8.
\item \textsuperscript{170} Friedman, \textit{supra} note 160, at 37.
\item \textsuperscript{171} A learning mechanism is, as its name implies, a way that the brain incorporates past experiences to apply them to future situations. Such mechanisms can include, for example, trial and error comparisons between an expected reward and an actual reward, and a model-based mechanism in which the brain makes predictions about an environment and then adapts that predictive model based on new experiences. \textit{See} Rick Nauert, \textit{Brain Images Reveal How Learning Strategies Work}, PSYCH CENT. NEWS (June 3, 2010),
\end{itemize}
the brain.\textsuperscript{172} Importantly, these changes are not quickly undone, even during abstinence, and are likely to be a significant factor contributing to relapse.\textsuperscript{173} In fact, these drug-induced changes may take many months or even years to reverse themselves.\textsuperscript{174} In addition, brain damage associated with drug addiction may harm parts of the brain responsible for making long-term decisions, such as those maximizing long-term welfare over short-term pleasure.\textsuperscript{175} Drug abuse can therefore lead to abnormal functioning in parts of the brain that would normally control compulsive behavior. These physical changes in a person’s brain can thus reduce an individual’s ability to resist a drug when exposed to it.\textsuperscript{176}

Unsurprisingly then, the risk of relapse is one of the most significant problems in treatment even among individuals who have sustained a prolonged abstinence period.\textsuperscript{177} Laboratory experiments on both humans and animals indicate that primary triggers of relapse include exposure to cues associated with previous drug taking, exposure to stressors, and exposure to the drug itself.\textsuperscript{178} Other evidence indicates that these factors do not necessarily operate independently. For example, one study on rats found that the most potent factors in precipitating relapse after both short and long periods of abstinence were exposure to a brief period of stress and exposure to the drug itself.\textsuperscript{179} It further found that exposure to the drug itself increases the effect of exposure to


\textsuperscript{172} Friedman, \textit{supra} note 160, at 35.

\textsuperscript{173} \textit{Id.}

\textsuperscript{174} \textit{Id.} at 36.

\textsuperscript{175} \textit{Id.}

\textsuperscript{176} Nora Volkow & Ting-Kai Li, \textit{Drug Addiction: The Neurobiology of Behavior Gone Awry}, 5 \textit{NATURE REVIEWS} 963, 965 (2004).


drug-related cues. Another study found that exposure to stressful events can similarly exacerbate the impact of exposure to drug-related cues on drug-seeking behavior, and vice versa. These factors are examined in further detail below.

1. Stress

Exposure to a stressful environment is a significant risk factor contributing to drug addiction relapse in humans. In one study, opiate-addicted individuals who were shown “stress related imagery” experienced heightened drug cravings. Researchers found similar results in cocaine-addicted individuals, for whom stress-induced hypothalamic-pituitary-adrenal axis responses predicted future drug use quantities. Animal research likewise indicates increased drug cravings in response to stress. Researchers studying relapse behavior found that opiate-addicted animals that experience foot shock (a small electrical shock to the foot to induce stress) are more likely to exhibit drug-seeking

180 Id.
182 Robyn M. Brown & Andrew Lawrence, Neurochemistry Underlying Relapse to Opiate Seeking Behavior, 34 NEUROCHEMICAL RES. 1876, 1879 (2009).
183 Id. This “stress-related imagery” was based on the participants’ descriptions of recent stressful personal events. After viewing the imagery, participants rated how vividly they could imagine the scenario, the extent of their opioid craving, and how anxious they felt. See Scott M. Hyman et al., Stress and Drug-Cue-Induced Craving in Opioid-Dependent Individuals in Naltrexone Treatment, 15 EXPERIMENTAL & CLINICAL PSYCHOPHARMACOLOGY 134.
behavior. This demonstrates a strong correlation between stressful experiences and heightened drug cravings.

Additionally, research shows that a stressful environment can actually increase a drug’s pleasurable effect. Stress does this by “priming” the brain’s reward pathways, meaning it increases the drugs’ efficacy and thus encourages the addict’s continued use. Such research is supported by clinical studies of drug abusers and alcoholics, in which subjects frequently cite stress as a reason for relapse. Other research indicates that a history of drug abuse can make individuals more sensitive to stressful events and thus more vulnerable to relapse. These relapse-inducing factors are clearly at play in the anesthesiology context, with its 70 to 80-hour workweeks and on-call duties.

2. Drug-related Cues

As a number of studies have shown, exposure to drug cues can precipitate relapse by increasing drug cravings. These drug cues fall into two general categories: “discrete” cues and “contextual” cues. A “discrete” cue is a physical object associated with drug-taking, such as drug paraphernalia. A “contextual” cue is one associated with a background setting, such as a room in which drugs have been previously used. The resultant heightened craving has been described as a form of

---

185 Brown & Lawrence, supra note 182, at 1881.
186 See The Reward Pathway Reinforces Behavior, supra note 11.
188 Id. at 351.
189 Id. at 351.
190 See generally Dan I. Lubman et al., Electrophysiological Evidence of the Motivational Salience of Drug Cues in Opiate Addiction, 37 PSYCHOL. MED. 1203 (2007); Shepard Siegel, Drug Tolerance, Drug Addiction, and Drug Anticipation, 14 CURRENT DIRECTIONS PSYCHOL. SCI. 296 (2005); Sinha, supra note 187, at 343.
191 Brown & Lawrence, supra note 182, at 1882.
192 Id.
193 Id.
Pavlovian conditioning\textsuperscript{194} in which drug-addicted organisms can experience withdrawal symptoms in the presence of the usual cues, even absent consumption.\textsuperscript{195} Furthermore, there is evidence that drug-related cues tend to capture the attention of drug addicts even when they are involved in an unrelated task.\textsuperscript{196} This suggests that the presence of drug cues may cause impulsive drug-seeking behavior.\textsuperscript{197}

