PLAYING GOD: WHO SHOULD REGULATE EMBRYO RESEARCH?*

Baroness Ruth Deech**

It is an honor to be able to celebrate the Belfer family: Ira M. Belfer, member of the Class of 1933, who went on to become a distinguished practitioner of law and a major supporter of Brooklyn Law School; his son Myron, Professor of Psychiatry at Harvard Medical School and Senior Consultant in Child Mental Health for the World Health Organization. The careers and actions of father and son together embody generosity of spirit, a love of legal education, and the desire to put the advances of medicine to use in the interest of children all over the world. These qualities are highly appropriate to the subject that I am to address today.

The title of my lecture was inspired by a comment made to me in 2002 by a Member of Parliament (MP) when I gave evidence to the House of Commons Select Committee on Science and Technology, which was examining embryo research. The MP queried the decision of the Human Fertilisation and Embryology Authority (HFEA), of which I was then the chairman, to permit a couple to benefit from the new technique of embryo selection. Mr. and Mrs. Hashmi have a son called Zain, who suffers from a family-inherited condition of beta-thalassaemia. The Hashmis sought to have preimplantation genetic diagnosis (PGD) carried out on embryos that they would produce in order to be able to select one that would be free of the inherited genetic condition and which would also, if it developed into a baby, provide tissue that was compatible with Zain. Zain needed bone marrow from a compatible donor to save his life. I explained to the Committee of Members of Parliament that the HFEA had felt confident in making this decision within the law and that it was ethically sound. Moreover, I explained, speed was of the essence in such a situation. Little Zain suffered every day from the invasive treatment that he required and his mother was getting no younger. I knew that if a decision on an issue of this magnitude and sensitivity were to be left to legislators, not only would it take a very long time depending on the legislative timetable but it would inevitably attract media and constituent interest. Legislators would vote not only with the broader issue in mind but

---

* This lecture was delivered at Brooklyn Law School on October 10, 2006, for the Ira M. Belfer Lecture Series. In drafting this lecture, the author has drawn upon previous speeches and publications.  
** Former Chair of the U.K. Human Fertilization & Embryology Authority; Former Principal, St Anne's College, Oxford University; Independent Adjudicator for Higher Education (England & Wales); Member of the House of Lords.
also with politics at the forefront. When I appraised this dilemma by saying that the HFEA undertook these decisions according to law designed in order to protect legislators from having to make them, the response of the MP was: “Who do you think you are, playing God?” But “playing God” was precisely, in my view, the role assigned to the HFEA by Parliament itself. Moreover, “playing God” may mean either assuming to oneself the power to make decisions that no one on earth should be making; or it may mean doing one’s human best to act as a partner with God in improving the lot of mankind and, where necessary, attempting to perfect God’s creation.

It was on that occasion that I realized that even though issues of PGD and stem cell research are discussed as if they were ones purely of ethics, law, and science, the realm of assisted reproductive technology is a battleground fought over by legislators jealous of their power, desperate patients, clinicians and companies with considerable earnings, and religious pressure groups. I concluded that the only way to keep the peace is by comprehensive regulation by as neutral and expert a body of people as can be assembled. Any other solutions result in distortions and inconsistencies.

These political and financial aspects of embryo research regulation are masked by various ethical and religious approaches. After introducing the arguments about ethics and embryo research, I will unmask them by explaining the nature of stem cell research and examining the stances of Germany, the United States, Italy, and the United Kingdom and their internal inconsistencies. I will weigh up regulation and autonomy and propose a framework of legislative regulation. I will take as examples the new areas of embryo research where the HFEA took decisions under my

---

1. The tussle between religious forces and parliamentary ones in this field is made clear by the comment of Pope Benedict at Easter 2006: “There is a move to reinvent mankind, to modify the very grammar of life as willed by God . . . . [T]o take God’s place, without being God is insane arrogance, a risky and dangerous venture.” Pope Benedict XVI, Good Friday Prayer at the Colosseum in Rome (Apr. 14, 2006). The same opinion was voiced in England when the license for PGD offered to the clinic that treated the Hashmi family was challenged as illegal by CORE, a group fundamentally opposed to any form of interference with the embryo at any stage. R (on the Application of Quintavalle) v. Human Fertilisation & Embryology Auth., [2005] UKHL 28, [2005] 2 A.C. 561. CORE stated that the reason they challenged the HFEA in court over the use of PGD for tissue-typing purposes was not to obtain a pronouncement on whether the technique was right or wrong in itself, but to make the point that such a policy decision should be made by a democratically elected government, not by an “unelected” organization, namely the HFEA. The gesture was too late because in 2001–2002 both houses of the British Parliament voted strongly in favor of extending the remit of the HFEA to cover research on embryos for the purposes of treating and learning about serious disease.
chairmanship, techniques which have adapted established In Vitro Fertilization (IVF) procedures to lifesaving and life-altering purposes.

I. EMBRYOS

At the heart of the new ethical debate lie attitudes towards embryos and in particular their use for research. It is very well-known that it is regarded as morally wrong by some because they regard the embryo as human from the moment of fertilization, and therefore one should not take its life or bring it into existence simply in order to be destroyed. This, they say, is as true of the embryo as of children and adults. Indeed, one may take the argument a step further, because if one should not destroy something with the potential to become a human, this could equally apply to eggs, for eggs alone can be converted into embryos through the cloning technique. Up to 70% of natural embryos in the body never succeed in implanting in any case and are lost. The *reductio ad absurdum* of protection of the embryo is that natural intercourse should be avoided because that is bound to lead to the loss of some fertilized embryos. Alternatively and equally unrealistic, if each embryo is a possible life and should not be wasted, then because all fertile adults are capable of combining to produce embryos at every moment of the day, we should do so incessantly in order to avoid any possibility of wasted potential.

The Warnock Committee, whose 1984 Report laid down the foundations of regulation, reached a typically British compromise: in English law the embryo has been given special status but not absolute protection. In Britain, the embryo is treated as an entity deserving due attention, which means that it is to be used only if there is no alternative, that its use must be ruled by informed consent of the donors, that there are restrictions on exporting embryos or mixing them with non-human material, and that there is considerable recordkeeping to ensure that every single embryo in research is accounted for.

The embryo is defined in section 1 of the Human Fertilisation and Embryology Act 1990 (HFE Act) as a “live human embryo where fertilisation is complete.” The United Kingdom was the first country in the world to acknowledge as legal and then regulate therapeutic cloning or

---


3. This compromise has been upheld by the *House of Lords Select Committee Report on Stem Cell Research*, Report, 2002, H.L. 83-I, § 4.21.

cell nuclear replacement (CNR). This technique of artificially fertilizing an egg which has been enucleated was challenged in the courts.\(^5\)

The House of Lords confirmed that embryos created by CNR did come within the regulatory purview of the HFEA; they reached this conclusion by taking a purposive approach to the interpretation of the word “embryo.”\(^6\) This decision on the one hand upheld the democratic or parliamentary approach to this new field in that it supported the remit of the existing law, but it also enabled the regulators to extend their reach.

The position in Britain now is that a wide range of embryo research may be permitted under license from the HFEA. The extended list of purposes for which research involving embryos may be permitted is: promoting advances in the treatment of infertility; increasing knowledge about the causes of congenital disease and the causes of miscarriage; developing more effective techniques for contraception; developing methods for detecting the presence of gene or chromosome abnormalities before implantation; increasing knowledge about the development of embryos, knowledge about serious disease, or enabling such knowledge to be applied in developing treatment for serious disease.\(^7\)

A word of explanation about the science behind the arguments over the status of embryos. Two non-fertility avenues are being explored. One is PGD and the other is research on adult and embryonic stem cells, derived respectively from adult cells and embryos. The latter can further be divided into research on embryos created by uniting eggs and sperm, and research on embryos created by CNR. The latter technique involves obtaining an egg, which is subsequently enucleated. The removed nucleus is replaced by the nucleus of a cell obtained from an adult, possibly an adult who requires treatment or a volunteer. The new nucleus is fused to the enucleated egg by an electric shock. This apparently causes the egg to believe that it has been fertilized as if by sperm, and it starts to grow. The advantage of CNR is that the resulting embryo will be identical in every respect to the genetic makeup of the adult who donated the nucleus. Were stem cell lines to be derived from this embryo and new organs and tissues created, the adult patient would not reject the new organ, for it would be 100% compatible with his own body. This would end the current position whereby recipients of donated organs have to spend the

---

5. The grounds were that the statutory definition did not appear to include embryos created by processes that do not involve the union of egg and sperm. CNR does not involve sperm in the creation of the embryo. R (on the Application of Quintavalle) v Sec’y of State for Health [2003] UKHL 13, [2003] 2 A.C. 687.

6. Id.

rest of their lives on immuno-suppressive drugs to prevent their bodies rejecting “stranger” organs.

The legislative approval of extended research purposes in England was based on the fact that the HFEA has a good record in ensuring that clinics comply with the law and therefore an extension would not represent a slide down the slippery slope. By way of contrast, much harm has been done to the scientific community by the falsity of the claims of the Korean Professor Hwang, who claimed to have created thirty cloned embryos and eleven stem cell lines. In 2006, it was reported that Professor Hwang used eggs donated by junior team members and faked his research. There are two ethical issues here: the faked research and also the likelihood that pressure was applied to female members of the team to donate the gametes needed for the research.

Bearing in mind the fate that befell Professor Hwang and his team, there is a need to govern and regulate the way in which eggs are obtained. The HFEA is undertaking a public consultation on whether women should be permitted to donate eggs purely for use in medical research. It is already legal in Britain to donate eggs for the purpose of pregnancy on an altruistic basis. Indeed, it is not uncommon for people willingly to give up organs and bone marrow to help others; voluntary kidney donation is quite widespread. This seems to point to the ethical or pragmatic acceptability of the donation of eggs for medical research, even though there is risk to the donor and no direct benefit to herself. The drugs used can be dangerous and the generous payment of expenses to donors, or even the possibility of sale of eggs, as in the United States, for large sums, may tempt the needy.

II. STEM CELLS

Stem cells present the most exciting possibilities for the future, although as yet undeveloped. They are the basic components of the rest of the body and capable of providing new cells. They are found in embryos, in the fetus, the placenta and umbilical cord, and in parts of the body. When the embryo has grown to the eight-cell stage soon after fertilization, each of the eight is totipotent; that is, it could develop into any and every type of cell needed for the body. Some days later when stem cells are present in the inner cell mass they are still pluripotent, that is have the capacity to develop into most types of cell. Because of their ability to reproduce themselves and develop into other types of cells, stem cells offer the prospect of growing new tissue to repair parts of the body damaged by accident or ill health and to treat a wide range of diseases that have developed because the cells have degenerated, e.g., Alzheimer’s, Parkinson’s, and diabetes. If those treatments can be found, there would
be a lifelong cure and new regenerative tissue when needed. Stem cell research has been widely publicized and is seen as a good focus for financial investment and medical research. Adult stem cells, unlike embryonic ones, are allegedly not as useful for growth and research, although opinions differ and knowledge is limited so far. The problem with adult stem cells is that if they are derived from an adult with a disease, the likelihood is that the stem cell may reinforce that disease rather than cure it. Nevertheless, adult stem cells are of significance because work on them does not present the ethical problems associated with work on embryos, whether surplus to requirement after IVF or deliberately created. Most stem cell research is carried out on embryos surplus to requirement, and it must be ethically preferable to use them for beneficial research than to allow them to perish.

In Britain, work on stem cells derived from surplus or created embryos is legal and was recommended by the HFEA. It has to be approved by scientific peers and ethics committees and the license from the HFEA takes into account quality approval, justification, legal compliance, consent, and yearly accounting for each one used. This is strict but effective and gives researchers an area of safety and respectability in which to proceed. So far fifteen licenses have been granted.

The research license is granted only if the use of embryos is necessary and the aim is necessary or desirable for the treatment of infertility or serious disease. “Serious” is a fluid concept depending on time and place. As time brings reduced tolerance of ill health and greater expectations of relief, the band of diseases regarded as serious will expand. Under British regulation a sample of every stem cell line derived from a cell taken from an embryo or adult must be deposited in the U.K. Stem Cell Bank, which was set up in 2004 by the Medical Research Council and the Biotechnology and Biological Sciences Research Council. Research is only licensed if it cannot be done from existing stem cell lines in the Bank. The potential benefits are considerable, ranging from organs available for transplantation and ending the current shortage, to tailored drugs, discoveries about the genetic origins of diseases, and cures for infertility. Gametes might be derived from stem cells to provide reproduction for the infertile.

III. NATIONAL ATTITUDES TO STEM CELL RESEARCH

Despite these advantages there has been a diversity of response to this research in Europe and the United States. This is because of cultural, religious, and economic reasons specific to each country. In some, feelings run so high that they cannot reach a compromise position, for example,
using only surplus embryos. Belgium and Holland permit embryo research and have no legislation; Portugal and Germany forbid it.  

A. Germany

The contrasting ethical attitudes can be seen most clearly in the case of Germany. Abortion is permitted but embryo research is not, despite the constitutional guarantee of freedom of research. No egg may be fertilized except for the pregnancy of the donor and, as in Italy, there may be no initial creation of surplus embryos. This has the effect of preventing PGD. Nevertheless, stem cell lines may be imported into Germany provided they were derived before 2002 and subject to restrictions. In that year German law prohibited the derivation of new stem cell lines, even if not defined as embryos. As in the United States, there is a certain amount of ethical hypocrisy in Germany in that there is reliance on stem cell lines created outside the country. The policy is set in the context of human rights and a historical perspective, a determination not to repeat the mistakes of the past, to integrate the disabled, and to protect women and children.

B. Italy

Italy was until recently the most unregulated country in Europe. It was the place to which one went for the treatment of women over sixty, for attempts at cloning, sex and race selection, and embryo splitting. However, in 2004 a comprehensive law was introduced. Gamete donation is banned, as is egg freezing. The number of embryos that may be used per cycle is limited to three, no scientific research on embryos is permitted, and there can be no PGD. Couples seeking IVF must be of two different sexes, married or living together, and no posthumous treatment is allowed. The law amounts to a set of prohibitions rather than the construc-

8. The Council of Europe Convention on Human Rights and Biomedicine, article 18, provides that where the law allows for embryo research it must ensure adequate protection for the embryo and that the deliberate creation of embryos for research is prohibited. Council of Europe, Convention on Human Rights and Biomedicine art. 18, Apr. 4, 1997, 36 I.L.M. 817, Europ. T.S. No 164. Cloning is prohibited and this is reinforced by the Charter of Fundamental Rights of the European Union. Charter of Fundamental Rights of the European Union art. 3, Dec. 7, 2000, 40 I.L.M. 266. There are few ratifications. This is because in England and some other countries acceptance of the research has moved on well beyond the European declarations, affected as they are by religious susceptibilities and history.


tion of a general regulatory framework for the conduct of assisted reproduction. It penetrates a long way into the area of medical discretion, for example banning embryo freezing and limiting the number of embryos that may be transferred in one cycle, thereby reducing the success rate. The law goes even further than the report of the Dulbecco Commission that preceded it, which had recommended that stem cell research should be allowed on surplus embryos. The Italian law has been condemned by some European research organizations in that it greatly reduces the chances of an infertile woman bearing a baby, but it was not voted down by the electorate when they were given the opportunity. The new law is of course wholly consistent with the dominant Catholic religion of the country.

C. The United States

The position in the United States is perhaps the most troubling and inconsistent of all, and its results affect the whole world. The stance is derived from a combination of politics, business, religion, and aversion to federal control. Any general regulation or ethics directive is seen as undermining the doctor-patient relationship and imposing bureaucracy on medicine. Freedom of state action has led to rule by market forces, a general free for all which in turn leads to doctor-patient-baby conflicts of interest, abuses, and dangerous genetic “cures.” The regulation that has been attempted by professional bodies in the United States has proved ineffective because no consensus has been reached on the main issues and the guidelines are unenforceable at law. These issues are all bound up with the very sensitive American position on abortion and the tension over childbearing issues between religious forces and the constitutional rights to privacy and liberty. Any legislation that interfered with, for example, the “right” to clone could be open to constitutional review by the courts. Much as the United States needs federal regulation of embryo research, it is particularly difficult to achieve because of the Constitution, guarantees of state and personal autonomy, and the political/religious lobby.

American law prohibits the use of federal funds for the creation of a human embryo for research purposes. This means that private clinics and company laboratories are free to undertake research. In 2001, President Bush announced that federal funding may be used for research on existing stem cell lines, but no others, and set up the Council on Bioethics under the chairmanship of Leon Kass, a conservative thinker. The effect of the law is that no live embryo may be destroyed for research pur-

poses. It is believed that there are seventy-eight existing stem cell lines eligible for federal funding for research, of which twelve have become available to the United States so far and are regarded as unsuitable. Federal funds may be spent on adult non-embryonic stem cells, to which no constraints apply. There is little open and informed debate of the issues in the United States because there is no public accountability for embryonic research and no unified voice speaking for it, only presidential dik-tat. Private companies are doing considerable unsupervised research, which is profitable and hard to square with concerns over public health. The current position is as unethical as it can be: the public sector is obstructed while private work is unregulated and driven by commercial concerns. There should be no pretence that the Bush pronouncement of 2001 was ethical as claimed: it was one of political expedience prevailing over all the scientific and logical arguments pointing the other way.

There are ways around the ban. A single cell detached from an early embryo can be grown in culture to create stem cell lines, without destroying the embryo. It could theoretically be used to attempt pregnancy, even with one cell removed, as occurs after preimplantation genetic diagnosis, although this would be contrary to English law. Keeping the embryo alive would address the ethical concerns of many, but it is not as efficient as culturing the entire eight-cell embryo. This might be a way to circumvent the federal funding ban. There is pressure from pro-life groups to find alternatives to destroying embryos to grow stem cells. It is a counter-productive search, however, as many embryos are destroyed in pursuit of this one result. Recently, in British laboratories scientists have taken cells from dead embryos and grown them to the stage where stem cells might be extracted. The scientists used embryos that had died naturally during IVF treatment. The embryos in this experiment had stopped developing a few days after fertilization. Nevertheless, the question was immediately raised of how one would know that the embryos were dead, and whether there was something wrong with those embryos that caused the arrest of their development. Other ethicists however regarded this new process as akin to organ donation from dead patients, with no more concerns than surround that process.