3. Drug Exposure and its Effects on Stressors and Cues

Greater levels of past drug use often correlate with greater levels of cue sensitivity.\textsuperscript{198} In addition, repeated drug use increases the brain’s stimulus-award associations and forms a type of “addiction memory” that increases craving.\textsuperscript{199} Drug exposure can also change sensitivity to future drug exposure and stressors.\textsuperscript{200} To make matters worse for recovering addicts, stimuli that lead to this type of conditioned response maintain their effect well into abstinence.\textsuperscript{201} Even after “extinction training,”\textsuperscript{202} in which the ability of cues to provoke relapse is

\textsuperscript{194} Pavlovian (classical) conditioning is a learned association between stimuli. “[T]he subject learns to associate a previously unrelated neutral stimulus with another stimulus that reliably elicits some kind of reaction.” \textit{Pavlovian (Classical) Conditioning,} IND. UNIV., http://www.indiana.edu/~p1013447/dictionary/pavcond.htm (last visited Mar. 29, 2014).

\textsuperscript{195} Siegel, \textit{supra} note 190, at 297.

\textsuperscript{196} Lubman et al., \textit{supra} note 190, at 1208. In this study, the participants’ task was to press a button as quickly as possible whenever a white cup was displayed. \textit{Id.} at 1205.

\textsuperscript{197} \textit{Id.} at 1208.


\textsuperscript{200} Stewart, \textit{Psychological and Neural Mechanisms, supra} note 178, at 3153.

\textsuperscript{201} \textit{Id.}

\textsuperscript{202} “Extinction training” refers to a process that attempts to disassociate the drug cue from the drug itself. In animal experimentation this is done, for example, by training an animal to perform a task that results in the drug’s
reduced or eliminated, exposure to stress or the drug itself can rejuvenate the effects of the conditioned response to environmental cues. As the above studies show, researchers consistently identify (1) cues associated with previous drug taking; (2) exposure to stressors; and (3) exposure to the drug itself, as primary triggers of relapse into drug use. Courts should take a practical approach and keep these factors in mind when considering whether a recovering addict should be awarded LTD benefits.

B. Additional Factor: Genetics

Genetics also plays a role in drug addiction. Alcoholism has been shown to have a strong genetic component, and recent research has provided evidence that drug addiction is also a heritable disorder. It is estimated that genetic factors are responsible for forty to sixty percent of drug addiction vulnerability, with environmental factors responsible for the remainder. Animal research indicates the heritability of drug addiction at 0.4 for hallucinogens, 0.7 for cocaine, and slightly above 0.5 for alcohol. As a result of such studies, addiction has come to be regarded as a “complex genetic disease.”

The specific genes involved in addiction are unknown, but recent data indicate a relationship between drug addiction and the genes that encode dopamine receptors. Specifically, a study of 2,364 current opiate abusers or dependents indicated that the administration and then performing “extinction training,” in which the previous task no longer provides the drug. Id. at 3148.

---

203 Id. at 3153–54.
205 Sellman, supra note 160, at 7; Vetulani, supra note 204, at 313.
206 See Chuan-Yun Li et al., supra note 164, at 28.
207 Sellman, supra note 160, at 7. “Heritability” is an estimate of the genetic component of a trait, and ranges from zero to one. David Goldman et al., The Genetics of Addictions: Uncovering the Genes, 6 NATURE REVIEWS: GENETICS 521, 522 (2005).
208 Sellman, supra note 160, at 7.
209 Gorwood et al., supra note 160, at 803.
dopamine D2 receptor bears a highly significant link to the risk of opiate addiction.\textsuperscript{210} Dopamine release is necessary for brain “reward,” and all addictive drugs produce enhanced dopamine levels.\textsuperscript{211} This process not only “hijacks” the system normally used to experience the rewarding effects of natural survival functions, such as eating, but creates a lasting effect that promotes further use of the substance.\textsuperscript{212}

As mentioned, however, more research is needed to determine the specific genes involved in drug addiction susceptibility.\textsuperscript{213} Mapping of the human genome at the beginning of the century stirred hopes of isolating a handful of genes primarily affecting drug addiction.\textsuperscript{214} These hopes have not yet been realized, and researchers are still examining hundreds of enormously complex, linked, and variant genes.\textsuperscript{215} Despite these challenges, the concept of addiction as an interaction of genetic and environmental factors is now the “dominant paradigm” over the traditional view of drug abuse as an exercise of free will.\textsuperscript{216}

Taken as a whole, this research indicates that drug relapse is anything but a choice. Instead, it provides strong support for the view that addiction and relapse are the products of genetics, stress, and external stimuli, including the drug itself. Notably, these factors are often unavoidable because they are a result of genetics or are inherent in the addict’s occupation.

\textsuperscript{210} \textit{Id.} at 810. “Abuse” is the recurrent use of drugs despite adverse consequences. “Dependence” is another word for addiction, and manifests itself through symptoms such as heightened tolerance and withdrawal symptoms. \textit{See U.S. DEPT. OF HEALTH & HUMAN SERVS., APPENDIX E: SUBSTANCE USE, ABUSE, DEPENDENCE CONTINUUM, AND PRINCIPLES OF EFFECTIVE TREATMENT, available at} http://www.ncsacw.samhsa.gov/files/SAFERR_AppendixE.pdf. The cases cited in this Note involve both abusers and dependents.

\textsuperscript{211} Feltenstein & See, \textit{supra} note 177, at 265.

\textsuperscript{212} \textit{Id.}

\textsuperscript{213} \textit{See Gorwood et al., supra} note 160, at 803.

\textsuperscript{214} Sellman, \textit{supra} note 160, at 7.

\textsuperscript{215} \textit{Id.}

\textsuperscript{216} \textit{Id.}
V. ANALYSIS

A. Why Stanford is Flawed

The Fourth Circuit’s Stanford decision is flawed for five reasons. First, Stanford contravenes ERISA’s underlying purpose. Second, it is at odds with the current understanding of addiction science. Third, it fails to distinguish the risk of relapse from other chronic ailments and thus fails to show why it should be treated differently than those ailments. Fourth, it runs counter to recently enacted legislation on mental health and addiction. Finally, it disregards strong public policy arguments supporting a contrary decision. For these reasons, courts should follow the First Circuit’s approach outlined in Colby and regard the risk of relapse into substance abuse as a current disability.

1. The Fourth Circuit’s Stanford Decision is Contrary to ERISA’s Purpose and Supreme Court Precedent Regarding the Interpretation of ERISA

Congress made its purpose clear when it passed ERISA. Its stated goal was to “protect interstate commerce and employee benefit plans and their beneficiaries . . . by establishing standards of conduct, responsibility and obligation for fiduciaries of employee benefit plans . . . .” As the Supreme Court noted in Firestone Tire & Rubber Co. v. Bruch, “ERISA abounds with the language of trust law,” and requires that plan administrators, as fiduciaries, uphold “certain principles developed in the evolution of the law of trusts.” When ERISA administrators violate their fiduciary duties, ERISA allows policyholders to bring a cause of action against them. In Firestone, the Court referred to ERISA section 1104, which states “a fiduciary shall discharge his duties with respect to a plan solely in the interest of the participants and beneficiaries” and “for the exclusive purpose of providing

---

218 Id.
220 Id.
benefits to participants and their beneficiaries.” The *Kufner* court explained that ERISA imposes “higher than marketplace quality standards on insurers.” The implication of this requirement is that insurers must sometimes interpret ERISA benefit plans in a way that does not maximize profitability. Insurers who wish to avoid covering particular conditions must write their plans in a way that clearly circumscribes their obligations.

This conception of ERISA is difficult to square with the Fourth Circuit’s contention that Mr. Stanford was not entitled to disability benefits unless he was actively abusing a drug. The court argued that Mr. Stanford’s inability to return to his former job as a nurse anesthetist was not a result of a physical or mental impairment but rather “the result of a license limitation and the prudence of employers.” The court’s narrow understanding of “mental impairment,” which excluded addiction, is in stark contrast to the plain language of ERISA. As the statute states, fiduciaries are to discharge their duties “solely in the interest of participants and beneficiaries” and “for the exclusive purpose of providing benefits to the participants and their beneficiaries.” This language indicates that addicts should be entitled to benefits during the recovery period. While administrators have some discretion in deciding whether a particular condition constitutes a current disability, administrators (and the courts reviewing their decisions) may not ignore the statute’s plain text. The insurer’s decision to deny benefits to Mr. Stanford unless he relapsed, upheld by the Fourth Circuit, was clearly not “solely” in his interest and was thus contrary to the statute’s plain language. *Stanford* also failed to properly account for the conflict of interest that resulted from the insurer’s dual role as the evaluator and payer of claims. In *Firestone*, the Court explained that ERISA plan administrators often operate under a conflict of

---

221 *Id.* (construing 29 U.S.C. § 1104(a)(1) and § 1104(a)(1)(A)(i) (1974)).
224 *Id.*
interest and that reviewing courts should therefore consider this as a factor in reviewing benefit denials. The Fourth Circuit determined that it was to review the insurer’s determination under a “modified abuse of discretion” standard: this standard required it to “reduce [its] deference only to the degree necessary to neutralize any untoward influence resulting from the insurer’s conflict of interest, as shown in the record.” Because Mr. Stanford did not demonstrate a conflict of interest, the court did not reduce its deference.

Several months after Stanford was decided, the Supreme Court clarified the “conflict of interest” addressed in Firestone. In Metropolitan Life Insurance Co. v. Glenn, Supreme Court explicitly stated that an administrator, which both evaluates claims and make payments, operates under a conflict of interest (which the plaintiff need not demonstrate). It further stated that the significance of this element is fact-specific. The Court noted that while ERISA’s trust law basis requires deference to the fiduciary’s determination, courts must take this conflict of interest into account.

While the Fourth Circuit later acknowledged in Champion v. Black & Decker that Glenn would have required it to weigh this conflict as a factor despite Mr. Stanford’s failure to demonstrate a conflict, it is unlikely that this weighing would have changed the

---

226 Firestone Tire & Rubber Co. v. Bruch, 489 U.S. 101, 115 (1989); see also Metro. Life Ins. Co. v. Glenn, 554 U.S. 105, 108 (2008) (“Often the entity that administers the plan, such as an employer or an insurance company, both determines whether an employee is eligible for benefits and pays benefits out of its own pocket. We here decide that this dual role creates a conflict of interest; that a reviewing court should consider that conflict as a factor in determining whether the plan administrator has abused its discretion in denying benefits; and that the significance of the factor will depend upon the circumstances of the particular case.” (citation omitted)).


228 Id. at 359.

229 See Metro. Life Ins. Co., 554 U.S. at 118.

230 Id. at 108.

231 See id. at 115.

result.\textsuperscript{233} In \textit{Stanford}, the court explained that a plaintiff must produce evidence that an administrator’s decision was motivated by a conflict of interest.\textsuperscript{234} Since Mr. Stanford failed to produce evidence of this motivation, the court stated that a decision to overturn the denial would be based upon the mere existence of a conflict, which would eliminate deference entirely.\textsuperscript{235} However, the court indicated that the result would have been the same even if Mr. Stanford had demonstrated a conflict: “[w]e cannot say that [the insurer’s] conclusion is unreasonable, even in light of [its] conflict of interest as insurer and administrator of the benefit plan . . . .”\textsuperscript{236} The court held that the insurer’s interpretation of the plan, that the plan did not cover the “potential risk of relapse,” was reasonable whether or not a conflict existed.