The moral arguments have been turned on their heads however by the fact that very recently some states, led by New Jersey and California, have legislated for state funding for stem cell research. This may lead

to a movement of scientists from restrictive states to those that are more liberal. California Proposition 71\textsuperscript{14} of 2004 granted $3 billion funding to stem cell research. The measure included the establishment of the California Institute for Regenerative Medicine to regulate and oversee stem cell research. The Act has a preamble stating that half of California families have a member who has or will suffer from a medical condition that could be treated with stem cell therapies. The Act is designed to plug the gap in federal funding and to shift the emphasis of health care towards prevention rather than expensive cures. It will presumably give a boost to the prestigious Californian universities’ research programs. Human reproductive cloning remains forbidden but otherwise all types of research on all types of embryos may be carried out. Accountability is to be achieved by open meetings and annual reports to the public. The Independent Citizens’ Oversight Committee of twenty-nine members governs the California Institute for Regenerative Medicine and will include patient, university, and research representatives. The Act directs the Committee to establish standards concerning informed consent, controls on human research, the prohibition of compensation to donors, privacy laws, and time limits for obtaining cells (eight to twelve days after fertilization). There are similar laws in New Jersey\textsuperscript{15} and a Stem Cell Institute of New Jersey, with a grant of state funds, albeit very small compared to California.

There will soon be a patchwork of regulation across the United States and a scattering of states where funding for stem cell research will be provided. This will lead to scientific tourism, a reflection of the world situation. This is a further argument in favor of national control and a national regulatory body. This need is not met by the President’s Council on Bioethics, because it is allegedly composed of members chosen for their conservative views and appointed by the President, and it has no powers. My experience on the HFEA convinced me that to mix the pragmatic with the philosophical is the best way of going forward. There is nothing so invigorating for a committee containing ethicists as to be confronted with a case needing an immediate practical solution. Questions such as what to do when embryos are mixed up or lost, the use of embryos that are subject to litigation, the payment of donors in a situation of shortage, all force committee members to apply policy and principles and come up with a legal and humane solution.

\begin{footnotes}
\item[14] \textit{Cal. Const.} art. XXXV, § 1 (approved by the voters of California).
\end{footnotes}
IV. THE STRUCTURE OF REGULATION

On the basis of examination of the inconsistencies and weaknesses of various national situations, what may be concluded about the ideal regulatory framework? There may be comprehensive regulation or private rights and prohibitions, such as in Italy. There may be regulation or a free market as in the United States. There may be regulation by independent committee or by legislators, which is the question in the United Kingdom. Britain was extremely fortunate in the timing of its legislation, as far back as 1990. It enacted a comprehensive framework before many of today’s issues emerged and before positions became entrenched. Hence when new issues arose, such as stem cell research, there was already a framework within which to control them, and the public and Parliament were accustomed to it. Britain was also fortunate in that the statute was well-drafted and was flexible enough to cope well to the present day, although reforms are now being considered.\(^\text{16}\)

The most interesting analysis of national structures of regulation has been made by D.G. Jones and C.R. Towns.\(^\text{17}\) The authors describe four types of regulation of stem cell research.

The first is the prohibition of all human embryo research: Ireland, Austria, Norway, and Poland. The second is permission to use stem cell lines already in existence before a certain date: the United States and Germany. The third is to use stem cells only from embryos surplus to IVF requirements: Canada, Greece, Finland, Hungary, Netherlands, Taiwan, and Australia. The fourth is to allow also the creation of embryos specifically for research: the United Kingdom, Belgium, Israel, Singapore, Japan, South Korea, and Sweden.\(^\text{18}\)

The authors then point out the inconsistencies.\(^\text{19}\) In some countries where all embryo research is forbidden because the embryo is sacrosanct, IVF is allowed. IVF cannot normally be carried out without engendering a certain number of embryos that are surplus to requirements and are eventually destroyed. This is more unethical than using them for research purposes. Germany and Italy are at least internally consistent in that they do not allow any embryos to be created that will not immediately be used for pregnancy, avoiding the creation of extra ones. This is at a cost of some detriment to the woman’s health. Instead of banking extra embryos,

\(^{16}\) SCIENCE AND TECHNOLOGY COMMITTEE, HUMAN REPRODUCTIVE TECHNOLOGIES AND THE LAW, 2004–5, H.C. 7-1, at 175–89.


\(^{18}\) Id. at 1113–14.

\(^{19}\) Id. at 1114–16.
she will have to undergo ovarian stimulation by drugs repeatedly. Multiple births are more likely to occur because all the embryos will be transferred in one attempt rather than using only one or two and keeping the remainder. Multiple births are dangerous to the health of mother and baby and costly to society. States that ban all embryo research are also quite likely to take advantage of stem cell research in other countries. In Germany the import of stem cell lines from other countries where they have been derived from surplus embryos is allowed under strict conditions.  

If permission is given to use stem cells created before a certain date (the second category) they may be inappropriate for research and will eventually be too old. Those countries that prohibit the creation of new stem cell lines on the basis of the sanctity of the embryo are also likely to allow the destruction and creation of embryos for the purposes of IVF. The use of surplus embryos (the third category) does at least make a utilitarian use of embryos that are in existence and would otherwise perish. However, the creation of embryos for research that are specific to the patient is blocked. This position accepts that surplus embryos will be created and destroyed during IVF and therefore accepts that embryos are disposable, despite imposing restrictions on research. The fourth category, the most liberal one, is based on not accepting that every single embryo has human potential. It maximizes research opportunity and secures a diversity of embryos for research, including ones that will be compatible for the purposes of transplantation.

Another way of categorizing the different national attitudes to embryo research is to look at the contenders, the winners and losers. When no research is allowed at all or only on old stem cell lines, this is an advantage for the religious/political factions. If prohibition can be evaded by import or export of gametes then commerce benefits but not patients and domestic researchers. If research is allowed only on surplus embryos, this inhibits scientists. If the most liberal attitude is taken, all interested parties benefit except that legislators are denied the control that they seek.

The ethical positions are further muddied by the existence of “reproductive tourism.” If patients, researchers, and gametes may move from country to country in search of the desired facilities that may be legally available to them, as is permitted to Europeans by the Treaty of Rome,  

---

there is little point in restriction in a few countries, other than to send a signal. The ethics of the entire continent of Europe have to sink to the lowest point because the scientist who is unable to carry out embryo research in Italy may move to England; the would-be patient who prefers to have an anonymous sperm donor may leave England or Sweden and go to Belgium; and so on. This situation highlights the need for a liberal regime in the knowledge that those who have principled objections need not take advantage of what is going on, but that it is safer and cheaper to have treatment in the home country.

I conclude from this brief overview of the national positions that there is no substantive ethical or qualitative difference between research directed to stem cell derivation and the use of embryos for IVF and general research. If one activity deserves attention and regulation, so do the others. Stem cell research alone cannot be usefully singled out for regulation, and indeed the regulation of stem cell research on its own is not efficient.

V. THE HUMAN FERTILIZATION AND EMBRYOLOGY AUTHORITY OF THE UNITED KINGDOM

It is often asked by what moral right do the members of the HFEA pronounce on these issues. It is because it embodies the democratic compromise between strongly held views in society, reached by the HFE Act 1990. The Authority works within the Act to reconcile opposing views and point to a way forward, with public accountability. The HFEA licenses and monitors clinics that carry out IVF treatments, donor insemination, and embryo research; it strives to ensure that treatment and research are undertaken with respect for human life and responsibility towards the parties, recognizing the vulnerability of patients and the expensive nature of the treatments. One in six couples seek infertility treatment and there may be many more who are infertile but do not seek medical advice. Care should be taken not to exploit them.

The HFEA regulates the storage of gametes, registers information about donors, treatments, and children, safeguarding the biggest database of its kind in the world. It issues a Code of Practice to clinics; gives advice and information to patients, donors, clinics, and the government; and keeps new developments under review. It has staff and premises in London, and twenty members selected openly on merit after advertisement. It issues an annual report, holds open meetings, and keeps in touch with all elements of the relevant professions. Its overall aim is to ensure that public understanding and reassurance move at the same pace as the new developments that it licenses. In the sixteen years of its existence, the HFEA has overseen major steps forward in infertility treatment, ex-
tending from “simple” IVF into matters of convenience, lifesaving, and life alteration. The HFEA has faced several legal challenges, all but one of which have failed. The HFEA is not universally popular with scientists, who resent the extra paperwork surrounding permission to pursue relevant research. It has however been successful in blocking the slippery slope. As far as is known, no one has tried to keep an embryo in vitro for more than fourteen days from fertilization, or has cloned, because policing of the laboratories is part of the system.

Legal regulation in Britain has probably served to protect clinicians and scientists not only from legal action for malpractice (where they have followed the HFEA Code of Practice) but also to give them a shield against accusations of ethical malpractice, for they are acting within the parameters agreed by Parliament and the HFEA. These safeguards clearly do not apply where there has been a total breakdown in the clinic controls and the deliberate flouting of the criminal law, which has happened once or twice.

Regulation has disadvantages too, as I shall show, but in general the history of regulation shows that work on embryos might never have been permitted in Britain at all had it not been for the existence of, at first, voluntary professional self-regulation and, subsequently, statutory controls. It has progressed with reliability and in tandem with public and peer acceptability. Judged by those criteria, regulation in principle has been a successful move. In addressing issues relating to public fear of new technologies, family issues, safety, and extent, regulation has been more of a success than a failure. The HFEA has tried to achieve a judicial use of its powers, combined with a consultative and consensual ap-

21. E.g., HFEA, First PGD Guidelines, 1999, CE13/08/1999; Press Release, HFEA, HFEA to Permit Use of Frozen Eggs in Fertility Treatment (Jan. 25, 2000); 1994 HFEA ANN. REP. 18 (noting that the first ICSI licenses were granted in 1993).

proach to regulation, and it aims to secure confidence and respect from Parliament, the public, patient groups, and the professions. Parliament, having set the framework, can rely on the Authority for the day-to-day management of the many legal and ethical issues that have emerged from the field.

VI. REGULATION FOR ALL?

Is regulation a good thing for the whole world? It has certainly achieved a great deal in Britain without too much dissent. There are some deeper issues that need consideration however. How does regulation square with human rights and autonomy? What issues should be settled in the legislation and what left to the discretion of the regulating committee or the patients? There are no general criteria for resolving this but different answers are given in different systems. They are linked to the competition for power to which I alluded, between market forces, religious forces, politicians, drug companies, doctors, and patients. It is rare for these competing forces to be recognized and addressed. They tend to be disguised by ethical and legal discussions.

There are many ways to justify legislative regulation. One could start with the simple issue of safety. No new treatment which affects the bringing into being of a new life should be allowed to proceed unless there is as much evidence as can reasonably be obtained that there will be no harm to mother or baby. Human reproductive technology has tended to proceed under a precautionary principle, balancing the benefits to be obtained against the possible harm. Thus the introduction of intracytoplasmic sperm injection (ICSI) and frozen egg technology was delayed in Britain while assurances about their safety were sought. Safety bears on issues such as the insemination of women over sixty, cloning, posthumous births, and donor gametes. I regard it as legitimate to curb reproductive autonomy when its exercise unreasonably impacts on the independence of others or threatens harm. Multiple births with their attendant private and public costs, sex selection with its effects on existing children, and PGD all have a long-lasting impact on society and may place a burden on others. It follows that society may have a legitimate reason to control through the democratic process the choices that may be made by individuals with their doctors. It is right to grasp the nettle and

23. Rosamund Scott, Choosing Between Possible Lives: Legal and Ethical Issues in Preimplantation Genetic Diagnosis, 26 OXFORD J. OF LEGAL STUDIES 153, 175 (2006); see also RONALD DWORKIN, FREEDOM’S LAW: THE MORAL READING OF THE AMERICAN CONSTITUTION 104 (1996); ONORA O’NEILL, AUTONOMY AND TRUST IN BIOETHICS 58 (2002); JOHN A. ROBERTSON, CHILDREN OF CHOICE ch. 2 (1994); John Harris, Rights and Reproductive Choice, in THE FUTURE OF HUMAN REPRODUCTION 5, 34 (John Harris &
accept that reproductive autonomy can co-exist with regulation rather than leaving many profound issues to be decided, slowly, by individual court cases.

Regulation is also called for to control the market forces in this field, in the way that any big business is legitimately a target of regulatory control. Some of Britain’s most well-known clinicians are reputed to be amongst the wealthiest men in the country. Because IVF and embryo research are big business, it is not wise to leave regulation to the professional bodies in the field because they will have conflicts of interest. Where there is a nationalized health system, decisions have to be made about resource allocation because resources spent on IVF and embryology will have an impact on other areas of medicine and their effectiveness has to be assessed. Britain accepts that it is dangerous to leave personal choice in IVF and the new genetics to market forces. In the United States it is accepted that IVF is a billion-dollar business, and even gametes and surrogacy are for sale. There is an urgent need to control sale of gametes and embryos, like the sale of other organs, bearing in mind the dignity and vulnerability of individuals and the health and welfare of the potential baby.

Identity and recordkeeping are especially important in regulation, in case IVF children as adults are entitled to seek information about the identity of the donors. A register of data is also important as an epidemiological tool, searching for factors affecting the success of IVF, and presenting data on multiple births, birth defects, and success rates. Confidentiality of data should not be so strict that follow-up studies are ruled out. Of course a regulatory authority needs sufficient resources, without which the purpose is defeated and dangerous mistakes may be made. It

Søren Holm eds., 1998). The state interest was spelled out in the case of Dickson v. United Kingdom, App. No. 44362/04, Eur. Ct. H.R. (2006). In this case a prisoner was refused the right to artificial insemination of his wife while he was in prison. The argument that society allows children to be born to single persons in poor circumstances was not regarded as sufficient to entitle the prisoner to achieve something similar. This was because the state was being asked to become an “active accomplice and participant” in the conception. Id. (Bonello, J., concurring). “I believe a responsible state to be right to require of itself standards higher than those beyond its control in the free procreation market.” Id. In Israel, Larissa Trimbobler sought artificial insemination by her husband Yigal Amir, who is in prison for life; the couple were granted conjugal visits.

24. The characteristics of potential gamete donors are available on the Internet, although whether their descriptions match the reality is another issue. It has even been pointed out that in an unregulated market, an embryo could be split, and once one of the twins has been born and the other embryo frozen, the frozen one could be offered for sale knowing what its twin looked like. George J. Annas, Some Choice: Law, Medicine, and the Market 11 (1998).
needs to be answerable to the legislature and to give a full account of its workings in a way that the public can access and understand. Its members need to be appointed in a way that gives people without a professional interest a chance to be represented. My own experience was that a majority of lay members over professionals on the authority was a good thing. Expert evidence can always be acquired from outside the membership. In particular, it needs to take evidence on the developments that lie ahead so that it is not caught unawares when a sudden application for new treatment is made or a new technique is problematic. It will be expensive and litigation can never be entirely avoided because there will be statutory interpretation issues arising from new developments and the regulator is bound to upset both clinicians and patients at some stage by its decisions.

The country that is in most urgent need of regulation is the United States, because it is the most advanced scientific nation with an impact on all world science. Such federal law as there is appears to be concerned more with funding than with substance. There is no supervision of activities in IVF and embryo research; disasters are bound to happen. The private companies that do research may do well, but this activity is not necessarily undertaken with concern for public health and for the inequalities between the medically insured and the uninsured Americans. There is a lack of clarity about standards and achievements because the private work is undertaken in conditions of commercial confidentiality and competitiveness. It must surely be possible for the United States to establish a public oversight body and comprehensive legislation. The Council on Bioethics is no substitute; in fact it may provide false assurance because its powers are non-existent.

But what is the reality of experiencing and administering regulation? Are there disadvantages?

Regulation has to be financed, and it is the patients who bear the expense, which is passed on to them by the clinics. Relatively little free IVF treatment is undertaken by the U.K. National Health Service (NHS) and, even in those cases, the NHS has to absorb the extra cost represented by regulation, and therefore ultimately it reduces the resources available elsewhere in the hospital system. Regulation also engenders avoidance, not in the sense of law breaking, but in the practice of reproductive tourism, going abroad to obtain a treatment banned at home. It stimulates constant discussion, criticism, and demands for reform. There

are effective lobby groups and political opponents of any decision made by the regulators.

In approaching a new decision in the regulation of embryology, the following are the factors that in practice determine the outcome in a regulated environment.

First, the legal framework. Every regulatory decision has to be taken in the knowledge that it is likely to be challenged in the courts, either by a disappointed individual or by a pressure group, and that it is important to the regulator to succeed in the litigation. There is in Britain no overarching written constitution that might ensure success on the basis of a higher principle of freedom. However, our detailed regulatory legislation is boxed in or, some would say, made porous, by recent human rights and European Treaty provisions. The application of the Human Rights Act 1998 to legislation enacted before that date has tilted interpretation towards individual rights and liberty in applying legislation that was already carefully drafted to balance public health demands and individual desires. The European Treaty principles of freedom of movement of goods and services, and the right to seek medical treatment abroad, may in the last analysis undo all the careful regulatory constraints applied in the home country.

Second, money, to fight and to enforce. The regulatory authority needs to be sufficiently well-financed. Litigation cannot be brought to enforce the measures of a regulatory authority unless the authority has the funds available to fight a case all the way to the House of Lords (our Supreme Court). It is often the case that a popular litigant will be better funded by a newspaper to which he or she has sold his or her story, than the government authority in opposition. The HFEA was occasionally chastised for mistakes that were in truth unavoidable as long as the technology is in the hand of humans. A well-known example was the birth of black twins to a white couple, where the wrong sperm had accidentally been used in the clinic treatment. The consequent reprimands and demands for ever tighter regulation and inspection were made even while the Authority was being denied the resources that would make seven-day-a-week supervision possible.

Third, the power of the media in a small country where everyone reads the same newspapers and in general watches the same TV news. Newspaper and television coverage is often wholly inaccurate, but if there is a good human story, the more attractive of the two litigants will carry greater weight. When a young woman is featured in the media, making an appeal for a baby (by some risky method), it is clear that the public will side with her and that deeper issues in the decision will not be considered.
Fourth, politics. There is no doubt that government departments and ministers are apt to take a certain view in relation to questions that are widely debated, whether it is genetically modified food, fluoridation, or reproductive technology. While the pressures they exert may be indirect, they are nonetheless forceful.

Finally, and only if there is any room at all left for debate and choice, ethics. The HFEA developed five ethical principles derived from the legislation and from our deliberations on real cases day-by-day. The ethical considerations were bolstered by widespread public consultation. First was the assurance of human dignity, worth, and autonomy. In line with international conventions, nobody should be used as a convenience or as a bank of spare parts: consent and counseling are vital. No comatose or dying person should ever have gametes removed from them without their prior consent or even knowledge. Second, the welfare of the potential child. Consideration of its need for a father is enshrined in the legislation, although this requirement is now open to debate. Hence, the concern about cloning and the difficult family relationships that might ensue. Third, safety was given the greatest weight. Newly discovered treatments, such as preimplantation genetic diagnosis and egg freezing, were sometimes delayed for safety checks and trials of viability. Despite public pressure and compassion for those seeking treatment, the safety of the child and mother must be considered. Fourth, respect for the status of the embryo. Legislation lays down the parameters of permitted research and prohibits the mixing of humans and animals, cloning, and research on embryos over fourteen days old. These principles informed decisions about, for example, the posthumous removal of sperm and the ban on sex selection for social reasons.