\textit{Stanford} therefore rested primarily upon the premise that the “risk of relapse” is not a “current disability.” Such an interpretation is inconsistent with an insurer’s requirement to “discharge [its] duties with respect to a plan solely in the interest of the participants and beneficiaries.”\textsuperscript{237}

2. Recent Psychology Research Further Undermines Stanford

A fundamental problem with the Fourth Circuit’s ruling in \textit{Stanford} is that it is based on the discredited notion that an addict’s decision to use drugs is the result of choice, and not of disease.\textsuperscript{238} According to the Fourth Circuit, “[w]hether [a recovering addict] succumbs to [the temptation to use drugs]
remains his choice; the heart-attack prone doctor has no such choice.”

This notion is contrary to current psychology research indicating that relapses can occur well into abstinence because of lasting physical changes to the brain that result in a loss of control over drug use. Recent evidence strongly undermines the notion that continued drug use is a choice. With this research in mind, it is clear that the court’s reasoning in Stanford is based on the flawed notion that addiction is a choice, not a disease with a strong genetic component.

In Stanford, the insurer argued that an addict’s decision to use drugs is a choice by defining “choice” in extremely narrow terms. The insurer contrasted an addict to “a patient with an unacceptably high susceptibility to suffering from a heart attack” and declared that the patient “cannot avoid such heart attack by choosing not to have it.” It is true that a recovering addict could presumably encounter a situation in which drugs are readily available, yet decide not to use them. In this sense, he has a “choice” that one who is susceptible to heart attacks does not. However, this “choice” evaporates when the addict with a genetic predisposition and a physically altered brain is placed in a situation in which drugs are readily available. His decision to use drugs in such circumstances seems, if anything, less of a “choice” than a heart attack-prone patient’s decision not to exercise or to eat fatty foods. The insurer’s interpretation of the word “choice” is thus extremely narrow and unfairly applied to recovering addicts.

Even if we do accept the Fourth Circuit’s notion of addiction as a choice, many other chronic medical conditions (for which

\(^{239}\) Stanford, 514 F.3d at 358.

\(^{240}\) See Friedman, supra note 160, at 35.

\(^{241}\) See generally Friedman, supra note 160; see also Gorwood et al., supra note 160 (describing drug dependence as a “chronic, relapsing disorder”); Sellman, supra note 160 (describing addiction as a “complex genetic disease”).

\(^{242}\) Stanford, 514 F.3d at 358.

benefits are generally provided) are also the result of choice.\footnote{See Friedman, supra note 160, at 36.} For example, atherosclerosis, type 2 diabetes, and hypertension are all chronic conditions that are partially the result of voluntary behavior, such as diet.\footnote{Id.} Unsurprisingly, the treatment for each of these conditions often involves voluntary lifestyle changes.\footnote{See id.} An addict’s “choice” to use drugs is not easily distinguished from the lifestyle choices that contribute to these ailments, so to treat addiction differently on this basis is simply unjust.

One could perhaps argue that an addict’s initial decision to use drugs was a choice, and the lasting changes that the drugs caused to his brain only occurred as a result of this initial choice. Leaving addiction’s strong genetic component aside, this argument fails to distinguish addiction for the same reasons described in the preceding paragraph. One could argue that the individual suffering from atherosclerosis, type 2 diabetes, or hypertension only developed his condition as a result of his initial unhealthy lifestyle choices, and that the continued risk of a heart attack, for example, is a result of those earlier choices.

As argued above, psychology research has demonstrated that drug addiction is a genetic disease, and future courts should consider this in making their decisions. Courts often rely on psychology research to support their holdings in areas where the law is not settled, and the Supreme Court has done so in some landmark decisions.\footnote{See, e.g., Miller v. Alabama, 132 S. Ct. 2455, 2464 (2012); Brown v. Board of Ed., 347 U.S. 483, 494 (1954).} Most notably, in Brown v. Board of Education, the Court cited psychology research indicating that “[s]egregation of white and colored children in public schools has a detrimental effect upon the colored children. The impact is greater when it has the sanction of the law . . . .”\footnote{Brown, 347 U.S. at 494.} More recently, in Miller v. Alabama, the Court held that the Eighth Amendment prohibits mandatory life imprisonment without parole for those who committed crimes prior to age eighteen.\footnote{Miller, 132 S. Ct. at 2460.} In support of its position, the Court stated “that developments in
psychology and brain science continue to show fundamental differences between juvenile and adult minds... in parts of the brain involved in behavior control.” 250  Stanford was based on the discredited notion that an addict’s use of drugs is a “choice” and courts should look to current research to support decisions that recognize addiction as a genetic disease, and thus treat it as a “current disability.” 251 Scientific research is particularly useful on this issue because the law remains unsettled.

 Stanford is further flawed because the court denied LTD benefits to an individual for whom returning to work would place his health and life, as well as the health and lives of his patients, at tremendous risk. 252 An overview of current addiction research provides overwhelming evidence that the greatest risk factors in precipitating drug relapse are (1) stress; (2) exposure to the drug itself; and (3) environmental cues. 253 In light of these factors, it is difficult to imagine a set of circumstances better able to precipitate relapse than the placement of an anesthesiologist back into a hospital setting where he previously succumbed to opioid addiction. The high-pressure hospital setting, combined with long hours and easily obtainable drugs of choice, makes relapse all too likely.

3. The Fourth Circuit Failed to Distinguish the Risk of Addictive Relapse from the Risk of Relapse of Other Chronic Ailments

The Stanford majority determined that the risk of relapse into drug use was fundamentally different from the risk of relapse into other chronic ailments. 254 According to the court,

250 Id. at 2464 (internal quotation marks omitted).
251 See generally sources cited supra note 160 and accompanying text.
253 See Stewart, Psychological and Neural Mechanisms, supra note 178, at 3153; see also Liu & Weiss, supra note 181, at 7856 (describing stress and conditioned responses to drug cues as “critical factors in relapse to drug use”); Stewart, Pathways to Relapse, supra note 179, at 125 (describing re-exposure to the drug and exposure to stress as the two most important factors in reinstating drug-seeking behavior).
254 Stanford, 514 F.3d at 358.
[a] doctor with a heart condition who enters a high stress environment . . . “risks relapse” in the sense that the performance of his job duties may cause a heart attack. But an anesthetist with a drug addiction who enters an environment where drugs are readily available “risks relapse” only in the sense that the ready availability of drugs increases his temptation to resume his drug use. Whether he succumbs to that temptation remains his choice. The heart-attack prone doctor has no such choice.255

In his dissent, Judge Wilkinson was harshly critical of this attempt to distinguish the risk of relapse from other chronic ailments.256 As noted in Part III.A, Judge Wilkinson described the majority’s attempt to distinguish drug addiction as “legally ungrounded.”257 He said their attempt was based on moral and medical considerations that the court had no authority to make when the plan “put[] addiction squarely on all fours with other impairments.”258 The court’s failure to distinguish addiction from other ailments, and yet still deny benefits to Mr. Stanford, indicates that the ongoing stigmatization of drug addiction played a role in the court’s decision.

The majority’s attempt to cast disability as a “reward for sobriety,” but only in the addiction context, is similarly unpersuasive.259 In ruling for the insurer, the court acknowledged that its denial of benefits to Mr. Stanford created a perverse incentive by denying benefits to those in recovery while providing them to those who relapse.260 Nevertheless, the court argued that such reasoning assumed that disability was a “reward for sobriety” when, in fact, the reward for sobriety was “the creation of innumerable opportunities that were closed to Stanford as long

255 Id.
256 Id. at 363 (Wilkinson, J., dissenting).
257 Id.
258 Id.
259 See id. at 359.
260 Stanford, 514 F.3d at 359.
as he continued to use drugs.” It is unclear why the Fourth Circuit apparently confined this logic to recovering addicts. The court’s logic seems to imply that, like the recovering addict, the heart attack-prone doctor should not be entitled to benefits because his reward for adopting a healthier lifestyle is the “innumerable opportunities that were closed to” him before he changed his ways. Despite this obvious inconsistency, the court suggested that a heart attack-prone doctor should be entitled to benefits.

The Stanford majority’s contention that Mr. Stanford could work “countless other jobs” and therefore should not be entitled to benefits is similarly flawed. First, as Judge Wilkinson argued in his dissent, the plan’s plain language defined disability as an inability “to perform the material and substantial duties of your regular occupation.” Even with this plain language issue aside, the situation the majority describes is not unique to individuals recovering from drug addiction. For example, in Evans v. UnumProvident Corp., the court held that an insurer’s denial of LTD benefits to a plaintiff who suffered from a form of epilepsy was arbitrary and capricious. In making its determination, the court noted that while the plaintiff was capable of performing sedentary work, she was still disabled because the stressful nature of her work contributed to her recurrent seizures. Presumably, the plaintiff was capable of performing other, less stressful jobs, but this fact did not render her ineligible for LTD benefits. In this sense, her condition was no different from that of an anesthesiologist who is physically capable of performing other jobs, yet cannot return to anesthesiology because the stressful nature of the job contributes to relapse. In either case, the individual lacks a functional impairment that renders him unable to physically perform some type of occupation, yet LTD benefits will still be provided; there is no

---

261 Id.
262 See id. at 359–60.
263 Id. at 362 (Wilkinson, J., dissenting) (emphasis in original).
265 Id. at 869.
266 Id. at 879–80.
compelling reason to treat the two conditions differently.

4. Recent Policy Enactments Support the View that Addiction Should be Treated Like Other Ailments

The notion that addiction should be treated like other ailments is supported by congressional legislation. On November 8, 2013, Secretary of the Department of Health and Human Services Kathleen Sebelius announced regulations that would enforce the Mental Health Parity and Addiction Equity Act of 2008 (Parity Act) and extend its reach to those receiving coverage under the Affordable Care Act. Congress enacted the Parity Act in order to prevent health plans from discriminating against individuals with mental and substance abuse disorders by requiring that the plans’ standards for those conditions be comparable to those for other medical conditions. The Parity Act prohibits “limits on the frequency of treatment, number of visits, days of coverage or other similar limits on the scope or duration of treatment” when such conditions are not imposed upon coverage for other medical conditions. It also prohibits plans from imposing more stringent financial requirements upon those suffering from mental health issues or addiction. This means that plans cannot impose different “deductibles, copayments, coinsurance, [or] out-of-pocket expenses” on mental health and addiction treatment.


270 Weber, supra note 267, at 207–08.

271 Id. at 210.

272 Id.

273 Id. It is important to note that the Parity Act does not require plans to cover mental health or substance disorder benefits. It only requires that when such benefits are provided, they must be on equal footing with medical and
The Parity Act’s legislative history reveals that Congress intended to curtail the widespread practice of insurer discrimination against those with mental illness and substance-related disorders. The Committee on Ways and Means issued a report stating: “[t]he Committee believes that the discrimination that exists under many group health plans with respect to mental health and substance-related disorder benefits must be prohibited. Diseases of the mind should be afforded the same treatment as diseases of the body.” The Committee went on to describe addiction and mental health disorders as “the only disorders that have been systematically and unfairly excluded from equal coverage.” The Parity Act and its legislative history demonstrate Congress’ intent to fight arbitrary and discriminatory treatment of those suffering from addiction or mental illness.