A fifth principle has now emerged, which is that the saving of life is a good use to which new advances in embryology may be put. An example is the decision to allow preimplantation genetic diagnosis and HLA-typing to attempt to create a sibling whose umbilical cord blood might save an older child. Another example is the legalization of stem cell research.

Despite these complications, there is no doubt in my mind that comprehensive regulation is urgently needed in every state of the United States. The visiting English regulator, listening to the debate in the United States about federal funding of stem cell research, finds herself on another planet.

It seems obvious from history that one cannot commence the process of regulation with stem cell regulation, as if building an inverted pyramid. One has to build up to it from a broad and tested base. One cannot regulate stem cell work more or less in a vacuum without a foundation of
data, sanctions, inspection, monitoring, and uniformity. This is all the more so since there is to my mind no genuine difference between stem cell research and other types of embryo research, with shared safety and ethical issues. There is no genuine moral distinction between the use of embryos for procreation, research, and stem cell growth.

In the United States there appear to be too many cross-sector rules that are unenforceable and overlap: the NIH on research, the FTC on advertising, the FDA on drugs, the ASRM on laboratory accreditation, and state licensing. Even within one state there is a proliferation of guidelines with no enforcement. The California Institute of Regenerative Medicine (CIRM) has guidelines for the use of its funds in research, but there is a second set of guidelines for non-CIRM research.

Overall in the United States there seems to be no uniformity, or only fragmented rules, and reliance on professional self-regulation which is inherently weak. There appears to be no monitoring of the health of mothers and babies in IVF; no regulation of the extent of preimplantation genetic diagnosis; no oversight of the creation and disposal of embryos, or of the move from technology to safe medical treatment; no regulation of commerce in gametes; and no safe register of the names of donors and the outcome of the treatments.

Casting a British eye over U.S. practice, there is a need for uniform substantive legislative prohibitions in relation to cloning, controls on experiments in the womb and genetic manipulation; there is a need for surveillance of laboratories and clinics, and enforcement of the fourteen-day rule for keeping embryos in the laboratory. There should be regulation of the buying and selling of gametes, and consideration should be given to legislation banning the patenting of embryological research.

There is a need for U.S.-wide legislatively guaranteed procedures and openness. There should be studies of the health of IVF children and there should be publicity for the adverse consequences, if any, of certain treatments. Acknowledging the dangers of competition between clinics, there should be standards to ensure the integrity of statistics and to enable comparison between clinics. There should be good patient information, a limit on the number of embryos to be used in any one treatment, uniform safety standards, and penalties for their breach. The United States needs laws about the destination of embryos after the expiration of the permitted storage period and in situations where previously given consents are unilaterally withdrawn, typically upon divorce. Patients’ and donors’ rights, information, and consent in relation to distant research will become increasingly important and must be addressed. There needs to be supervision by an independent, central, and transparent body of people empowered to grant licenses, monitor and permit research, and
impose sanctions backed by criminal penalties. Britain is not alone in having confidence in this method. It has been adopted in Canada, Australia, France, and Japan to some degree.

Whatever the political disadvantages of, and the political jealousies engendered by British-style regulation of embryo research and use, these are minimum standards that are necessary.

In conclusion, I argue that the benefits of regulation are overwhelming. Scientists are sometimes mistrusted; there is unacknowledged competition between the politicians who would like to control every move in this interesting and momentous area, and the clinicians who have, as one would expect, personalities to match the tremendous strides forward into the unknown they have made. (I heard one of them say: “I have made a thousand women pregnant.”) There is also the rich commercial market to be considered and the desires of patients who may be under pressure and uninformed. Only comprehensive regulation can hold the ring and bring order and consensus to this topic.
A TALE OF THREE TAKINGS:
TAKING ANALYSIS IN LAND USE REGULATION IN THE UNITED STATES, AUSTRALIA, AND CANADA

Donna R. Christie*

I. INTRODUCTION

The roots of the American, Australian, and Canadian legal systems spring from a common English heritage. Similarly, their conceptions of property draw on a common heritage, including the Magna Carta¹ and the writings of Blackstone² and Locke,³ with the result that protection of property is a prominent feature of all three legal systems. When the government expropriates private property within these democratic societies, there is a presumption and, in some cases, a constitutional compulsion to compensate the owner.

Beginning in the 1920s, the U.S. Supreme Court deviated from the principle that compensation is only required when the government takes possession of or acquires a legal interest in property. In Pennsylvania Coal Co. v. Mahon,⁴ the Supreme Court held that regulating the use of property, in that case Pennsylvania Coal’s mineral rights, may also require compensation if the regulation “goes too far.”⁵ “Regulatory taking” claims did not become common, however, until the 1970s when land use and environmental regulation became pervasive; because these regulations seriously devalued or limited the use of land, property rights advocates, not only in the United States but also in Australia and Canada, began to seek more extensive protection. By the 1990s, cases in both Aus-

---

* Elizabeth C. & Clyde W. Atkinson Professor of Law, Florida State University College of Law, B.S. Chem. 1969, University of Georgia; J.D. 1978, University of Georgia; Post Doc. 1978–1980, Marine Policy and Ocean Management Program, Woods Hole Oceanographic Institution. The author is grateful to Simon Evans, Russell Brown, Philip Gerard, Lee Godden, Nat Stern, and Rob Atkinson for reviewing and commenting on a draft of this Article and to the Florida State University College of Law and Griffith Law School, Brisbane, for supporting research for the Article. She would also like to thank Jennifer Paterson, Bradley Milkwick, and Payal Shah for their excellent assistance in researching materials and editing the seemingly endless footnotes.

2. E.g., WILLIAM BLACKSTONE, COMMENTARIES ON THE LAWS OF ENGLAND (1765).
4. 260 U.S. 393 (1922).
5. Id. at 415.
tralia and Canada seemed to follow Mahon’s lead by requiring compensation when land use regulations seriously devalued mineral rights.

This article surveys and compares the development of the concept of regulatory taking in the United States, Australia, and Canada in the context of land use regulation. Part II discusses the seminal cases in the three countries, all coincidentally involving mineral rights. Part III examines the constitutional underpinnings of protection of property in the three countries. Finally, Part IV looks at each country’s struggle to balance important public interests reflected in land use and environment regulation with protection of private property and to develop a consistent theory of regulatory taking.

II. THREE SEMINAL CASES

A. The United States: Pennsylvania Coal Co. v. Mahon

The provision in the Fifth Amendment of the U.S. Constitution stating that “private property [shall not] be taken for public use, without just compensation” limits the federal government’s “tacit . . . pre-existing power” of eminent domain by requiring that such takings of property be compensated. Prior to 1922 and the Supreme Court’s decision in Pennsylvania Coal Co. v. Mahon, this limitation had been generally understood as applying to circumstances where the government actually appropriated land or ousted the private landholder. Justice Holmes’ determination in Mahon that property might be “taken” without physical appropriation is credited with introducing the concept of regulatory tak-

6. U.S. CONST. amend. V.
9. The so-called Takings Clause “does not prohibit the taking of private property, but instead places a condition on the exercise of that power.” First English Evangelical Lutheran Church v. County of Los Angeles, 482 U.S. 304, 314 (1987). The Fifth Amendment also provides that no “person . . . shall . . . be deprived of property, without due process of law.” U.S. CONST. amend. V.
10. 260 U.S. 393 (1922).
11. See Lucas v. South Carolina Coastal Council, 505 U.S. 1003, 1014 (1992) (“Prior to Justice Holmes’ exposition . . . it was generally thought that the Takings Clause reached only a direct appropriation of property, or the functional equivalent of a practical ouster of [the owner’s] possession.”) (internal citations omitted).
—what has been termed as “[b]y far, the most intractable constitutional property issue” of modern American jurisprudence.

In *Mahon*, the Pennsylvania Coal Company challenged the constitutionality of the Kohler Act, a Pennsylvania statute prohibiting the mining of coal in a manner that would cause subsidence of “any structure used as a human habitation” and other structures. These included roads and railroads; public buildings, such as schools and hospitals; and commercial buildings, such as factories and stores. The Mahons attempted to invoke the statute to prevent the coal company from mining under their home in a way that would remove support and cause it to sink. The landowners held the property, however, under a deed from the coal company which conveyed the surface rights, but expressly reserved the right to remove all the underlying coal. The deed also stated that the landowners assumed the risk of subsidence and waived all claims for damages arising from future coal mining. The trial court invalidated the statute as unconstitutional, but the Pennsylvania Supreme Court upheld the law as a valid exercise of the police power.

Justice Holmes agreed with the coal company that the police power could not be stretched so far as to extinguish the existing property and contract rights of the company without compensation. Holmes’s often-quoted maxim “that while property may be regulated to a certain extent, if a regulation goes too far it will be recognized as a taking” is far from self-explanatory. His less aphoristic commentary did little to clarify further the extent to which government regulation can interfere with property rights before compensation is required. Holmes explained as follows:

Government hardly could go on if to some extent values incident to property could not be diminished without paying for every such change in the general law. As long recognized some values are enjoyed under an implied limitation and must yield to the police power. But obviously the implied limitation must have its limits or the contract and due proc-

17. *Id.*
18. *Id.*
19. See *id.* at 413.
20. *Id.* at 415 (emphasis added).
ness clauses are gone. One fact for consideration in determining such limits is the extent of the diminution. When it reaches a certain magnitude, in most if not in all cases there must be an exercise of eminent domain and compensation to sustain the act. So the question depends upon the particular facts.21

While Holmes highlighted the complete diminution of value of the coal company’s property rights, he also considered in his analysis the weight to be given the public interest element of the law.22 He adjudged, however, that the public interest in “a single private house”23 was not “sufficient to warrant so extensive a destruction of the defendant’s constitutionally protected rights.”24 Further, in considering the general validity of the statute, Holmes also discounted the overall public interest in a problem created by the “short sighted[ness]” of public officials who had used eminent domain to acquire surface rights without a right of support.25 In holding that the Kohler Act was not a valid exercise of the police power, Holmes summed up with the reminder that “a strong public desire to improve the public condition is not enough to warrant achieving the desire by a shorter cut than the constitutional way of paying for the change.”26

Justice Brandeis’ equally famous dissent in the case focused on the public interest aspects of the Kohler Act in preventing harm to the citizens of the state. Brandeis argued that legislation restricting use of land to protect the public health, safety, or morals from threatened harms is not a taking.27 Brandeis contrasted regulation intended to confer benefits on the public, noting that to legitimate such a restriction on individual landowners, a certain reciprocity of advantage might be necessary.28 “But where the police power is exercised, not to confer benefits upon property owners, but to protect the public from detriment and danger, there is . . . no room for considering reciprocity of advantage.”29

22. Holmes paid some lip service to the judgment of the legislative body in acting in the public interest by stating: “The greatest weight is given to the judgment of the legislature but it is always open to interested parties to contend that the legislature has gone beyond its constitutional power.” Mahon, 260 U.S. at 413.
23. Id.
24. Id. at 414.
25. Id. at 415.
26. Id. at 416.
27. Mahon, 260 U.S. at 417 (Brandeis, J., dissenting).
28. Id. at 422.
29. Id.
B. Australia: Newcrest Mining (WA) Limited v. The Commonwealth of Australia

Australia’s Constitution of 1901 in section 51(xxxi) authorizes the Commonwealth to acquire property on “just terms” for “any purpose in respect of which the Parliament has power to make laws” under that section. The Australian High Court’s Newcrest Mining decision in 1997 has been described as “the first time under Australian law [that] land use restrictions have been held to constitute an acquisition of property for which there is a constitutionally guaranteed right to compensation.”

In Newcrest Mining, the company held mining leases at Coronation Hill in the Northern Territory granted by the Commonwealth. Proclamations in 1989 and 1991 made under the National Parks and Wildlife Con-

31. The states of Australia are not subject to the constitutional requirement to provide just terms. See id. at 650.
32. “Just terms” here differs from “just compensation” under the U.S. Constitution. In Nelungaloo Proprietary Ltd. v. Commonwealth, (1948) 75 C.L.R. 495 (Austl.), the High Court confirmed that just terms does not guarantee full compensation. Justice Dixon’s opinion stated that just terms “appears to refer to what is fair and just as between the community and the owner . . . unlike ‘compensation,’ which connotes full money equivalence.” Id. at 569. It should be noted, however, that just terms often requires full compensation, and a significant number of justices on the High Court have taken the position that full compensation is a requirement. Tom Allen, The Acquisition of Property on Just Terms, 22 SYDNEY L. REV. 351, 371–75 (2000).
33. AUSTL. CONST. § 51(xxxi). The section specifically enumerates the areas in which Parliament has the “power to make laws for the peace, order, and good government of the Commonwealth.” Id. § 51. The just terms clause is a restriction on Parliament’s legislative authority in that “it forbids the making of laws with respect to the acquisition of property . . . on terms that are not just.” Mutual Pools & Staff Pty. Ltd. v. Commonwealth, (1994) 179 C.L.R. 155, 169 (Austl.). By implication, section 51(xxxi) also limits the power of Parliament to enact laws for compulsory acquisition of property under other authorities, referred to as other “heads of power” in Australian jurisprudence. See id. at 177.
36. Id. at 145.
servation Act of 1975\textsuperscript{37} had the effect of extending the Kakadu National Park to include land covered by a number of Newcrest Mining’s unexpired mineral leases. The Act had been amended in 1987 to provide: “No operations for the recovery of minerals shall be carried on in Kakadu National Park.”\textsuperscript{38} The amendments further provided that “the Commonwealth is not liable to pay compensation to any person by reason of the enactment of this Act.”\textsuperscript{39} Newcrest Mining sought to have the proclamations invalidated on the grounds that they constituted an acquisition of Newcrest’s property without provision of “just terms” as required by section 51(xxxi) of the Constitution.\textsuperscript{40}

In contrast to Holmes’ succinct opinion for the U.S. Supreme Court in \textit{Mahon},\textsuperscript{41} the High Court decision in \textit{Newcrest Mining} consisted of seven separate opinions spanning over one-hundred pages. The primary focus of the opinions, however, was not on whether there had been an acquisition of property by the Commonwealth, but rather on the issue of whether section 51(xxxi) was applicable at all to the proclamations. The Commonwealth argued that it acted under the plenary power over the Northern Territory conferred by section 122 of the Constitution which was not qualified by the requirement of just terms.\textsuperscript{42} Four of the seven justices—Toohy, Gaudron, Gummow, and Kirby—found, for varying reasons, that section 51(xxxi) was applicable and found the proclamations invalid for failure to provide just terms in respect of the acquired property. Dissenting Justice Brennan joined them in finding that the proclamations would have effected an acquisition of property requiring just terms if section 51(xxxi) had been applicable.\textsuperscript{43}

The issue of whether the proclamations actually acquired property received relatively little attention by the High Court. Only Justice McHugh, a dissenter, directly discussed the effect of the proclamations in terms of the \textit{Tasmanian Dam Case}.	extsuperscript{44} Justice McHugh expressed no doubt that the mining leases were property within the purposes of section 51(xxxi), but suggested that just terms refers not only to the payment of compensation, but also to the Commonwealth’s receiving some “benefit
or advantage.” McHugh reasoned that although Newcrest Mining’s right to mine was adversely affected, the Commonwealth received no proprietary interest in the minerals. Even a total extinguishment of Newcrest’s interest required no compensation if there was no gain by the Commonwealth. Regardless, McHugh concluded that Newcrest had not forfeited any property interest—its right to exploit those interests had been “merely . . . impinge[d]” upon by the proclamations.

Justice Gummow conceded that the proclamations’ language did not effect a direct acquisition of property, but he had no difficulty concluding that the results of proclamations were not simply an extinguishment of a statutory privilege under a licensing system or “merely an impairment of the bundle of rights constituting the property of Newcrest.” He stated that the Conservation Act had “the effect, as a legal and practical matter, of denying to Newcrest the exercise of its rights under the mining tenements . . . [resulting in] sterilisation of the rights constituting the property in question.” The “identifiable benefit or advantage” acquired by the Commonwealth did not have to be identical to what was taken from Newcrest. The benefit to the Commonwealth was not the right to the minerals prior to the expiration of the leases, but the ability to operate the park unhampered by any mineral operations by Newcrest.

Justice Kirby agreed that the prohibition on mining operations resulted in an acquisition of Newcrest’s mining tenements, but his strong position on the relation of the Constitution to individual property rights distinguished his opinion. Kirby categorized property as a fundamental right, and argued that the Constitution should be interpreted in a man-

45. Id. See also The Tasmanian Dam Case, (1983) 158 C.L.R. at 145 (“The emphasis in s. 51(xxxi) is not on a ‘taking’ of private property but on the acquisition of property for purposes of Commonwealth.”).
47. Id.
48. Id.
49. Justice Gummow was joined by Justices Toohey, Gaudron, and Kirby on this issue.
50. Id. at 633.
51. Id. at 635.
52. Id.
53. Id.
54. Id. at 634.
55. Id.
56. Id. at 658–59.
ner that preserves fundamental rights\(^58\) and produces a result that is not “manifestly unjust.”\(^59\)

**C. Canada: The Queen v. Tener\(^60\)**

Canada’s Constitution Act\(^61\) distributes power between the federal government and the provinces,\(^62\) and while the provinces have broad power to expropriate property pursuant to their authority over property and civil rights,\(^63\) the federal government’s power of expropriation is limited to taking property in relation to its specific legislative authorities.\(^64\) Unlike the United States and Australia, Canada has no federal constitutional provision to guarantee compensation for expropriation of property.\(^65\) The federal government and all the provinces and territories have, however, enacted land expropriation acts\(^66\) and other statutes to include compensation provisions. A right to compensation depends on whether there has been an expropriation of property within the scope of a specific law which provides for compensation.\(^67\)

---

58. “Where the Constitution is ambiguous, this Court should adopt the meaning which conforms to the principles of fundamental rights rather than an interpretation which would involve a departure from such rights.” *Id.* at 657. “Where there is an ambiguity in the meaning of the Constitution, as there is here, it should be resolved in favour of upholding such fundamental and universal rights.” *Id.* at 661.

59. *Id.* at 639.

60. [1985] 1 S.C.R. 533 (Can.).

61. Constitution Act, 1867, 30 & 31 Vict. Ch. 3 (U.K.), *as reprinted in R.S.C., No. 5 (Appendix 1985).*

62. *Id.* § 32.