In Stanford, the insurer had apparently discriminated against Mr. Stanford because his impairment was “mental.” The plan required benefits once the claimant established an “injury or sickness caus[ing] physical or mental impairment to such a degree of severity that [he is] . . . continuously unable to perform the material and substantial duties of [his] regular occupation.” The insurer explained that Mr. Stanford did not suffer “a physical or mental impairment as a result of his drug use or recovery” and that being an addict did not render him unable to perform the material duties of his occupation. The insurer’s narrow understanding of “mental impairment” was unjustified, and the Fourth Circuit should not have upheld it.

5. Providing LTD Benefits to Recovering Addicts is Good Public Policy

There are compelling public policy arguments for providing LTD benefits to recovering drug addicts. The Committee on surgical benefits. See 29 U.S.C. § 1185(a)–(b) (2012).

275 Id.
276 Id.
277 Id.
279 Id.
Ways and Means’ reasons for passing the Parity Act are equally applicable to the “addiction as a current disability” debate. The Committee cited a 2006 study that described the prevalence of mental and substance abuse-related disorders, which affected nearly a quarter of the U.S. population and cost more than $300 billion annually. A recent study found that 22.2 million Americans suffered from substance abuse or dependence in 2012, a number that had remained stable over the prior decade.

As mentioned, a number of courts have also made compelling public policy arguments for treating the risk of relapse into drug addiction as a “current disability.” Even in Stanford, the majority acknowledged that its denial of benefits to Mr. Stanford “create[d] a somewhat troubling—some might say perverse—incentive structure: an addict who continues to abuse drugs will be entitled to long-term benefits, but upon choosing sobriety will lose those benefits unless he again begins to use drugs.” As Judge Wilkinson argued in his dissent, “[f]orcing Stanford to relapse into addiction or lose his benefits would thwart the very purpose for which disability plans exist: to help people overcome medical adversity if possible, and otherwise to cope with it.” Few would argue that a bartender who was forced to leave work as a result of alcoholism should be compelled to return to work during recovery because his benefits would discontinue. Disability plans should not force addicted individuals to choose between losing benefit payments on the one hand and relapsing on the other.

The Kufner court also considered public policy implications in

---

281 Id.
285 Id. at 362 (Wilkinson, J., dissenting).
its decision to treat the risk of relapse as a current disability.\textsuperscript{286} It described the insurer’s implication that the plaintiff could return to work until he suffered a relapse as “untenable given the serious risk this poses to public health and safety, which the Court considers an additional factor weighing against defendant’s benefits determination.”\textsuperscript{287} As described in Part III.B, the court labeled this risk “a form of ‘benefits Russian roulette’ with plaintiff’s career and his patients’ lives at risk.”\textsuperscript{288} The court further described the insurer’s position as “tantamount to a breach of the public trust” and clearly contrary to its duties under ERISA.\textsuperscript{289} While danger to the public is particularly acute in the anesthesiology context, it is also a serious concern in other areas as well. For example, a relapsed crane operator, air traffic controller, or train engineer could pose tremendous risks to the public. None should be forced to choose between relapsing and losing benefits.

\textit{B. A Middle Ground Between Colby and Stanford?}

In \textit{Colby}, the First Circuit proposed an untenable middle ground between its own holding and that in \textit{Stanford}. According to the court, on remand, the insurer could have examined whether the risk of relapse decreased over time and, if it did, argued for a corresponding benefit reduction.\textsuperscript{290} Instead, the insurer took a categorical approach and argued that any risk of relapse, no matter how severe, did not constitute a current disability under the plan.\textsuperscript{291} The result of this all-or-nothing approach was the court’s award of a full thirty-six months of benefits\textsuperscript{292}

This argument, that the risk of relapse progressively diminishes over time, is unsupported by current psychology research.\textsuperscript{293} As noted previously, substance abuse causes lasting

\begin{footnotesize}
\begin{itemize}
\item \textsuperscript{286} Kufner, 595 F. Supp. 2d at 796.
\item \textsuperscript{287} Id.
\item \textsuperscript{288} Id.
\item \textsuperscript{289} Id. at 796–97.
\item \textsuperscript{290} Colby v. Union Sec. Ins. Co., 705 F.3d 58, 64 (1st Cir. 2013).
\item \textsuperscript{291} Id.
\item \textsuperscript{292} Id. at 68.
\item \textsuperscript{293} Stewart, \textit{Psychological and Neural Mechanisms}, supra note 178, at
\end{itemize}
\end{footnotesize}
changes in the brain, and these changes play a significant role in precipitating relapse. Research also demonstrates that “exposure to a drug can initiate neurochemical changes with enduring molecular and anatomical consequences that affect subsequent responses to events that induce relapse.” These changes “continue to manifest themselves well into abstinence and may be a cause of the relapses into compulsive drug use that can occur long after the drug has been cleared from the body.” While a recent study of methamphetamine-dependent individuals found some evidence that impulsive decision making decreases over time, the study also found evidence that cue-induced cravings increase over time, and therefore concluded that the risk of relapse does not decline with abstinence.

The First Circuit’s proposed alternative argument finds little support in recent psychology research.

VI. CONCLUSION

Under LTD plans, there is no principled reason to differentiate the risk of addictive relapse from other medical impairments. Accordingly, other courts should follow the First Circuit’s Colby decision. The Fourth Circuit’s Stanford decision contravenes Supreme Court precedent, which had illustrated the high standards that ERISA places upon plan administrators. These standards require administrators to discharge their duties “solely in the interest of the participants and beneficiaries,” not in their own financial self-interest, as the insurer apparently did in Stanford. Stanford also contravenes current psychology

3147.