64. *Id.*

65. *Id.* at 31–32. Professor Todd notes that section 1(a) of the Canadian Bill of Rights, which applies only to national legislation, does recognize “the right of the individual to . . . enjoyment of property and the right not to be deprived thereof except by due process of law.” *Id.* at 34. He states that due process, however, “seems to mean nothing more than ‘in accordance with the common law and statute law as it exists at any particular time.’” *Id.* (internal citations omitted). Further, section 2 of the Bill of Rights allows the Parliament, through express declaration in legislation, to infringe upon or abrogate rights. *Id.*


67. Further, Canadian common law provides no basis for compensation. *See infra Part IV.C.*
In *Tener*, the plaintiffs were successors in title to grants of mineral rights issued by the Crown in lands that in 1939 were incorporated into Wells Gray Provincial Park.\(^{68}\) Originally, Wells Gray was designated a Class B park, which allowed for some mining exploration and development to continue.\(^{69}\) With the Park Act\(^ {70}\) in 1965, holders of mineral rights were required to obtain a park use permit to exercise those rights within a park.\(^ {71}\) Then in 1973, Wells Gray was upgraded to a Class A park,\(^ {72}\) which under the Provincial Parks Act\(^ {73}\) meant that park use permits could be issued only when “necessary to the preservation or maintenance of the recreational values of the park involved.”\(^ {74}\) The Teners unsuccessfully applied for several permits between 1974 and 1977, and were finally notified in 1978 that no new exploration or development would be allowed in the park.\(^ {75}\) The Teners then filed suit seeking compensation for acquisition of the mineral claims, past expenditures on the claims, and the present value of the loss of opportunity to exploit the minerals.\(^ {76}\)

Much of the *Tener* opinion is dedicated to a discussion of which act or acts were relevant and whether the acts provided for compensation. Justice Estey, writing for the majority of the Canadian Supreme Court,\(^ {77}\) reminded that there is a “longstanding presumption of a right to compensation”\(^ {78}\) and noted the principle stated in *Attorney-General v. De Keyser’s Royal Hotel, Ltd.*\(^ {79}\) that “[u]nless the words of the statute clearly so demand, a statute is not to be construed so as to take away the property of a subject without compensation.”\(^ {80}\) The Park Act, it was concluded, provided authority for expropriation of land\(^ {81}\) and incorporated

---


\(^{69}\) *Id.* at 555.

\(^{70}\) R.S.B.C., ch. 31 (1965) (Can.).


\(^{72}\) *Id.* at 555.

\(^{73}\) R.S.B.C., ch. 211 (1936) (Can.). In 1973, the Mineral Act, R.S.B.C., ch. 244 (1960) (Can.), was also amended to require authorization by the Lieutenant Governor in Council for exploitation of mineral rights in a park.

\(^{74}\) Park Act, R.S.B.C., ch. 31 § 9(1)(a).


\(^{76}\) *Id.* at 538.

\(^{77}\) Justice Estey wrote for the five-member majority. The two other justices also concurred in the majority’s conclusions.


\(^{80}\) *Tener*, [1985] 1 S.C.R. at 559; *see also* The Queen v. Fisherman’s Wharf Ltd., [1982] 144 D.L.R.3d 21, 28–33 (N.B.C.A.) (discussing the development of this principle in Canadian case law).

by reference the compensation provisions of the Ministry of Highway and Public Works Act.\textsuperscript{82} In determining whether there had been an expropriation, the majority focused on what the government had acquired through the denial of permits, i.e., the benefit to the park.\textsuperscript{83} “The denial of access to these lands . . . amounts to a recovery by the Crown of a part of the right granted to the respondents in 1937. This acquisition by the Crown constitutes a taking from which compensation must flow.”\textsuperscript{84} Justice Estey carefully distinguished this type of regulation from zoning and regulation of activities on land which generally will not trigger compensation.\textsuperscript{85} He stated that zoning and other land use regulations do not add to the value of public property, but that the denial of the permits did add value to the park and constituted an expropriation of the Teners’ interest in the land.\textsuperscript{86} The case was therefore referred to a tribunal for determination of the amount of compensation.

III. IN THE BEGINNING: PROPERTY AND THE CONSTITUTIONS

A discussion of the degree to which property is recognized and receives protection may logically start with a nation’s conception of the role of property in its society and the extent to which property is given constitutional status.\textsuperscript{87} There is a rich literature on the nature of property and the justification for private property,\textsuperscript{88} but in the context of regula-

---

\textsuperscript{82} Id. at 560.
\textsuperscript{83} Id. at 565.
\textsuperscript{84} Id. at 563.
\textsuperscript{85} Id. at 557, 564. Justice Estey noted an exception when zoning is used to “depress the value of property as a prelude to the compulsory taking of the property for a public purpose.” Id. at 557.
\textsuperscript{86} Id. at 565.
\textsuperscript{87} With a plethora of emerging democracies in the last two decades, this historical issue has reemerged as an issue of contemporary debate. See Bruce Ackerman, \textit{The Rise of World Constitutionalism}, 83 VA. L. REV. 771 (1997).
tory takings and the constitutions of the three countries discussed in this article, two models seem most relevant. The first model rejects the absolutist notions of Blackstone and ascribes both private and public aspects to property. The private aspects are bound up in the concept of personal liberty and autonomy, while the public aspect focuses on the role that property serves in society. If government follows a classically republican tradition, private property is subordinate to the public interest; liberalism, however, espouses that government exists to protect individual liberty and property. The second, related model equates property with power. To the extent that the right of property may limit


89. Blackstone defined the “right of property” as “that sole and despotic dominion which one man claims and exercises over the external things of the world, in total exclusion of the right of any other individual in the universe.” William Blackstone, 2 Commentaries *2. His discussions of property also included the following observations:

The third absolute right, inherent in every Englishman, is that of property: which consists in the free use, enjoyment, and disposal of all his acquisitions, without any control or diminution, save only by the laws of the land . . . . So great moreover is the regard of the law for private property, that it will not authorize the least violation of it; no, not even for the general good of the whole community.

William Blackstone, 1 Commentaries *134–35.

90. Gregory Alexander explains that while property serves an individual function, “securing a zone of freedom for the individual in the realm of economic activity,” it also:

serve[s] the public good. This conception does not so much socialize ownership as it conceives the ends which property is to serve as being social. . . . That is, property is privately owned just insofar as this serves the common welfare, where the common welfare is conceived as more than merely the aggregate of individual preference satisfaction.


91. See id.

92. Ely, supra note 88, at 33 (“The sacrifice of individual interests to the greater of the whole formed the essence of republicanism.”).

93. “Government . . . is instituted no less for protection of property, than for the protection of individuals.” William Michael Treanor, The Origins and Significance of the Just Compensation Clause of the Fifth Amendment, 94 Yale L.J. 694, 710 (1985) [hereinafter Treanor, Origins of Just Compensation].
the power of the government and the majority over individuals, property serves to preserve freedom and liberty. This need to limit government and safeguard property is inconsistent with republicanism’s faith in legislative bodies as the “voice of the people . . . to perceive the common good and to define the limits of individual rights.” While these conceptions of property are not necessarily contradictory, they do create what Professor Carol Rose calls a “fundamental tension.”

The current constitutions of the United States, Australia, and Canada were each adopted approximately a century apart, giving the framers of the latter two not only the benefit of the prolific scholarship on property, but also the ability to draw upon the constitutional history of the United States in considering the extent to which property would receive constitutional protection. Each country’s constitutional history provides some insights into the perceptions of the country’s Founders or Framers concerning the nature of property and degree of protection it would be afforded in their society.

A. The Constitutional Provisions of the United States

Neither the first state constitutions nor the Articles of Confederation included a just compensation clause. During America’s colonial and revolutionary period, compensation for government appropriation of land

94. See Nedelsky, supra note 88, at 223 (“Property sets bounds between a protected sphere of individual freedom and the legitimate scope of government authority.”). Property and power also come together in the realm of democratic political participation, as property provides the power to participate effectively in the polity. See Treanor, Origins of Just Compensation, supra note 93, at 699.

95. Treanor, Origins of Just Compensation, supra note 93, at 701.


97. The United States Constitution was adopted by a constitutional convention in 1787 and came into force in 1788; the Australian Constitution came into effect in 1901; Canada’s Constitution Act was adopted in 1982, but it describes the “Constitution of Canada” as comprising the Constitution Act, 1982 and thirty other acts and orders included in a schedule. Constitution Act § 52(2), 1867, 30 & 31 Vict. Ch. 3 (U.K.), as reprinted in R.S.C., No. 5 (Appendix 1985).

98. Expropriation without compensation in these times was common. See Treanor, Origins of Just Compensation, supra note 93, at 698 (“Loyalist property was seized. Undeveloped land was taken for roads. Goods of all types were impressed for military use.”).
was not a firmly entrenched principle.\textsuperscript{99} State legislatures confiscated the land of Loyalists, created obstacles to payment of British debts, issued paper money, and passed debtor relief laws.\textsuperscript{100} The insecurity of property rights contributed to the dislocation of economic relations and instability of society, and led many political leaders to call for a stronger national government.\textsuperscript{101}

The debate in the United States leading to the 1788 Constitution directly pitted liberal conceptions against the republican values of the revolutionary era.\textsuperscript{102} It had been no coincidence that the Declaration of Independence drafted by Thomas Jefferson referred to “life, liberty, and the pursuit of happiness” as “unalienable rights,”\textsuperscript{103} rejecting Locke’s well-known formulation of “life, liberty, and property.”\textsuperscript{104} Jefferson’s republican principles valued property “not as an end in itself but as a foundation for republican government.”\textsuperscript{105} Property was among the private interests that should be subordinate to the common good.\textsuperscript{106} But it was James Madison who set the agenda for the constitutional convention and who was the architect of the Bill of Rights. Jennifer Nedelsky aptly summarized the Madisonian and liberal Federalist views of property as follows:

\begin{quote}
[P]roperty was the central instance of rights at risk in a republic. It was property that had alerted them to the inherent vulnerability of minority rights in popular government, and thus property that became the focal point for the broader problem. And property was not just an abstract symbol. It was a right whose security was essential to the economic and political success of the new republic. If property could not be protected, not only prosperity, but liberty, justice, and the international strength of the new nation would ultimately be destroyed.\textsuperscript{107}
\end{quote}

\textsuperscript{99.} \textit{Id.} at 695–98; see also William Michael Treanor, \textit{The Original Understanding of the Takings Clause and the Political Process}, \textit{95 Colum. L. Rev.} 782, 785 (1995) [hereinafter Treanor, \textit{Takings Clause}]. But see Ely, \textit{supra} note 88, at 25 (“The compensation principle, although recognized, was only imperfectly realized before the Revolution. Yet the colonists generally regarded just compensation as a fundamental principle.”).

\textsuperscript{100.} See Treanor, \textit{Origins of Just Compensation, supra} note 93, at 704.

\textsuperscript{101.} See Ely, \textit{supra} note 88, at 41.

\textsuperscript{102.} See id. at 33–44.

\textsuperscript{103.} Id. at 6; see also Ely, \textit{supra} note 88, at 42 (“Harboring little faith in the people, the [Federalist] framers were not democrats in any modern sense. Indeed, they viewed popular government as a potential threat to property rights.”).
Madison was unsuccessful, however, in inserting a statement of government’s purpose in the Constitution that included protection of private property. While the Constitution does include numerous provisions on economic interests relevant to property, there is no specific “property clause” affirming a fundamental right of property.

In the Bill of Rights, ratified three years after the Constitution, Madison was able to incorporate specific protections for property. As adopted, the Fifth Amendment provides, in relevant part, that “[n]o person shall . . . be deprived of life, liberty, or property, without due process of law; nor shall private property be taken for public use, without just compensation.” The placement of property in the Fifth Amendment, rather than the First Amendment, raises questions about the degree of protection afforded by its inclusion in the Constitution. James W. Ely, Jr. argues that the placement of property in the Fifth Amendment along with the “criminal justice protections” creates a “close association of property rights with personal liberty,” linking property to fundamental civil rights. Richard Epstein, on the other hand, asserts that “all rights [in

108. ELY, supra note 88, at 54.
109. Id. at 43–47.
110. See id. at 46–47; Alexander, Constitutionalising Property, supra note 90, at 90.
111. The Constitution was ratified in 1788, while the Bill of Rights was ratified in 1791.
112. Interestingly, the just compensation clause was the only provision in the Bill of Rights that had not been requested by any of the states. See Treanor, Origins of Just Compensation, supra note 93, at 708–09.
113. U.S. CONST. amend. V.
114. ELY, supra note 88, at 54. Ely goes so far as to state that “[t]he Fifth Amendment explicitly incorporated into the Constitution the Lockean conception that protection of property is a chief aim of government.” Id.
115. After the adoption of the Bill of Rights, Madison expanded upon his definition of property in an essay simply entitled “Property.” In that essay, Madison expressly linked property and civil rights:

There he explicitly contended that the term “property” has two meanings: it simultaneously embraces a private, Blackstonian conception and a public, if not civic, conception. The private meaning is [Blackstone’s] familiar legal conception of property . . . . Madison emphasized that a second, “juster” meaning must be added to this common-law understanding in which property “embraces every thing to which a man may attach a value and have a right.” The first meaning includes a man’s “land,” “merchandize,” and “money.” The second sense extends the meaning of property to include a person’s opinions and “the free communication of them.” It also includes “the free use of [one’s] faculties and the free choice of the objects on which to employ them.” This second meaning encompasses what we today would think of as “civil” rights.
the Bill of Rights] are . . . fundamental."116 Gregory Alexander, however, draws attention to the fact that “property [has] not been treated as a fundamental right, equal in status to the Due Process Clause’s liberty interest or rights under the Equal Protection Clause.”117 He observes that “[c]ourts treat liberty interests as ‘fundamental,’ vigorously protecting them against all governmental encroachments save those undertaken for ‘compelling’ reasons. Property interests, on the other hand, cannot resist any governmental encroachment that passes a weak ‘rationality’ standard.”118

ALEXANDER, COMMODITY & PROPRIETY, supra note 105, at 68. Laura S. Underkuffler also argues, to a different end than Ely, that the Framers had a broad view of property:

The most apparent difference between this [comprehensive] conception of property and most contemporary [absolutist] formulations is its explicit inclusion of a very broad range of individual interests under the rubric of property. The comprehensive approach to property goes far beyond property as physical objects and their analogs to include freedom of expression, freedom of conscience, free use of faculties, and free choice of occupations. The inclusion of a broad range of individual rights, liberties, powers and immunities in the conception of property changes the nature of property’s concreteness. Its concreteness lies not in an actual or symbolic tie to corporeal or incorporeal objects, but in the mediating function that property serves between individual rights and governmental power.

Underkuffler, supra note 88, at 139.

116. EPSTEIN, supra note 88, at 143.


118. Alexander, Fundamental Rights, supra note 117, at 735; see also Carol M. Rose, Property as the Keystone Right?, 71 NOTRE DAME L. REV. 329, 332 (1996) (“Other doubts about property are raised in the double standard of rights introduced by the notorious footnote four of United States v. Carolene Products, according to which ‘mere’ economic rights like property take a constitutional back seat to the rights associated with political participation or the avoidance of majoritarian oppression.”); Laura S. Underkuffler-Freund, Property: A Special Right, 71 NOTRE DAME L. REV. 1033, 1043 (1996) (arguing that property must receive less protection than other rights, such as freedom of speech or religion). Professor Alexander’s understanding has recently been reinforced by Kelo v. New London, 545 U.S. 469 (2005), and Lingle v. Chevron U.S.A. Inc., 544 U.S. 528 (2005). In both cases, the Supreme Court refused to apply a higher standard review
In spite of this disagreement among modern scholars about whether or not the “constitutionalization” of property in the Fifth Amendment establishes its status as a fundamental right, most commentators agree that Madison and his contemporaries expected the just compensation clause to have limited legal consequences. Madison intended that his proposals for the Bill of Rights be noncontroversial, and, for the most part, he simply incorporated rights already recognized in state constitutions or common law. In the case of the property provisions of the Fifth Amendment, he apparently succeeded in his intent, because there is no evidence of any debate or opposition. A just compensation provision that applied only to the federal government and only to physical takings for property interests, rejecting a substantive due process role for the Court in reviewing legislative determinations.

119. See Treanor, Origins of Just Compensation, supra note 93, at 708, 710–13; see also John F. Beggs, The Theoretical Foundations of the Takings Clause and the Utilization of Historical Conceptions of Property in the Ecological Age, 6 FORDHAM ENVTL. L.J. 867, 894 (1995) (“[T]he inclusion of a takings clause in the Constitution was not a significant concern of the Framers and therefore no such provision was included in the document. Even after the Convention, none of the states actually lobbied for the inclusion of a takings clause in the Bill of Rights.”); John F. Hart, Land Use Law in the Early Republic and the Original Meaning of the Takings Clause, 94 NW. U. L. REV. 1099, 1132–33 (2000) (stating that, at the time it was ratified, not much was said about the Takings Clause, and the Framers’ silence indicates that they understood the takings clause to be a “confirmation of the status quo”); Charles P. Lord, Stonewalling the Malls: Just Compensation and Battlefield Protection, 77 VA. L. REV. 1637, 1668 (1991) (stating that the Framers did not contemplate the Takings Clause playing a part in health or safety regulations or nuisance law).

120. See Treanor, Origins of Just Compensation, supra note 93, at 710 n.92 (“Madison proposed only amendments that would not ‘displease the adverse side [the Anti-Federalists].’”).

121. ELY, supra note 88, at 53.

122. Id. at 55; Treanor, Origins of Just Compensation, supra note 93, at 713. In fact, “[t]here are apparently no records of discussion about the meaning of the clause in either Congress or, after its proposal, in the states.” Treanor, Takings Clause, supra note 99, at 791.

123. See ELY, supra note 88, at 54; Treanor, Origins of Just Compensation, supra note 93, at 708. In a later Article, Treanor suggested that:

Madison would have liked the Takings Clause to have regulated state, as well as federal, actions. As a practical matter, however, Madison could not achieve this end directly. The movement to secure a bill of rights came from Anti-federalists who wanted to limit the national government’s power. Madison’s proposal to include in the Bill of Rights an amendment preventing the states from infringing freedom of the press and freedom of conscience and from denying jury trials was defeated by the Senate. Presumably, an attempt to make a takings requirement—a fairly novel right—binding against the states would
ings could not have been particularly controversial at the time since the “federal government did not, in fact, exercise its eminent domain power until the 1870s.”

Indeed, both Federalists and Anti-Federalists believed protection of private property was necessary to promote prosperity. But the Federalists saw the people, particularly the unpropertied majority, as the greatest threat to property. Fearing “Democratic excesses,” the Federalists sought reliance on the judiciary, rather than the legislatures, to protect property rights. The Fifth Amendment property provisions were intended to

have met with a similar fate... [I]t seems that for political reasons he did not try to extend the Takings Clause to the states.

Treanor, Takings Clause, supra note 99, at 843 n.308 (internal citations omitted). The just compensation requirement of the Fifth Amendment was not found to be applicable to the states until rather late in its history. Treanor explains as follows:

It is clear that the Takings Clause is now deemed applicable to the states through the Fourteenth Amendment. Because the line between substantive due process and incorporation was not always a clear one, the precise case that incorporated the Takings Clause is a matter of dispute. The standard citation is to Chicago, B. & Q. Ry. v. Chicago, 166 U.S. 226, 239 (1897). Justice Stevens, however, has argued that this was a substantive due process case. Professor Siegel has stated that incorporation occurred in Fallbrook Irrigation Dist. v. Bradley, 164 U.S. 112 (1896). In Reagan v. Farmers’ Loan & Trust Co., 154 U.S. 362, 399 (1894), the Court appears to hold that the Takings Clause is incorporated through the Fourteenth Amendment’s Equal Protection Clause.

Id. at 860 n.369 (internal citations omitted).

124. See id. at 798 (“The predecessor clauses to the Fifth Amendment’s Takings Clause, the original understanding of the Takings Clause itself, and the weight of early judicial interpretations of the federal and state takings clauses all indicate that compensation was mandated only when the government physically took property.”); see also Lord, supra note 119, at 1668 (observing that Madison did not intend for the compensation clause to apply to anything other than “direct, physical takings”); Bernard Schwartz, Takings Clause—’Poor Relation’ No More?, 47 Okla. L. Rev. 417, 420 (1994) (“Madison’s word choice indicates that he ‘intended the clause to apply only to direct, physical taking of property.’”); Mark Tunick, Constitutional Protections of Private Property: Decoupling the Takings and Due Process Clauses, 3 U. Pa. J. Const. L. 885, 886 n.10 (2001) (“[P]roposing texts of early amendments to the Constitution, Madison suggested the following formulation of the Takings Clause: ‘No person shall be... obliged to relinquish his property, where it may be necessary for public use, without a just compensation.’”).

125. Treanor, Origins of Just Compensation, supra note 93, at 709 n.78. Until Kohl v. United States, 91 U.S. 367 (1875), which clearly recognized the right of the federal government to condemn land, the federal government had traditionally requested the states to condemn land for federal use.

126. ALEXANDER, COMMODITY & PROPRIETY, supra note 105, at 405 n.83.

127. Id. at 405–06.
provide judicially enforceable limitations on government. In his speech to Congress proposing the Bill of Rights, Madison conjectured that “[i]ndependent tribunals of justice will consider themselves in a peculiar manner guardians of those rights . . . .” In this respect, Ely notes, “Madison proved to be a good prophet.”

Although the liberal ideology of the Federalists dominated the constitutional debate in the United States, even this brief overview of the history and politics reveals that Madison was constrained by competing republican values from completely realizing his vision of property in the Constitution and Bill of Rights. Many scholars reject the notion advocated by neoconservatives that the Constitution embodies a monolithic, classical, liberal view of property. Both republican and liberal views of property are represented in the history and politics of the constitutional era and “have been present in [United States] property tradition at least since the adoption of the Constitution.”

B. Australia’s Constitutional Property Provisions

The Australian Constitution was drafted during a series of conventions in the 1890s. The provision of the Constitution dealing with compulsory acquisition of property was developed late in the debates and arose from concerns distinctly different from those addressed by the...

128. See Ely, supra note 88, at 56; Treanor, Origins of Just Compensation, supra note 93, at 710–11; Treanor, Takings Clause, supra note 99, at 837.
130. Ely, supra note 88, at 56.
131. See supra text accompanying notes 107–10.
132. See, e.g., Epstein, supra note 88, at 107–70 (exemplifying the liberal, neoconservative conception of property).
133. Alexander, Constitutionalising Property, supra note 90, at 92; see also Margaret Jane Radin, The Liberal Conception of Property: Cross Currents in the Jurisprudence of Takings, 88 Colum. L. Rev. 1667 (1988) (“The liberal conception is only incompletely embodied in our constitutional practice.”); Treanor, Takings Clause, supra note 99, at 818 (“The power of the republican view of property during [the framing] period shows that there was no consensus among the framers that majority decisionmakers could not be trusted to determine the appropriate level of protection for property interests.”).
135. This Article’s discussion will focus on the power of compulsory acquisition of property from persons by the Commonwealth under section 51(xxxi) of the Australian Constitution.
property provisions of the Fifth Amendment of the U.S. Constitution.\footnote{136} Having rejected without debate the proposition that property should be a subject matter of Commonwealth legislative authority,\footnote{137} some of the Framers expressed concern about whether provisions of the proposed Constitution or power incidental to express legislative authorities created authority for the Commonwealth to acquire property for public purposes.\footnote{138} The brief debates on the subject concluded that “there must be a power of compulsory taking property for the purposes of the Commonwealth.”\footnote{139} That power was provided by section 51(xxxi), which empowers the Commonwealth Parliament to enact laws for the “acquisition of property on just terms from any State or person for any purpose in respect of which the Parliament has power to make laws.” The terms “acquisition,” “property,” and “just terms” were not defined.

Justice Dixon of the Australian High Court has posited that the roots of section 51(xxxi) are found in the U.S. Constitution, stating: “The source of [section] 51(xxxi) is to be found in the fifth amendment of the Constitution of the United States, which qualifies the power of the United States to expropriate property by requiring that it should be done on payment of fair compensation.”\footnote{140} The constitutional debates and the context of the inclusion of section 51(xxxi), however, provide no evidence of this rationale.\footnote{141} Neither by reference nor by context does the Australian provision relate to the circumstances surrounding the adoption of the Fifth Amendment. The Australian property provision is included in a section of the Constitution segregating and enumerating the powers of Parliament, rather than being included in a bill of rights providing spe-

\footnote{136.}{See R.W. Baker, \textit{The Compulsory Acquisition Powers of the Commonwealth}, in \textit{ESSAYS ON THE AUSTRALIAN CONSTITUTION} 194 (2d ed. 1961); Evans, \textit{Drafting, supra} note 134, at 128; P. H. Lane, \textit{Lane’s Commentary on the Australian Constitution} 216 (1986) (“The sole source of Commonwealth legislative power to acquire compulsorily is s. 51(xxxi).”).}

\footnote{137.}{Evans, \textit{Drafting, supra} note 134, at 125–26. Land law and property law are primarily a state responsibility. Under other heads of authority, however, the Commonwealth does to some extent regulate property. Australian Constitution section 51(xxvi) has allowed regulation of indigenous property rights, while section 51(xxix) has allowed environmental regulation associated with international treaty obligations. \textit{Id.} at 125 n.9.}

\footnote{138.}{See Allen, \textit{supra} note 32, at 352–53 (“In Australia, there was some doubt whether the Commonwealth was sovereign, and hence whether it would have the power of eminent domain in the absence of an express provision to that effect.”); see also Evans, \textit{Drafting, supra} note 134, at 128–29; R. L. Hamilton, \textit{Some Aspects of the Acquisition Power of the Commonwealth}, 5 \textit{FED. L. REV.} 265, 266–67 (1973) (Austl.).}

\footnote{139.}{Evans, \textit{Drafting, supra} note 134, at 129.}


\footnote{141.}{Evans, \textit{Drafting, supra} note 134, at 130.}
specific recognition of individual rights.\textsuperscript{142} Section 51(\textit{xxx}i) is a positive grant of power limited by the purpose of the acquisition and the requirement of “just terms.” The Fifth Amendment is stated in the negative, limiting an inherent right of government.\textsuperscript{143} Section 51(\textit{xxx}i) was intended to assure that the government had the authority to exercise eminent domain, while the Fifth Amendment was intended to limit an inherent power of eminent domain by creating rights in individuals. The wordings of the two provisions, though somewhat similar, have distinct differences.\textsuperscript{144} Justice Dixon later noted that American experience and judicial decisions, though helpful, “cannot be applied directly to s. 51(\textit{xxx}i).”\textsuperscript{145}

The lack of debate surrounding section 51(\textit{xxx}i) does not mean that property was not a central political issue in development of the Australian Constitution, rather that the context was quite different from the U.S. experience. Fear of a democratic government that would subject property rights to redistribution at the will of an unpropertied majority was not a significant issue in the Australian debates.\textsuperscript{146} Commentators have attributed the lack of formal protections for property and civil rights in the Australian Constitution to the Framers’ confidence in the “institution of responsible government.”\textsuperscript{147} Indeed, one objection raised to the inclusion of “just terms” in section 51(\textit{xxx}i) was that it was it was improper to include words in a constitution that suggest that the Parliament might not “act strictly on the lines of justice.”\textsuperscript{148} Further, Australia did not go through a period of revolutionary upheaval with widespread confiscations of the property of British and Loyalists.\textsuperscript{149} Rather, the protection of British property and encouragement of continued British investment in Australia was an issue of major concern for the country’s economic fu-

\textsuperscript{142} See Lane, \textit{supra} note 136, at 225.
\textsuperscript{143} See \textit{supra} text accompanying notes 6–9.
\textsuperscript{144} The references to “acquisition” rather than “taking,” and “just terms” rather than “just compensation” are significant departures from the Fifth Amendment. Graham L. Fricke, \textit{Compulsory Acquisition of Land in Australia} 8–9 (2d ed. 1982); see also Baker, \textit{supra} note 136, at 220 (stating that their “intentions and wording” are different).
\textsuperscript{145} Grace Bros. Pty. Ltd. v. Commonwealth, (1946) 72 C.L.R. 169, 290 (Austl.).
\textsuperscript{146} Evans, \textit{Drafting, supra} note 134, at 131.
\textsuperscript{147} Hilary Charlesworth, \textit{The Australian Reluctance About Rights}, 31 Osgoode Hall L.J. 195, 197 (1993) (Can.). Charlesworth, however, argues that reliance on responsible government is inadequate to protect interests of minorities. \textit{Id.} at 198.
\textsuperscript{148} Evans, \textit{Drafting, supra} note 134, at 128; see also Robert C. L. Moffat, \textit{Philosophical Foundations of the Australian Constitutional Tradition}, 5 Sydney L. Rev. 59, 85–86 (1965) (“[T]he protections to individual rights provided by the traditions of acting as honourable men were quite sufficient for a civilised society.”).
\textsuperscript{149} See Evans, \textit{Drafting, supra} note 134, at 131.
ture. In light of the “longstanding British and colonial tradition” in Australia of providing compensation when the government compulsorily acquired property, the debate centered not on the issue of a compensation clause, but on the “clarity and certainty of property rights and the impartial determination of property disputes” and on whether issues concerning property and investment would be subject to appeal to Britain’s Privy Council.

While the debates contain numerous references to protection of life, liberty, and property as the proper role of government, it is clear that property did not serve the symbolic role that it did in the Madisonian conceptualization of property as a proxy for personal liberty. In his analysis of the drafting of section 51(xxxi), Simon Evans concludes that “it is difficult to fix on an American constitutional provenance for . . . [section 51(xxxi)] and attribute to the Framers an intention that it reflect an American constitutional guarantee and that there is “no basis in the Debates” to determine that the section was intended “as a broad guarantee of individual rights.” Although section 51(xxxi) was the first provision of its kind in the constitution of a Commonwealth country, Evans suggests that the section draws more on the British tradition than on the U.S. Constitution. In England, although the King and, later, Parliament could have compulsorily acquired property through the power of eminent domain without compensation, R. W. Baker points out that there is no historical evidence of the British Crown or Parliament having done so. Robert Moffat also agrees that Australia’s “legal and philosophical traditions have been largely inherited from England” and suggests that references in the constitutional debates to the U.S. Constitution were not for purposes of imitating the United States, but “simply happened to serve the purposes of the debaters who used it to bolster

150. Id. at 138–39.
151. Id. at 132.
152. Id. at 138.
153. Id. at 138–40.
154. See id. at 129.
155. See supra text accompanying note 107.
156. Evans, Drafting, supra note 134, at 132.
157. Id. at 131.
159. Evans, Drafting, supra note 134, at 132.
161. Moffat, supra note 148, at 77.
their arguments.\textsuperscript{162} Both Baker\textsuperscript{163} and Moffat\textsuperscript{164} warn, however, that English, as well as American, authority must be used cautiously in interpreting section 51(xxxi). “[T]he particular interrelationships of institutional operation, tradition, and legal philosophy which have developed in the Australian circumstance have been clearly, if not distinctly, its own.”\textsuperscript{165}

**C. Canada’s Rejection of the Constitutionalization of Property**

Like Britain, Canada does not have a constitution embodied in a single document.\textsuperscript{166} The British North America (BNA) Act of 1867,\textsuperscript{167} which in 1982 was renamed the Constitution Act, 1867,\textsuperscript{168} was the country’s first constitutional framework document. The BNA Act was enacted by the British Parliament and created the Dominion of Canada, which continued to be a British colony.\textsuperscript{169} The Act did not attempt to incorporate all of Canada’s constitutional principles, but merely to provide the rules necessary to establish the confederation and allocate power between the federal Parliament and the provincial legislatures.\textsuperscript{170} The preamble of the Act itself stated that Canada’s Constitution was to be similar in principle to Britain’s.\textsuperscript{171} The BNA Act provided no process for amending Canada’s constitution domestically, so the country’s constitutional law continued to draw on a number of sources, including British law and tradition.\textsuperscript{172}

It is hardly surprising then that the BNA Act contained no U.S.-style bill of rights or specific provisions for protection of property. The Fathers of the Confederation were, however, quite aware of the tensions between property and democracy, and strongly supported liberal conceptions assuring the primacy of property.\textsuperscript{173} The Confederation Debates also reflect the concerns for protection of property from majority rule.\textsuperscript{174}

---

\textsuperscript{162} Id. at 79.
\textsuperscript{163} Baker, supra note 136, at 198–99.
\textsuperscript{164} See Moffat, supra note 148, at 88.
\textsuperscript{165} Id.
\textsuperscript{166} See Peter W. Hogg, Constitutional Law of Canada § 1.2 (1997).
\textsuperscript{167} Constitution Act, 1867, 30 & 31 Vict. Ch. 3 (U.K.), as reprinted in R.S.C., No. 5 (Appendix 1985).
\textsuperscript{168} Hogg, supra note 166, § 53(2).
\textsuperscript{169} Id. § 1.2.
\textsuperscript{170} Id.
\textsuperscript{171} Id.
\textsuperscript{172} Id.
\textsuperscript{174} Id.
But rather than memorializing individual rights in the BNA Act, the Canadian Fathers chose to protect property by establishing a Senate whose members were appointed from the propertied class. The Senate’s veto power over legislation passed by the House of Commons was intended to guarantee that property rights would not be subjected to majoritarian rule.

Canada’s Senate proved to be ineffective in this role. Time and inflation made the property ownership requirement virtually meaningless. Senators voted on the basis of party affiliation rather than other interests, and the public resented the idea of an appointed body exercising a veto power over legislation enacted by elected officials. The model of responsible government adopted by Canada dictates that the ministers and cabinet are responsible to the elected House of Commons, not the Senate. With “no obvious place in this scheme of things,” the Senate exercises restraint in “recognition that, as an appointed body, it has no political mandate to obstruct the House of Commons.” Thus, property rights in Canada were subordinated to the dictates of democratic institutions.

Support for more explicit protection of property and other individual rights and freedoms did not emerge until after World War II. In 1960,

175. Constitution Act para. 17, 1867, 30 & 31 Vict. Ch. 3 (U.K.), as reprinted in R.S.C., No. 5 (Appendix 1985). Qualifications for appointment to the Senate included ownership of property at least “[f]our thousand dollars over and above his debts and liabilities.” Id. para. 23(4). The nature of the Senate and the franchise in relation to property were major issues in the Australian constitutional debates. In the end, it was resolved that both the Senate and House of Representatives would be elected, and that franchise would be the same for both. Evans, Drafting, supra note 134, at 144–47.


177. See id. at 315 (“The Senate . . . became . . . a rubber stamp for Commons legislation and a burial ground for political patrons.”).

178. See id; see also Hogg, supra note 166, § 9.4(c).

179. Hogg, supra note 166, § 9.4(c). “If government policy is defeated by the House of Commons, then the government must resign and make way, either for a new government that can command the support of the House of Commons, or for an election that will provide a new House of Commons.” Id.

180. Id.

181. Alvaro notes that the property debate continued on the question of whether the federal or provincial legislatures had jurisdiction over property issues, not on “whether or not property ownership ought to have been subordinated to the dictates of legislatures.” Alvaro, supra note 173, at 316.

182. See id.; see also Philip W. Augustine, Protection of the Right to Property Under the Canadian Charter of Rights and Freedoms, 18 OTTAWA L. REV. 55, 61 (1986). Augustine notes that in addition to the political and academic support that arose after World War II for adoption of a bill of rights, some members of the Supreme Court of Canada between 1938 and 1957 had begun to formulate an “implied bill of rights” based
the Canadian Bill of Rights was enacted, not as a constitutional measure, but as an ordinary statute applicable only to the federal government. Section 1(a) of the Canadian Bill of Rights provides:

> It is hereby recognized and declared that in Canada there have existed and shall continue to exist without discrimination . . . the following human rights and fundamental freedoms, namely:

(a) the right of the individual to life, liberty, security of the person and the enjoyment of property and the right not to deprived thereof except by due process of law.

Federal legislation that conflicts with Bill of Rights protections is in-operative, unless the conflicting legislation expressly states that it operates notwithstanding the Bill of Rights. Canadian courts have been reluctant to read the Bill of Rights broadly, however, because it is simply a statute, not a constitutional document. It does “not reflect a clear constitutional mandate to make judicial decisions having the effect of limiting or qualifying the traditional sovereignty of Parliament.”

Section 1(a) of the Bill of Rights contains no requirement that the federal government compensate for the appropriation of property, but the language bears a close resemblance to the Fourteenth Amendment of the U.S. Constitution, which provides in relevant part: “nor shall any State deprive any person of life, liberty, or property without due process of law.” The Fourteenth Amendment similarly does not incorporate a specific obligation for states to compensate individuals when property is appropriated, but the U.S. Supreme Court has found that the requirement of “just compensation” can be implied from the due process clause.