294 Id.
295 Id.
296 Friedman, supra note 160, at 35.
298 With the exception of Allen, this Note has focused on ERISA claims. While the arguments in supra Part V.A.1. are applicable only to ERISA claims, the other arguments are equally applicable to non-ERISA claims.
research, which shows that addiction is not a “choice” but a disease that physically changes the brain in ways that last well into abstinence.\textsuperscript{300} Thus, a recovering addict who is not actively using drugs is still “currently disabled” and should be entitled to LTD benefits under ERISA-governed LTD benefit plans.

Furthermore, \textit{Stanford} could have dire ramifications. It incentivizes recovering addicts to return to work before they are ready. With regard to anesthesiology, this inhibits the recovery process by placing addicts into an environment that is extremely conducive to relapse due to high stress levels and easily accessible drugs. Patients are similarly put at risk because anesthesiology is a crucial and complex component of many medical procedures. The risks of \textit{Stanford}-like decisions are not limited to anesthesiology, but extend to any occupation that affects public safety and health.

The First Circuit’s \textit{Colby} decision, holding that the risk of relapse into drug addiction can be so severe as to constitute a current disability, avoids these potentially disastrous consequences. Further, \textit{Colby} holds true to the congressional intent behind ERISA and the Parity Act. The Parity Act reflects a larger societal trend that recognizes the devastating effects of addiction and sees it as a disease rather than a choice or lack of willpower. Additionally, society has increasingly come to recognize that the way to deal with the pervasive problem of drug addiction is not to stigmatize users or blame them for poor decision-making, but to treat their condition as a chronic ailment on par with any other. Courts should therefore follow the First Circuit’s lead and do their part to move society forward on this issue.

\textsuperscript{300} Friedman, \textit{supra} note 160, at 35.
BROOKLYN LAW SCHOOL FACULTY 2013-2014

Faculty Emeriti

RICHARD ALLAN, B.A., LL.B., Professor of Law Emeritus
JOSEPH CREA, B.A., J.D., LL.M, LL.D., Professor of Law Emeritus
MARY FALK, B.A., J.D, Associate Professor of Legal Writing Emerita
MARTIN HAUPTMAN, LL.B., LL.M., Professor of Law Emeritus
WILLIAM HELLERSTEIN, B.A., J.D., Professor of Law Emeritus
NORMAN POSER, A.B., LL.B., Professor of Law Emeritus
GARY SCHULTZ, B.A., J.D., Professor of Law Emeritus

Faculty

NICHOLAS ALLARD, B.A., B.A./MA., J.D., Joseph Crea Dean and Professor of Law
WILLIAM ARAIZA, B.A., M.S., J.D., Professor of Law
CAMERON W. ARNOLD, B.A., M.A., J.D., Instructor of Legal Writing
JONATHAN ASKIN, A.B., J.D., Associate Professor of Clinical Law
MREM BAIR, A.B., J.D., Associate Professor of Law
CHRISTOPHER BEAUCHAMP, B.A., M.Phil, Ph.D., Assistant Professor of Law
DEBRA BECHTEL, B.A., J.D., Associate Professor of Clinical Law
URSULA BENTLE, B.A., J.D., Professor of Law
ANITA BERNSTEIN, B.A., J.D., Anita and Stuart Subotnick Professor of Law
BRADLEY BORDEN, B.B.A., M.B.A., J.D., LL.M., Professor of Law
DANA BRAXMAN REISER, B.A., J.D., Vice Dean and Professor of Law
MICHAEL CAHILL, B.A., M.P.P., J.D., Vice Dean and Professor of Law
I. BENNETT CAPERS, B.A., J.D., Professor of Law
STACY CAPLOW, B.A., J.D., LL.M, Associate Dean for Professional Legal Education and Professor of Law
NEIL B. COHEN, S.B., J.D., Jeffrey D. Forchelli Professor of Law
STEVEN DEAN, B.A., J.D., Professor of Law
ROBIN EFFRON, B.A., J.D., Associate Professor of Law
ELIZABETH EJAF, B.A., M.A., Ph.D., Associate Professor of Legal Writing
MICHAEL FANTO, B.A., M.A, Ph.D., J.D., Professor of Law
RICHARD FARRELL, B.A., LL.B., Wilbur A. Levin Distinguished Service Professor of Law
LINDA FELDMAN, B.A., M.A., J.D., Assistant Professor of Legal Writing and Director, Academic Success Program
MARYELLEN FULLERTON, B.A., J.D., Professor of Law
MARTHA GARRISON, B.A., J.D., Suzanne J. and Norman Miles Professor of Law
MICHAEL GEBRER, B.A., J.D., Professor of Law
CYNTHIA GODSOE, A.B., J.D., Assistant Professor of Law
JOEL GORA, B.A., LL.B., Professor of Law
SUSAN GREENE, B.A./B.S., J.D., Instructor of Legal Writing
SUSAN HERMAN, B.A., J.D., Centennial Professor of Law
EDWARD JANGER, B.A., J.D., David M. Barse Professor of Law
BERYL JONES-WOODIN, B.A., J.D., Professor of Law
JOY KANWAR, B.A., J.D., M.S.EL, Instructor of Legal Writing
ROBERTA KARMEL, B.A., LL.B., Centennial Professor of Law
ADAM KOLBER A.B., J.D., Professor of Law
MINNA KOTKIN, A.B., J.D., Professor of Law
BAILEY KURLIN, B.S., J.D., Professor of Law
REBECCA KYSAR, B.A., J.D., Associate Professor of Law
BRIAN LEE, B.A., M.A., Ph.D., J.D., Associate Professor of Law
GREGG MACEY, B.A., M.A., J.D, Ph.D., Associate Professor of Law
ROSEANN MACKECHNIE, B.A., J.D., Instructor of Legal Writing
GARY MINDA, B.A., M.A., S.J.D., Professor of Law
SAMUEL MURUMBA, LL.B, LL.M., Ph.D., Professor of Law
MINOR MYERS, B.A., J.D., Associate Professor of Law
OGNA PETERSON, B.A., J.D., Instructor of Legal Writing
ARTHUR PINTO, A.B., J.D, Professor of Law
ROBERT PITLER, B.B.A., LL.B., LL.M., S.J.D., Professor of Law
KAREN PORTER, B.A., M.S., J.D., Associate Professor of Clinical Law
DAVID REISS, B.A., J.D., Professor of Law
JAYNE RESSLER, B.A./B.S., J.D., Associate Professor of Legal Writing
ELIZABETH SCHNEIDER, B.A., M.Sc., J.D., Rose L. Hoffer Professor of Law
JANET SINDER, A.B., J.D., M.S., Director of Library and Associate Professor of Law
LISA SMITH, B.A., J.D., Assistant Professor of Clinical Law
DAN SMULIAN, B.A., M.Phil., J.D., Associate Professor of Clinical Law
LAWRENCE SOLAN, B.A., Ph.D., J.D., Don Forchelli Professor of Law and Director of Graduate Education
WINNIE TAYLOR, B.A., J.D., LL.M., Professor of Law
NELSON TEBBE, A.B., J.D., Ph.D., Professor of Law
CARRIE TETCHER, B.A., J.D., Assistant Professor of Legal Writing
AARON TVERSKEI, B.S., J.D., Irwin and Jill Cohen Professor of Law
MARILYN WALTER, B.A., J.D., Professor of Law
JOAN G. WEXLER, B.S., M.A.T., J.D., Dean and President Emerita, Professor of Law
Visiting Faculty