One Canadian commentator notes, however, that:

---

184. Hogg states that the government was “reluctant to resort to the anachronistic [amendment] procedures” and that the provinces would not have agreed to a bill of rights that applied to them. HOGG, supra note 166, § 32.1. Alvaro cites the possibility of the weakening of democratic institutions by subjecting legislation to review by a non-elected court as an additional factor. See Alvaro, supra note 173, at 316–17.
185. The Canadian Bill of Rights, S.C., ch. 44 § 1(a) (1960).
188. U.S. CONST. amend. XIV.
Not only is this American jurisprudence irrelevant because its legal foundation is fundamentally different than ours, but in addition, our legal tradition is quite different. Before the enactment of the Canadian Bill of Rights and the Charter, our courts followed the English concept of “due process of law” which because of the doctrine of parliamentary sovereignty prevented any review of legislative action...\(^\text{190}\)

Courts interpreting the due process provision have also rejected the notion of importing the American concept of substantive due process into interpretation of section 1(a) of the Canadian Bill of Rights. In \textit{R. v. Appleby (No. 2)},\(^\text{191}\) the New Brunswick Supreme Court rejected the argument that due process should be interpreted under the Canadian Bill of Rights as having substantive, in addition to procedural, content.\(^\text{192}\) In \textit{Curr v. The Queen},\(^\text{193}\) Justice Laskin of the Canadian Supreme Court demonstrated his wariness of applying American substantive due process analysis to the Canadian Bill of Rights, stating:

The very large words of s. 1(a), tempered by a phrase (“except by due process of law”) whose original English meaning has been overlaid by American constitutional imperatives, signal extreme caution to me when asked to apply them in negation of substantive legislation validly enacted by a Parliament in which the major role is played by elected representatives of the people.\(^\text{194}\)

Justice Laskin did not, however, totally exclude the possibility of reading substantive content into the due process rights if given “compelling reasons” and “objective and manageable standards” to guide the court.\(^\text{195}\)

The Canadian Supreme Court most recently addressed the question of the substantive content of section 1(a)’s property provisions in \textit{Authorson v. Canada (Att’y General)}.\(^\text{196}\) While stating that “Canadian courts have been wary of recognizing such protections,”\(^\text{197}\) the Supreme Court again did not completely reject the proposition that section 1(a) may embody

\(^{190}\) Jean McBean, \textit{The Implications of Entrenching Property Rights in Section 7 of the Charter of Rights}, 26 \textsc{Alberta L. Rev.} 548, 567 (1988) (Can.).
\(^{192}\) Id.; \textit{see also} Smith, Kline & French Laboratories Ltd. v. Can. (Att’y General), [1986] 1 F.C. 274 (Fed. Ct.).
\(^{194}\) Id. at 902.
\(^{195}\) Id. at 899–900.
\(^{196}\) [2003] 2 S.C.R. 40 (Can.).
\(^{197}\) Id. at 56. The Court attributed this to the American experience during the Lochner era. \textit{Id.} The Court also noted, however, that recent jurisprudence under the Charter of Rights and Freedoms might lead to more substantive protections. \textit{Id.} at 58.
The Court was clear, however, that a right to compensation for expropriation of property was not among any possible protections. The Bill of Rights recognizes and protects only rights that existed in 1960 prior to its enactment. The Court found that “[a]t that time it was undisputed, as it continues to be today, that Parliament had the right to expropriate property if it made its intention clear.” The Court bolstered its conclusion by referring to Justice Riddell’s oft-quoted statement in a turn-of-the-century case:

In short, the Legislature within its jurisdiction can do everything that is not naturally impossible, and is restrained by no rule human or divine. If it be that the plaintiffs acquired any rights, which I am far from finding, the Legislature had the power to take them away. The prohibition, “Thou shalt not steal,” has no legal force upon the sovereign body. And there would be no necessity for compensation to be given.

Consequently, property rights, which were not protected before 1960, receive no additional protection by the Bill of Rights from legislation that unambiguously provides for expropriation without compensation.

For those who had advocated a Canadian bill of rights, the 1960 enactment was a distinct disappointment: it applied only to the federal government and not to the provinces; it was merely a statute subject to repeal, amendment, or override by the Parliament; and its application and interpretation by the courts was narrow and non-interventionist.

Among those who sought more protections was Pierre Trudeau, Prime Minister of Canada for most of the period between 1968 and 1984. For the Trudeau government, amendment of the Constitution to include an effective bill of rights was a primary goal. This was achieved with en-

198. See id. (“It is unnecessary to decide now exactly what other substantive protections, if any, might be conferred by the Bill of Rights’ s. 1(a)’s property guarantees.”).
199. Id. at 52; see also R. v. Miller, [1977] 2 S.C.R. 680 (Can.) (holding that no absolute right to life existed prior to the Bill of Rights, so a death penalty statute was not inoperative); R. v. Burnshine, [1975] 1 S.C.R. 693 (Can.) (holding that a right to uniform sentencing across different regions of Canada did not exist prior to 1960, and was therefore not protected by the Bill of Rights).
202. See HOGG, supra note 166, § 33.1. Hogg notes that in the twenty-two years between the Bill of Rights’ enactment and the Charter’s adoption, only one Supreme Court case held an act of Parliament to be inoperative for breach of the Bill of Rights. Id. § 32.5.
203. See id. § 33.1.
actment through constitutional procedures of the Charter of Rights and Freedoms as Part I of the new Constitution Act, 1982.\textsuperscript{204}

The Charter is applicable to both the federal government and the provinces.\textsuperscript{205} Section 52(1) of the Constitution Act, 1982 provides that the Constitution is the “supreme law of Canada.” Section 52(3) “entrenches” the Charter by providing that it can only be changed through procedures laid out in the Constitution Act, 1982 and not through ordinary legislation.\textsuperscript{206} The entrenchment of the Charter laid the foundation for the courts rather than legislatures to exercise the primary responsibility for the protection of individual rights and freedoms. The scope of the Charter’s guaranteed rights is, however, limited by section 1 which “implicitly authorizes the courts to balance the guaranteed rights against competing societal values.”\textsuperscript{207} Section 1 provides:

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.\textsuperscript{208}

The Charter’s expanded degree of judicial review allows the courts to determine “whether the enacting legislative body has made an appropriate compromise between the civil libertarian values guaranteed by the Charter and the competing social or economic objectives pursued by the law.”\textsuperscript{209}

Providing the ability for a non-elected court to overturn the decisions of democratically elected officials was not uncontroversial,\textsuperscript{210} and the

\begin{footnotes}
\begin{enumerate}
\item \textit{Id.}
\item \textit{Id.}
\item \textit{Id.}, \textsection 1.4.
\item \textit{Id.}, \textsection 33.4(c).
\item Hogg, \textit{supra} note 166, at \textsection 33.4(c); \textit{see also} Reference re Section 94(2) of the Motor Vehicle Act (\textit{Motor Vehicle Act Reference Case}), [1985] 2 S.C.R. 486 (Can.).
\item The Canadian Supreme Court in \textit{The Motor Vehicle Act Reference Case}, summarized the controversy as follows:

Yet, in the context of s. 7, and in particular of the interpretation of “principles of fundamental justice,” there has prevailed in certain quarters an assumption that all but a narrow construction of s. 7 will inexorably lead the courts to “question the wisdom of enactments,” to adjudicate upon the merits of public policy. From this have sprung warnings of the dangers of a judicial “super-legislature” beyond the reach of Parliament, the provincial Legislatures and the electorate. The Attorney-General for Ontario, in his written argument, stated that, “the judiciary is neither representative of, nor responsive to the electorate
\end{enumerate}
\end{footnotes}
controversy was perhaps most dramatically played out in the case of property rights. Section 7 of the Charter, as adopted, reads:

Everyone has the right to life, liberty, and security of the person and the right not to be deprived thereof except in accordance with the principles of fundamental justice.211

The section has some conspicuous omissions. First, the right to enjoyment of property is not included as in section 1(a) of the Bill of Rights. Second, the words “due process” are omitted and replaced by the concept of “fundamental justice.” The Trudeau government originally proposed to include property rights within section 7. Attempts were made before enactment of the Charter to amend the section to entrench property rights, but proposed amendments were defeated.212

While the precise influence of the U.S. Constitution on the protection of property rights in the Australian Constitution is not clear, there is no doubt of its effect in the case of the Constitution Act, 1982. In spite of the Canadian Supreme Court’s conservative interpretation of the Bill of Rights’ section 1 “due process” provision213 and additional checks and balances included in the Charter,214 Canada’s Department of Justice feared that inclusion of property rights and due process in a constitutional enactment might lead to “extreme substantive intervention” by the courts that would interfere with social and economic regulation.215 The specter of the U.S. Supreme Court’s Lochner era216 “cast its shadow over Canada as well”217 and “demonstrated the hazard of granting to the

---

211. Canadian Charter of Rights and Freedoms, supra note 208, art. 7.
212. See Alvaro, supra note 173, at 320–26; Augustine, supra note 182, at 67–68; McBean, supra note 190, at 550.
213. The U.S. Supreme Court has rejected substantive due process analysis in the context of economic rights. See West Coast Hotel v. Parrish, 300 U.S. 379 (1937).
214. See supra text accompanying notes 208–10 for a description of the limits imposed by section 1 of the Charter.
215. See Augustine, supra note 182, at 67.
217. HOGG, supra note 166, § 44.7(b). Sujit Choudhry provides interesting insights into the meaning of Lochner to Canadian constitutional thought. He states that:
judges the power to review legislation on a ground as inherently indeterminate as substantive due process.” The provinces viewed constitutional entrenchment of property rights as encroaching upon their jurisdictions, and they were concerned that foreign ownership laws and land use regulation might be found by activist courts to violate property rights guarantees. The New Democratic party took the position that inclusion of property rights in section 7 would “generally render the provinces incapable of effectively legislating with respect to land utilization.” Professor Peter Hogg concludes:

The framers of Canada’s Charter of Rights deliberately omitted any reference to property in s. 7, and they also omitted any guarantee of the obligation of contracts. These departures from the American model, as well as the replacement of “due process” with “fundamental justice”... were intended to banish Lochner from Canada.

The banishment was not quite complete, because the Supreme Court of Canada has interpreted section 7 of the Charter as having substantive as well as procedural aspects in the interpretation of the term “fundamental justice.” In the Motor Vehicle Reference Case, the Court rejected the procedural/substantive dichotomy, stating that the Court had al-

The Lochner era looms as parable or a cautionary tale. Lochner is regarded as a powerful symbol of deep and profound constitutional failure ... The invocation of Lochner operates rhetorically as an epithet, as a signal of potential danger. Indeed, the ongoing citation of Lochner ... bears testimony to its rhetorical power.

Choudhry, supra note 216, at 53.
218. HOGG, supra note 166, § 44.10(a).
221. See McBean, supra note 190, at 550.
222. HOGG, supra note 166, § 44.7(b).
223. The Court came to this conclusion in spite of clear evidence that the framers intended the phrase to be synonymous with “principles of fundamental justice” which is understood in Canadian jurisprudence to mean procedural due process. McBean, supra note 190, at 568–69.
224. [1985] 2 S.C.R. 486 (Can.).
225. Justice Lamer, writing for the majority, explained, as follows:

The substantive/procedural dichotomy narrows the issue almost to an all-or-nothing proposition. Moreover, it is largely bound up in the American experience with substantive and procedural due process. It imports into the Canadian context American concepts, terminology and jurisprudence, all of which are inextricably linked to problems concerning the nature and legitimacy of adjudica-
ways engaged in review of the content of legislation and that the Charter simply broadened the scope of that review. Justice Lamer stated the “interpretation should be . . . a generous rather than a legalistic one, aimed at fulfilling the purpose of the guarantee and securing for individuals the full benefit of the Charter’s protection.” Fundamental justice is guaranteed, however, only in relation to the rights incorporated in section 7. Professor Hogg asserts that the omission of property rights from section 7 precludes the interpretation of the terms “liberty” and “security” to include economic rights, “otherwise, property, having been shut out of the front door, would enter by the back.”

The Bill of Rights was not repealed by the Charter, but it now has limited effect. Only two provisions of the Bill of Rights were not duplicated in the Charter: section 1(a), the due process clause which protects enjoyment of property; and section 2(e), which guarantees a fair hearing in designated circumstances. Section 1(a) remains operative and continues to provide procedural protections for property at the federal, but not provincial, level.

Id. at 498.
226. Id. at 495–96.
227. Id. at 500.
228. Hogg, supra note 166, § 44.9; see also Augustine, supra note 182, at 75–76. Although discussion of the effects of the North American Free Trade Agreement (NAFTA) is beyond the scope of this Article, it should be noted that some scholars suggest that a new “back door” for property rights protection has been created by NAFTA. These scholars argue that NAFTA’s provisions on compensation to investors for direct or indirect expropriations will lead to an Americanization of Canadian constitutional law and the “potential rise of Canada’s own period of Lochnerism.” David Schneiderman, NAFTA’s Takings Rule: American Constitutionalism Comes to Canada, 46 U. Toronto L.J. 499, 536 (1996) (Can.); see also Ricardo Grinspun & Robert Kreklewich, Consolidating Neoliberal Reforms: Free Trade as a Conditioning Framework, 43 Stud. in Pol. Econ. 33 (1994) (arguing that NAFTA is a “conditioning framework” through which external policies and constraints imposed by transnational actors become internalized outside the democratic process).
229. See Hogg, supra note 166, § 32.1.
The issue of constitutional property rights remains controversial in Canada, but relatively recently the Manitoba Court of Appeal succinctly assessed the current status of such rights: “Section 1(a) of the Canadian Bill of Rights, which protects property rights through a ‘due process’ clause, was not replicated in the Charter, and the right to ‘enjoyment of property’ is not a constitutionally protected, fundamental part of Canadian society.”

D. Overview

While the inclusion of property in a country’s constitution does not answer the question of whether compensation is required when government regulation affects only the value or use of property, simply the fact that property rights have been included in a nation’s constitution has consequences. Professor Nedelsky points out that “constitutionalizing property is an extremely powerful symbol of the public/private divide which designates governmental measures affecting property as public ‘interferences’ with a sacred private realm—which then bear the burden of justification.” Furthermore, she contends that entrenchment of a constitutional property guarantee creates a protected sphere of private property that is increasingly insulated from regulation and that reinforces the myth of property as a pre-political and fundamental right.

She also argues convincingly that it bolsters libertarian arguments that market forces, rather than state intervention, should control property use and distribution. Thus, Professor Nedelsky concludes that constitutional entrenchment of property creates an irresolvable conflict between protection of private property interests and promotion of public interests.

The constitutionalization of property enshrines the courts, rather than legislative bodies, as the primary arbiters of the private property/public interest conflict. It is then within the competence of courts to determine the scope of private property protection through their definitions of “property” and interpretations of the meaning of “taking,” “acquisition,”

230. Some of the provinces, commentators, and legislators have continued to argue for inclusion of property protections in the Constitution. See Alvaro, supra note 173, at 328–29; Augustine, supra note 182, at 80; McBean, supra note 190, at 582–83.


233. See id. at 422–23.

234. Id.

235. Id. at 426–28.
Property is shielded from redefinition and redistribution through democratic processes to the extent courts choose to apply constitutional property clauses as a “fundamental barrier-guarantee.”

Property played a markedly different role in the constitutional history of the three countries. Although republican views are represented in U.S. constitutional history, there is no doubt that the Federalist framers did view property as a symbol of personal liberty—a metaphor for a realm of individual autonomy beyond the reach of the democratic will. The guarantees of the Bill of Rights were intended to be interpreted by the courts to protect the individual from an overreaching government. Australia’s Constitution, on the other hand, contains no bill of rights, only constitutional authorization to acquire property on just terms, buried deep in a long list of areas subject to federal government jurisdiction and regulation, and generally intended to empower the federal government within certain limits. There is little in the constitutional history to suggest that it was intended to serve the role of a constitutional guarantee that had been consecrated in a bill of rights.

Australia’s legal traditions concerning compulsory acquisition of property seem to be drawn largely from Britain, but another particularly fundamental constitutional tradition seems distinctly and importantly drawn from the American constitutional experience. The power of judicial review exercised by Australia’s High Court is “a striking and fundamental contrast with the power of the English courts.” Through its broad power of judicial review, Australia’s High Court has had the opportunity to interpret section 51(xxxi) as creating a constitutional guarantee of individual property rights, moving beyond the limited roots of the

236. In the United States, for example, the Supreme Court expanded the protection of property through its interpretation of “due process” in the Fourteenth Amendment to include a requirement of compensation for state property acquisition. See supra note 7.

237. Van der Walt, supra note 220, at 125.

238. Neither the U.S. nor Australian constitutions specifically provide for judicial review. Australia’s constitutional debates, however, indicate that its framers intended the kind of review exercised by the U.S. Supreme Court. For example, speaking in the debates of the 1898 Convention, Sir Isaac Isaacs stated:

We are taking infinite trouble to express what we mean in this Constitution; but as in America so it will be here, that the makers of the Constitution were not merely the Conventions who sat, but the Judges of the Supreme Court. Marshall, Jay, Story, and all the rest of the renowned Judges, who have pronounced on the Constitution, have had just as much to do in shaping it as the men who sat in the original Conventions.

Moffat, supra note 148, at 84.

239. Id.

240. See LANE, supra note 136, at 222.
tion designed to assure that the federal government could exercise a limited power of eminent domain to acquire property.  

Certain justices on the High Court and U.S. Supreme Court have recently expressed views that seem to reinforce Professor Nedelsky’s concerns about the constitutionalizing of property leading to increasingly broad protections. Chief Justice Rehnquist of the U.S. Supreme Court indicated that he saw “no reason why the Takings Clause of the Fifth Amendment, as much a part of the Bill of Rights as the First Amendment or Fourth Amendment, should be relegated to the status of a poor relation.” Justice Kirby of the High Court has also stated that although the Australian Constitution “may sometimes fall short of giving effect to fundamental rights,” when ambiguous, the Constitution should be interpreted to protect such “universal and fundamental rights” as the right to a “legal process which includes provision for just compensation” for government acquisition of private property.

Property has been broadly defined in both countries’ constitutional contexts, but not to the extent that “use,” in and of itself, is considered “property” that is constitutionally protected. Certainly, when the framers of the American and Australian Constitutions crafted their provisions for protection of property, they could not have imagined the current regulatory state and the extent to which property use and values might be affected by land use and environmental laws. Canada, however, drafted the Constitution Act, 1982 fully aware of the need for environmental and land use regulation in the modern world and the tensions that such regulations create with absolutist theories of private property. Further, Canada was cognizant of the history of U.S. courts in developing a theory of regulatory taking and in reviewing legislative actions on the basis of substantive due process. Canada’s decision not to constitutionalize property

241. See van der Walt, supra note 220, at 129–30. Simon Evans’ analysis also concluded that the provision, as it was originally included, was assumed to cover only the appropriation of land. See Evans, Drafting, supra note 134, at 130.
244. Id. at 658.
245. P. H. Lane synthesizes from a number of High Court cases the following definition of property: “[P]roperty is the most comprehensive term that can be used . . . comprising any interest in any property . . . not only conventional estates and interests acknowledged at law and in equity whether in reality or in personality, but also innominate and anomalous interests.” LANE, supra note 136, at 222 (internal quotations omitted). The U.S. Supreme Court has defined property as “everything capable of private ownership.” San Francisco Nat’l Bank v. Dodge, 197 U.S. 70, 90 (1905).
246. For example, general zoning laws that restrict certain uses of property do not normally require compensation. See infra Part IV.B.
manifested the nation’s intention that both the public and private aspects of property should continue to be fully realized through federal and provincial legislative action. This is not to say that in Canada compensation for taking of property does not continue to be generally presumed, but that compensation as a constitutional guarantee is not considered fundamental to a just and democratic society.

IV. MODERN DEVELOPMENT OF THE REGULATORY TAKINGS DOCTRINE IN THE LAND USE CONTEXT

A. The United States Experience: The Search for Categorical Rules

Within four years of the U.S. Supreme Court’s decision in *Mahon*, the Court decided the case of *Village of Euclid v. Ambler Realty Co.*, which found that governments could limit the use and affect the value of land through zoning ordinances that were not “clearly arbitrary and unreasonable, having no substantial relation to the public health, safety, morals, or general welfare.” Because the property was alleged to have decreased in value by seventy-five percent, the case is often cited for the proposition that diminution of value alone does not establish a taking of property requiring compensation. Recognizing the need for governments to be able to respond to the needs of a complex and evolving society, the Court applied an extremely deferential, “fairly debatable” standard to judicial review of the means legislative bodies used to address such issues. Very little additional practical or theoretical guidance emerged from the Supreme Court in the next fifty years to assist governments or property owners in determining when a regulation “goes too far” and requires compensation until the 1970s, when land use and environmental regulations exploded.

248. *See supra* Part II.A for a discussion of the Court’s decision in *Mahon*.
249. 272 U.S. 365 (1926).
250. *Id.* at 395. Two years later, however, in *Nectow v. Cambridge*, 277 U.S. 183 (1928), the Supreme Court did strike down a zoning ordinance as violative of due process. Citing *Euclid*, the Court found that the ordinance had “‘no foundation in reason and [was] a mere arbitrary or irrational exercise of [the police] power.’” *Id.* at 187 (citing *Village of Euclid*, 272 U.S. at 395).
253. *Id.* at 388.
With its Penn Central Transp. Co. v. New York City decision in 1978, the Supreme Court acknowledged the challenge of creating a “set formula” for determining when a regulation requires compensation and opted to continue the “essentially ad hoc, factual inquir[y]” that characterized the few earlier cases. The Court did, however, identify factors of “particular significance” in evaluating a takings claim, including the “economic impact of the regulation on the claimant . . . , particularly, the extent to which the regulation has interfered with distinct investment-backed expectations,” and the “character of the governmental action.” The identification of these factors, however, was only a prelude to the application of a complex and unpredictable balancing test to determine “when ‘justice and fairness’ require that economic injuries caused by public action be compensated by the government, rather than remain disproportionately concentrated on a few persons.”

In the next decade, the Supreme Court undertook to provide clearer guidelines for identifying regulatory action that required compensation under the Fifth Amendment. With Agins v. City of Tiburon in 1980, the Court attempted to summarize the state of the law by pronouncing that the “application of a general zoning law to particular property effects a taking if the ordinance does not substantially advance legitimate state interests or denies an owner economically viable use of his land.”

---

254. 438 U.S. 104 (1978). Penn Central involved a claim that New York had taken the valuable “air rights” above Grand Central Station when, based on designation of the site as a landmark under New York City’s Landmarks Preservation Law, the city refused permission to build a fifty-five story office tower over the railroad terminal.

255. Id. at 124 (quoting Goldblatt v. Hempstead, 369 U.S. 590, 594 (1962)).

256. Id.

257. Id.

258. Id.

259. The Court further described the effect of the character of the government action as follows: “A ‘taking’ may more readily be found when the interference with property can be characterized as a physical invasion by government than when interference arises from some public program adjusting the benefits and burdens of economic life to promote the common good.” Penn Central Transp. Co., 438 U.S at 124 (internal citations omitted).

260. Id.


262. Id. at 260.

263. Id. (citing Penn Central Transp. Co., 438 U.S. at 138 n.36). Rather than referring to Mahon, Tiburon oddly refers to the concession by the City of New York that Penn Central would be entitled to relief if the Terminal ceased to be economically viable. Penn Central’s footnote 13 describes the scheme of “relief” available under New York’s landmark legislation when the owner of a landmark cannot make “a reasonable return on the landmark site.” Penn Central Transp. Co., 438 U.S. at 112 n.13. Possible relief included the acquisition of a protective interest in the property by the city through eminent domain. Id.
Two years later, in *Loretto v. Teleprompter Manhattan CATV Corp.*\(^{264}\), the Supreme Court announced the first categorical taking rule: When government requires a property owner to suffer a permanent physical invasion of her property—no matter how minor—compensation must be provided.\(^{265}\) The Court created a second categorical taking rule in the 1992 case *Lucas v. South Carolina Coastal Council*.\(^{266}\) There the Court found a per se taking in the unusual circumstance where a regulation deprives an owner of all “economically beneficial uses” of the property.\(^{267}\)

Then in 2005 in *Lingle v. Chevron U.S.A. Inc.*\(^{268}\) the Supreme Court revisited the rules announced in *Agins* in order to clarify the distinction between regulatory takings and violations of due process. The Court succinctly surveyed the history of regulatory taking jurisprudence and concluded as follows:

> Although our regulatory takings jurisprudence cannot be characterized as unified, these three inquiries (reflected in *Loretto*, *Lucas*, and *Penn Central*) share a common touchstone. Each aims to identify regulatory actions that are functionally equivalent to the classic taking in which government directly appropriates private property or ousts the owner from his domain. Accordingly, each of these tests focuses directly upon the severity of the burden that government imposes upon private property rights. The Court has held that physical takings require compensation because of the unique burden they impose: A permanent physical invasion, however minimal the economic cost it entails, eviscerates the owner’s right to exclude others from entering and using her property—perhaps the most fundamental of all property interests. In the *Lucas* context, of course, the complete elimination of a property’s value is the determinative factor. And the *Penn Central* inquiry turns in large part, albeit not exclusively, upon the magnitude of a regulation’s economic impact and the degree to which it interferes with legitimate property interests.\(^{269}\)

---

265. *Id.* at 438, 441. The Court in *Loretto* held that a law requiring landlords to allow installation of cable facilities on apartment buildings constituted a taking.
266. 505 U.S. 1003 (1992). *Lucas* involved coastal regulations that limited or prohibited building within certain areas based on erosion and coastal hazard evaluations. All of Lucas’ lots fell within a zone where no habitable structures could be built, and the trial court had found that the property had been rendered “valueless.” *Id.* at 1007.
267. *Id.* at 1019. The Court also found, however, that the regulation would not constitute a per se taking if, based on background principles of property law and nuisance, the regulation proscribes a use that was not part of the owner’s title to begin with. The burden of establishing this exception is, however, on the government. *Id.* at 1027.
269. *Lingle*, 544 U.S. at 539 (internal citations omitted).
The Court noted that “government regulation—by definition—involves the adjustment of rights for the public good,”270 but the role of the Takings Clause, the Court reiterated, is to “bar Government from forcing some people alone to bear public burdens which, in all fairness and justice, should be borne by the public as a whole.”271

The Court found that in Agins it had inappropriately commingled due process and takings inquiries. While takings analysis focuses primarily on the effect of a regulation on private property, the “substantially advances” formula of Agins “probes the regulation’s underlying validity.”272 The Court concluded that Agins’s “substantially advances” test is “not a valid method of identifying regulatory taking for which the Fifth Amendment requires just compensation.”273 In striking down this part of the Agins formula, the Court rejected heightened substantive due process analysis of land use regulation. Such a rule, the Court stated:

would require courts to scrutinize the efficacy of a vast array of state and federal regulations—a task for which courts are not well suited. Moreover, it would empower—and might often require—courts to substitute their predictive judgments for those of elected legislatures and expert agencies. . . . [W]e have long eschewed such heightened scrutiny when addressing substantive due process challenges to government regulation. The reasons for deference to legislative judgments about the need for, and likely effectiveness of, regulatory actions are by now well established.274

The unanimous decision in Lingle does not disturb the three primary theories for asserting a compensable taking: the two relatively narrow categorical taking circumstances—a physical occupation or a taking of

270. Id. at 538.
271. Id. at 542–43.
272. Id. at 543. The Court explained:

In stark contrast to the three regulatory takings tests discussed above, the “substantially advances” inquiry reveals nothing about the magnitude or character of the burden a particular regulation imposes upon private property rights. Nor does it provide any information about how any regulatory burden is distributed among property owners. In consequence, this test does not help to identify those regulations whose effects are functionally comparable to government appropriation or invasion of private property; it is tethered neither to the text of the Takings Clause nor to the basic justification for allowing regulatory actions to be challenged under the Clause.

Id. at 542. The Court further noted that no amount of compensation can authorize a regulation that violates due process requirements. Id. at 543.
273. Lingle, 544 U.S. at 545.
274. Id. at 544 (internal citations omitted).
all value—and the Penn Central factors that the Court noted serve “as the principal guidelines for resolving regulatory takings claims.” The case does seem to signal, however, somewhat of a retreat in takings jurisprudence. Future regulatory takings cases will focus on whether a regulation is “functionally comparable to government appropriation or invasion of private property,” limiting analysis to “the magnitude [and] character of the burden a particular regulation imposes on private property rights . . . [and] how the regulatory burden is distributed among property owners.”

The Supreme Court demonstrated in Lingle that the principles applicable to the law of regulatory taking can be succinctly summarized, but their application remains enigmatic. While the Court has identified two categories of regulations that will be considered per se takings, the categories are narrow and subject to exceptions that are as indeterminate in application as the Penn Central factors. In her concurring opinion in Palazzolo v. Rhode Island, Justice O’Connor emphasized that “essentially ad hoc, factual inquiries” would continue to be the norm in regulatory taking challenges, with Penn Central “provid[ing] important guideposts that lead to the ultimate determination of whether just compensation is required.” Significantly, Justice O’Connor noted that these guideposts include elements of the public interest as well as protection of private property: “The purposes served, as well as the effects produced, by a particular regulation inform the takings analysis.”

275. Id. at 539.
276. Id. at 542. For a discussion of the Supreme Court’s apparent retreat from Justice Scalia’s views of property rights, see Laura Underkuffler, Tahoe’s Requiem: The Death of the Scalian View of Property and Justice, 21 CONST. COMMENT. 727 (2004).
277. See Lucas v. South Carolina Coastal Council, 505 U.S. 1003 (1992), where the Court found that even if the regulation denied all economic value of property (a per se taking), no compensation would be due if the prohibited use was not part of the owner’s title because the use was restricted by nuisance and other background principles of property law. For a discussion of the development of this concept, see Michael C. Blumm & Lucus Ritchie, Lucas’s Unlikely Legacy: The Rise of Background Principles as Categorical Takings Defenses, 29 HARV. ENV’T’L L. REV. 321 (2005). In the case of finding an exception based on nuisance law, the complicated balancing test applied is at least as complex and unpredictable as Penn Central’s taking analysis. It has been suggested that “there is perhaps no more impenetrable jungle in the entire law than that which surrounds the word ‘nuisance.’” W. PAGE KEETON ET AL., PROSSER & KEETON ON THE LAW OF TORTS 616 (5th ed. 1984).
279. Id. at 633–34.
280. Id. at 634.
B. Australia: Constitutional Protection Versus Responsible Government

Cases involving the just terms requirement of section 51(xxxi) of the Australian Constitution were rare until the 1990s, and few have involved regulation of land use. The legal inquiry necessary to determine whether a regulation requires compensation under the section is perhaps even more complicated than U.S. regulatory taking analysis. Because the Australian Constitution strictly circumscribes the authority of the federal government, analysis typically begins with an arduous consideration of whether a regulation is a valid exercise of federal authority within the scope of section 51 or some other head of power. Section 51(xxxi) provides for compensation on just terms for acquisition of property for “any purpose in respect of which the Parliament has power to make laws” under section 51. There are cases, however, where regulation falls squarely within authority granted under the section, but an alleged or apparent appropriation of property will not require just terms. In some cases, this is the result of inelegant constitutional drafting in placing the acquisition power in a long list of other constitutional authorities including taxation, confiscation of property for breach of laws, and confiscation of the property of enemy aliens. Compensation for such acquisitions of property would be “irrelevant or incongruous” and would undermine the objective of the law.

In other cases where the just terms requirement has been found to be inapplicable, the High Court has not developed a consistent approach. Instead, the Court has relied upon a number of rationales to explain why just terms are not due in particular circumstances. Several of these ra-

282. See Simon Evans, When Is an Acquisition of Property Not an Acquisition of Property?, 11 PUB. L. REV. 183 (2000) [hereinafter Evans, Acquisition of Property].
283. See id. at 198.
285. See Evans, Acquisition of Property, supra note 282, at 187–89.
286. Like many commentators on Australian regulatory takings jurisprudence, Professor Evans has found the uncertainty surrounding the requirement of just terms problematic in that it “decreases the ability of governments, legislators, and legal advisers to predict confidently how the High Court will respond to proposed or enacted legislation.” Id. at 186. It should be noted that the finding that an acquisition requires just terms can have broad consequences. If an acquisition falls within the scope of section 51(xxxi) and just terms are not provided, then the legislation as well as the attempted acquisition is “null and void.” Baker, supra note 136, at 219.
287. Professor Evans catalogs these rationales as follows: (1) necessary to achieve the (non-acquisitory) objective of the power; (2) an adjustment of competing rights; (3) directed to prevention of a noxious use; (4) adversely affects or terminates rights rather than acquiring them; (5) interest allegedly acquired by the Commonwealth is not prop-
tionales could be relevant to the denial of just terms in the context of land use or environmental regulation. These include: the regulation prevents a noxious use of property rights; the regulation was a necessary or characteristic means to achieve an authorized objective, not chiefly directed at the acquisition of property; or the action adversely affects or terminates property rights rather than acquiring them. But only two section 51(xxxi) cases have addressed regulations that are arguably categorized as land use regulation: Newcrest Mining and the Tasmanian Dam Case.

In the Tasmanian Dam Case, the Commonwealth sought to stop construction of a dam by the Tasmanian Hydro-Electric Commission on the basis that the dam would “inundate significant Aboriginal archaeological sites, and . . . cause damage to a wilderness area which is of great natural value; (6) acquisition was not “for any purpose in respect of which the Parliament has power to make laws”; (7) affects rights that are “inherently susceptible of statutory modification or extinguishment”; and (8) acquisition was under a head of power that authorizes laws that do not provide just terms. Evans, Acquisition of Property, supra note 282, at 186.

291. A third case, Commonwealth v. Western Australia, (1999) 196 C.L.R. 392 (Austl.), involved the establishment of a defense practice area in Western Australia under the Defense Regulations. The defense practice area included areas where conflicts might occur with provincial and private mining interests, but the majority of the justices found that no inconsistency or acquisition of property had yet occurred. Notable in the case, however, is Justice Callinan’s proposal that a test be adopted for section 51(xxxi) that recognizes individual interests in property as compensable property. Justice Callinan adopted the following approach advanced by R. L. Hamilton:

A necessary first step in formulating a test for s. 51(xxxi) . . . is for Australian courts firmly to grasp the principle that the various separate rights of user of property are in themselves property. The court in Dalziel’s case recognised that by taking away some rights of user, in particular the right to possession, the Commonwealth could make property practically worthless . . . . What needs to be recognised is that property is a bundle of rights, and each right in that bundle is itself property the subject of acquisition. Whenever the Commonwealth seeks to control the exercise of one of the rights in the bundle a question of acquisition is on the threshold.