HILARY ALLEN, B.A., LL.B., LL.M., Visiting Associate Professor of Law
JEFFREY ATIK, A.B., J.D., Ph.D., Dr. Jur., Visiting Professor of Law
EMILY BEKMAN, B.A., J.D., LL.M., Visiting Assistant Professor of Law
NEIL COHEN, B.A., J.D., LL.M., Diploma in Criminology, Visiting Professor of Law
LOUISE CROWLEY, B.C.L., LL.M., Ph.D., Visiting Professor of Law
STEPHAN LANDSMAN, B.A., J.D., Visiting Professor of Law
ROGER MICHALSKI, B.A., Ph.D., J.D., Visiting Assistant Professor of Law
MELISSA MORTAZAVI, B.A., J.D., Visiting Assistant Professor of Law
ANDREW NAPOLITANO, B.A., J.D., Distinguished Visiting Professor of Law
THANE PITTMAN, B.A., M.A., Ph.D., Visiting Professor of Law
GERALD SHARGEL, B.A., J.D., Practitioner-in-Residence
BRAD SNYDER, A.B., J.D., Visiting Associate Professor of Law
ALAN M. TRAMMEL, B.A., M.Sc., M.St., J.D., Visiting Assistant Professor of Law
SHONNA TRINCH, B.A., M.A., Ph.D., Visiting Professor of Law

Librarians

ROSEMARY H. CAMPAGNA, B.A., M.S.L.I.S., Reference/Government Documents Librarian
GILDA CHIU, B.A., M.L.S., Collection Development/Acquisitions Librarian
KATHY DARVIL, A.B., M.S.I., J.D., Access Services/Reference Librarian and Adjunct Assistant Professor of Law
JEAN J. DAVIS, B.A., J.D., M.S.L.I.S., Associate Librarian for International Law and Adjunct Professor of Law
SARA GRAS, B.S., J.D., M.S.L.I.S., Reference Librarian and Adjunct Assistant Professor of Law
LINDA HOLMES, B.A., M.S., Associate Librarian
HAROLD O’GRADY, B.A., J.D., M.L.S., Reference Librarian and Adjunct Professor of Law
JANET SINDER, A.B., J.D., M.S., Director of Library and Associate Professor of Law
HAINAN YU, B.A., M.S., Systems Librarian
BROOKLYN LAW SCHOOL BOARD OF TRUSTEES

STUART SUBOTNICK ’68
Chairman of the Board
President and Chief Executive Officer
Metromedia Company

HON. EDWARD R. KORMAN ’66
Vice Chairman of the Board
Senior Judge
United States District Court for the Eastern District of New York

BURNSIDE E. ANDERSON III ’76
Senior Corporate Counsel, Legal Division
Pfizer Inc. (Ret.)

FRANCIS J. AQUILA ’83
Sullivan & Cromwell LLP

HON. HERMAN BADILLO ’54

DAVID M. BARSE ’87
Chief Executive Officer
Third Avenue Management, LLC

FREDERICK COHEN ’67
Duane Morris LLP

JEFFREY J. FEIL ’73
Chief Executive Officer
The Feil Organization

MARTIN A. FISCHER ’64
West Center Associates

JEFFREY D. FORCHELLI ’69
Forchelli, Curto, Deegan, Schwartz, Mineo & Terrana, LLP

SHANNON M. HALEY ’08
Sullivan & Cromwell LLP

ROBERT M. KAUFMAN ’57
Proskauer Rose LLP

LEO KITTAY ’08
Fross Zelnick Lehrman & Zissu, P.C.

EILEEN T. NUGENT ’78
Skadden, Arps, Slate, Meagher & Flom LLP & Affiliates
FLORENCE SUBIN ’75
Subin Associates

STEVEN L. ZELKOWITZ
Chief Executive Officer
Sycamore Energy Consulting

Members Emeriti

ROBERT B. CATELL
Chairman
Advanced Energy Research and Technology Center (AERTC) at
New York State University at Stony Brook

FRANCES MARGOLIS FRIEDMAN ’39