Id. at 489. This approach, referred to in American legal literature as conceptual severance or the “denominator issue,” has been generally rejected by the U.S. Supreme Court. See Penn Central Transp. Co., 438 U.S. at 130–31 (holding that air rights were not a severable compensable property and that the impact of the regulation must determined in relation to the property as a whole).
significance, and which satisfies the criteria for listing on the World Heritage List." 292 After finding that the Commonwealth had authority to prohibit construction of the dam without the Tasmanian Minister’s consent, the Court addressed whether the denial of consent “effect[ed] an acquisition of property otherwise than on just terms.” 293 Citing U.S. regulatory taking jurisprudence, Tasmania argued that although the Commonwealth legislation did not divest title of the land, it so restricted the use of the land and conferred such extensive rights of control over the property in the federal Minister that there had been an acquisition of property. 294 Justice Mason found “no direct relevance to [section] 51(xxxi)” in the U.S. Supreme Court’s interpretation of the Fifth Amendment to find a compensable taking of property if a regulation goes too far in restricting use or abridging property rights. 295 He contrasted the purposes of the constitutional provisions, noting that the American provision seems to be interpreted to assure appropriate allocation of the costs of regulation, 296 while section 51(xxxi) focuses on the “acquisition of property for purposes of the Commonwealth.” 297 Justice Mason acknowledged that maintaining the wilderness character of the land essentially dedicated the land for public use, i.e., protection and conservation, but emphasized that the Commonwealth did not acquire “a proprietary interest of any kind.” 298 For the requirement of just terms to come “into play it is not enough that legislation adversely affects or terminates a pre-existing right that an owner enjoys in relation to his property; there must be an acquisition whereby the Commonwealth or another acquires an interest in property, however slight or insubstantial it may be.” 299

Undoubtedly, the more recent Newcrest Mining case is difficult to reconcile with the Tasmanian Dam Case. A distinction may lie in the fact that the Commonwealth owned the land that benefitted from the regula-

293. *Id.* at 144.
294. *Id.*
295. *Id.*
296. *Id.* at 144–45 (“It seems that the Supreme Court has proceeded according to the view that the object of the clause is to prevent government from forcing some people alone to bear public burdens which should be undertaken by the entire public.”).
298. *Id.* at 146.
299. *Id.* at 145. Justices Murphy and Brennan joined Justice Mason in the conclusion that there had been no acquisition of property. Justices Wilson and Dawson and Chief Justice Gibbs did not decide on the issue, and Justice Deane found that the Act and regulations effected an acquisition. Justice Deane did state, however, that “laws which merely prohibit or control a particular use of, or particular acts upon, property plainly do not constitute an ‘acquisition.’” *Id.* at 283.
tion in Newcrest, but a clear, quantifiable benefit to the Commonwealth could not be established in the Tasmanian Dam Case. Commentators suggest that recent cases are interpreting “acquisition” of property more broadly. But even in an early section 51(xxxi) case, Bank of New South Wales v. Commonwealth, the High Court had found section 51(xxxi) was “not to be confined pedantically to the taking of title by the Commonwealth to some specific estate or interest in land recognized at law or in equity . . ., but that it extends to inommiate and anomalous interests.” The Court there was interpreting Minister of the Army v. Dalziel which involved the question of whether just terms were required when the Army for an indefinite period took possession of vacant land used by Dalziel, a tenant, to operate a parking lot. Although the Army acquired no legal or equitable interest in the land, the acquisition of possession for an indefinite time was held to be an acquisition of property. Dalziel does, however, perhaps provide the best basis to distinguish Newcrest Mining and the Tasmanian Dam Case. In both Dalziel and Newcrest Mining, the Commonwealth took complete possession and control of the property, leaving the owner with none of the rights that constituted the property in question. In the Tasmanian Dam Case, Justice Mason of the High Court emphasized the continued possession and residuary rights of the state, explaining that:

[W]hat is important in the present context is that neither the Commonwealth nor anyone else acquires by virtue of the legislation a proprietary interest of any kind in the property. The power of the Minister to refuse consent under the section is merely a power of veto. He cannot positively authorize the doing of acts on the property. As the State remains in all respects the owner, the consent of the Minister does not overcome or override an absence of consent by the State in its capacity as owner. The fact that the Minister has a power of veto of any develop-

300. See Allen, supra note 32, at 355–58. Simon Evans states that “the distinction between acquisition and deprivation has been progressively eroded.” Simon Evans, Constitutional Property Rights in Australia: Reconciling Individual Rights and the Common Good, in Protecting Rights Without a Bill of Rights 197, 199 (Tom Campbell et al. eds., 2006) [hereinafter Evans, Constitutional Property Rights]. Since Mutual Pools, the High Court has looked not at whether the government has acquired title, but to whether the government has gained “some identifiable and measurable countervailing benefit or advantage.” Id.
301. (1948) 76 C.L.R. 1 (Austl.).
302. Id. at 349.
303. (1944) 68 C.L.R. 261 (Austl.).
304. Id.
development of or activity on the property does not amount to a vesting of possession in the Commonwealth.\textsuperscript{305}

From this perspective, merely restricting the use of land by federal legislation or regulation does not engage the requirement of just terms. This comports with the notion that the interest the government receives must be able to be conceived as “property.” No property has been acquired if the “Commonwealth merely receives the satisfaction of its regulatory goals or if the economic position of the property owner is adversely affected.”\textsuperscript{306}

Land use regulation is more generally the subject of state law, rather than federal law, and it has been recently reaffirmed that state parliaments are not constrained in enacting legislation by the requirement of just terms or just compensation. In Durham Holdings Pty. Ltd. v. New South Wales,\textsuperscript{307} the Parliament of New South Wales passed legislation that vested title in the Crown to all coal deposits within the state. Because the legislation did not provide full compensation to the holders of the largest coal deposits, including Durham Holdings, Durham sought to have the law invalidated.\textsuperscript{308} The just terms provision of the federal Constitution does not apply to the Australian states, and the High Court has held on numerous occasions that the state parliaments have the authority to acquire property even without payment of compensation.\textsuperscript{309} Durham argued, though, that the law must be read in light of the presumption in Australian law that the legislature does not intend to acquire property without compensation\textsuperscript{310} and, further, that some common law rights are so fundamental and embedded in the legal system that parliament cannot override them.\textsuperscript{311} While acknowledging that a presumption against compulsory taking of property without compensation is a recognized principle of construction passed down from England and applied by Australian courts, the High Court also noted that the presumption could not stand in

\begin{footnotesize}
\begin{enumerate}
\item \textsuperscript{305} Tasmanian Dam Case, 158 C.L.R. at 146.
\item \textsuperscript{306} Evans, Constitutional Property Rights, supra note 300, at 200.
\item \textsuperscript{307} (2001) 205 C.L.R. 399 (Austl.).
\item \textsuperscript{308} Id. at 399–400.
\item \textsuperscript{310} Durham Holdings, (2001) 205 C.L.R. at 400.
\item \textsuperscript{311} Id. at 401, 409–10.
\end{enumerate}
\end{footnotesize}
the face of a clearly expressed intent of the Parliament in this case not to provide full compensation to certain holders of coal mining rights.\textsuperscript{312}

The High Court also rejected any role for the courts in this case in reviewing or limiting, on the grounds of deeply rooted, common law rights, the legislative powers of Parliament acting within its constitutional powers.\textsuperscript{313} While not denying the possible existence of some non-constitutional limits on the legislative power, the majority opinion overwhelmingly declined to recognize the definition of “just compensation” as being within that “field of discourse.”\textsuperscript{314} Even Justice Kirby, who emphasized the fundamental nature of property rights in \textit{Newcrest Mining}, stated: “It does not rest upon judicial pronouncements to accord, or withhold, recognition of the law in question by reference to the judge’s own notions of fundamental rights, apart from those constitutionally established.”\textsuperscript{315} Justice Kirby further pointed out the incongruity of the High Court imposing such a constitutional limitation on states when only twelve years earlier in 1988, the electors in all Australian states had soundly rejected a proposal to amend the federal Constitution to make acquisitions by states subject to provisions similar to section 51(xxxi).\textsuperscript{316}

In spite of the absence of constitutional protections,\textsuperscript{317} compensation for the effects of land use and environmental regulation on land value is a prominent feature of state land use and environmental laws in Australia. Legislation in Australian states often reflects acceptance of the principle that owners should be compensated for “injurious affection” to land. The term “injurious affection” was originally found in English legislation compensating an owner for damage or loss of enjoyment of re-

\textsuperscript{312} Id. at 400, 409.
\textsuperscript{313} Id. at 409–10, 427–28.
\textsuperscript{314} Id. at 410.
\textsuperscript{315} Id. at 427.
\textsuperscript{317} The Northern Territory is the only Australian state or territory with a compensation provision for acquisition of property comparable to the federal constitutional provision. See Northern Territory (Self-Government) Act, 1978, which provides:

\begin{quote}
Acquisition of property to be on just terms
\end{quote}

(1) The power of the Legislative Assembly conferred by section 6 in relation to the making of laws does not extend to the making of laws with respect to the acquisition of property otherwise than on just terms.

(2) Subject to section 70, the acquisition of any property in the Territory which, if the property were in a State, would be an acquisition to which paragraph 51(xxxi) of the Constitution would apply, shall not be made otherwise than on just terms.

\textit{Id.} § 50.
tained land caused by activities on a part of the land that had been taken by an eminent domain action or, in Australian terminology, on the land that had been resumed through compulsory acquisition. In Australia, the concept of injurious affection has been applied more broadly to include cases where legislation or regulation affects the value of property without the government actually acquiring a property interest in any of the owner’s land. In this respect, compensation for regulatory taking is much more readily available in Australia than in the United States for mere limitations on use resulting in diminution of value of land by planning, zoning, or environmental regulation.

A comprehensive review of such state compensation provisions is beyond the scope of this article, but a few examples are instructive. Some provisions of planning and environmental legislation continue to use the term “injurious affection.” For example, the Western Australia Town Planning and Development Act 1928 provides:

Any person whose land or property is injuriously affected by the making of a town planning scheme shall, if such person makes a claim within the time, if any, limited by the scheme (such time not being less than 6 months after the date when notice of the approval of the scheme is published in the manner prescribed by the regulations), be entitled to obtain compensation in respect thereof from the responsible authority.

319. Id.
321. Town Planning and Development Act, 1928 (W. Austl.).
322. Id. § 11(1). An interesting feature of the Act is the “betterment” provision that authorizes sharing of the benefit if a land use planning scheme increases the value of the land. Although not apparently utilized, the “betterment” provision provides:

Whenever, by the expenditure of money by the responsible authority in the making and carrying out of any town planning scheme, any land or property is within 12 months of the completion of the work, or of the section of the work affecting such land, as the case may be, increased in value, the responsible authority shall be entitled to recover from any person whose land or property is so increased in value, one half of the amount of such increase.

Id. § 11(2).
The term “injurious affection” has often been difficult to interpret and apply,\(^{323}\) and other statutes have clarified the bases for compensation. For example, Victoria’s Planning and Environment Act, 1987\(^ {324}\) states:

98. Right to compensation

(1) The owner or occupier of any land may claim compensation from the planning authority for financial loss suffered as the natural, direct and reasonable consequence of —

(a) the land being reserved for a public purpose under a planning scheme;

....

(2) The owner or occupier of any land may claim compensation from a responsible authority for financial loss suffered as the natural, direct and reasonable consequence of a refusal by the responsible authority to grant a permit to use or develop the land on the ground that the land is or will be needed for a public purpose.\(^ {325}\)

Queensland’s Integrated Planning Act, 1997\(^ {326}\) provides compensation in regard to development or changes in a planning scheme in the following circumstances:

5.4.2 Compensation for reduced value of interest in land

An owner of an interest in land is entitled to be paid reasonable compensation\(^ {327}\) by a local government if —

(a) a change reduces the value of the interest; and

(b) a development application (superseded planning scheme) for a development permit relating to the land has been made; and

(c) the application is assessed having regard to the planning scheme and planning scheme policies in effect when the application was made; and

(d) the assessment manager, or, on appeal, the court —

(i) refuses the application; or

(ii) approves the application in part or subject to conditions or both in part and subject to conditions.


\(^{324}\) Planning and Environmental Act, 1987 (Vic.).

\(^{325}\) Id. § 98.

\(^{326}\) Integrated Planning Act, 1997, ch. 5 (Queensl.).

\(^{327}\) It is important to note that the legislation does not provide for “full” or “just” compensation.
5.4.3 Compensation for interest in land being changed to public purpose

An owner of an interest in land is entitled to be paid reasonable compensation by a local government if because of a change, the only purpose for which the land could be used (other than the purpose for which it was lawfully being used when the change was made) is for a public purpose.\(^\text{328}\)

Compensation may also be available in legislation regulating areas for nature protection. Tasmania’s Threatened Species Protection Act of 1995\(^\text{329}\) creates the following compensation scheme:

45. Compensation

(1) A landholder is entitled to compensation for financial loss suffered directly resulting from an interim protection order or a land management agreement.

(2) A person who is required to comply with a notice under section 36 is entitled to compensation for financial loss as a result of being required to comply with that notice.

(3) The holder of a license, permit or other authority limited under section 38 is entitled to compensation for financial loss.

. . . .

(6) In making a determination [of compensation], the Minister must have regard to the following matters:

(a) the amount by which the value of the land will be increased or decreased as a result of the interim protection order;

(b) the amount of financial loss, including loss of profit, loss occasioned by breach of contract, loss of production and other consequential loss, to the landholder or other person which would result from compliance with the order;

(c) any increase in the value of the land which would result from the carrying out of works for the purposes of this Act;

(d) the cost of any works required to be carried out on the land;

(e) any change in the value of chattels or improvements which would occur because the land use or activity to which they relate is to be restricted or prohibited by the order;

(f) any other matter which the Minister considers relevant.\(^\text{330}\)

\(^{328}\) Integrated Planning Act, 1997, ch. 5 §§ 5.4.2–5.4.3.

\(^{329}\) Threatened Species Protection Act, 1995 (Tas.).
Similarly, Queensland’s Nature Conservation Act, 1992\(^3\) provides compensation under the following circumstances:

67. Compensation when protected area declared

(1) This section applies if—

(a) a nature refuge is declared under section 49; or

(b) a regulation giving effect to a management plan for a World Heritage management area or international agreement area commences.

(2) If a land holder’s interest in land is injuriously affected by a restriction or prohibition imposed under the declaration or regulation on the land holder’s existing use of the land, the land holder is entitled to be paid by the State the reasonable compensation because of the restriction or prohibition that is agreed between the State and the land holder or, failing agreement, decided by the Land Court.\(^3\)

The provision for compensation in a number of state land use and environmental statutes has not made property rights and regulatory taking less controversial in Australia. Recent High Court decisions confirming the right of states to acquire property without full compensation\(^3\) and the ability of states to condition subdivision and development of land on the transfer of property to the government\(^3\) have heightened concerns of property rights advocates as states have increasingly pursued land use and environmental regulation to protect biodiversity, sensitive ecosystems, and cultural heritage.\(^3\) Environmental advocates, on the other hand, feel that providing compensation for regulation of land raises serious policy issues, including creating an expectation of compensation for regulation in spite of the public interests at stake and creating a climate that inhibits effective regulation because of the financial consequences.\(^3\) But the controversy about the availability of compensation for regulations affecting land value is fought out primarily in the political process in Australian states, rather than in the courts.

---

30. Id. § 45.
32. Id. § 67.
C. Canada: A (Very) Limited Foray into Regulatory Taking

The Canadian experience in finding compensable regulatory takings, usually referred to in Canadian cases and literature as de facto expropriation, is extremely limited.\(^{337}\) Two cases in addition to Tener\(^{338}\) have found a compensable regulatory taking in the land use context.\(^{339}\) Casamiro Resource Corp. v. British Columbia\(^{340}\) and Rock Resources, Inc. v. British Columbia\(^{341}\) are remarkably similar to Tener in that the cases involved the loss of the right to exploit Crown-granted mineral rights within a provincial park.\(^{342}\) The majority opinion of the Canadian Supreme Court in Tener characterized the situation as the recovery by the Crown of an interest in land, i.e., the access rights which were necessary to recover the minerals, without which the claims were “virtually useless.”\(^{343}\) The court in Casamiro found the factual circumstances were equivalent to Tener’s and that the mineral rights had been reduced to “meaningless pieces of paper.”\(^{344}\) In Rock Resources, although the interest taken was classified as personalty, the court relied on Tener’s analysis

---

337. “A de facto expropriation, or as it is known in United States constitutional law, regulatory taking, does not have a long history or clearly articulated basis in Canadian law.” Mariner Real Estate Ltd. v. Nova Scotia, [1999] 177 D.L.R.4th 696, 712 (N.S.C.A.).

338. See supra Part II.C, for a discussion of the Tener decision.

339. In Canada, only five cases have found regulation of property rights a compensable expropriation of property. Three of these cases involved the extinguishment of mining rights by the creation of parks. The groundbreaking case was the decision of the Canadian Supreme Court in Manitoba Fisheries Ltd. v. The Queen, [1979] 1 S.C.R. 101, which required compensation for appropriation of the good will of a fish exporting firm when federal legislation created a monopoly in a statutory corporation to carry on such a business. In that case, the Supreme Court applied the maxim that compensation should be given if property is expropriated unless the statute clearly manifests the intention not to compensate. Professor Hogg notes that the “Crown corporation had in effect acquired the business of exporting fish,” and therefore arguably had acquired something. Hogg, supra note 166, § 28.5(d). Another case has held that destroying contract rights can be equivalent to taking property, and, relying on Manitoba Fisheries, awarded compensation. See Wells v. Newfoundland, [1999] 3 S.C.R. 199 (Can.).


342. Id.

343. Tener, [1985] 1 S.C.R. at 563–64. In Casamiro, the British Columbia Court of Appeal noted that in both Tener and Casamiro, the owner of the Crown-granted mineral claims had the right to the use and occupation of the surface, a right ancillary to the grant of a fee simple, a right granted by statute, and a right also in the nature of a profit a prendre. The surface rights were referred to in the Tener grants, but there was no Crown grant of the surface in either case. Casamiro, 80 D.L.R.4th at 19–20.

344. Id. at 22.
to find a taking. These cases create extremely limited circumstances in which a right to compensation has been recognized, and Canadian courts have rejected expansion of the doctrine in the realm of land use regulation beyond the narrow scope of these cases.

Initially, de facto expropriation cases must be distinguished from “administrative law challenges to the legality or appropriateness of planning decisions.” Similar to the U.S. Supreme Court’s decision in *Lingle*, the Nova Scotia Court of Appeal in *Mariner Real Estate* identified the issue in such circumstances as not whether the regulation constitutes an expropriation, but whether the regulation was lawfully enacted and not ultra vires. Further, in *Hartel Holdings Co. Ltd. v. Council of City of Calgary*, the Supreme Court of Canada stated that it is “clear that municipalities cannot abuse their powers by using them for an improper purpose.” Such actions do not give rise to a claim of de facto expropriation, however, but to damages for tort.

---

345. *Rock Resources*, 229 D.L.R.4th at 155. A fourth mineral rights case, *Cream Silver Mines Ltd. v. British Columbia*, [1993] 99 D.L.R.4th 199 (B.C.C.A.), denied compensation to the holder of mineral claims within a provincial park because Cream Silver held only “bare mineral claims” which were not registrable and did not include absolute access rights. Unlike the Crown-granted mineral rights granted in *Tener* and *Casamiro*, bare mineral claims are not considered “land” for purposes of the Park Act which expressly provided for expropriation of land with compensation. In Chief Justice Finch’s opinion in *Rock Resources*, he argued that *Manitoba Fisheries* did not distinguish personalty and realty in applying a presumption that the legislature intends compensation when property rights are taken. *Rock Resources*, 229 D.L.R.4th at 152.

346. *Mariner Real Estate Ltd.*, 177 D.L.R.4th at 717–18; see also Canadian Pacific Railway Co. v. Vancouver (City), [2006] 1 S.C.R. 227 (Can.) (challenging a land use regulation as beyond the city’s statutory authority and subject to procedural irregularities).

347. See supra text accompanying notes 268–76, for a discussion of *Lingle*.


349. [1984] 1 S.C.R. 337 (Can.).

350. *Id.* at 354. The Supreme Court provided the following examples in the case of down-zoning:

For example, in *Tegon Developments Ltd. v. Edmonton (City of)* (1977), 5 Alta. L.R. (2d) 63 (C.A.), affirmed [1979] 1 S.C.R. 98, the Edmonton City Council was disallowed from putting a freeze on the development of certain property in the hope that the province would designate the property under The Alberta Heritage Act. Similarly, in *Hauff v. Vancouver (City of)* (1980), 12 M.P.L.R. 125, the City of Vancouver’s attempt to pass a by-law for the express purpose of limiting property values with an eye to future acquisition was struck down.

*Id.* at 355. It should be noted that it is not considered bad faith to freeze zoning or further development in anticipation of subsequent acquisition of the land. So long as the actions are taken pursuant to a legitimate and valid planning purpose, “the resulting detriment to
The lack of a constitutional grounding for a right to compensation in Canada for government expropriation of property means that any right must found in the common law, a statute, or a rule of statutory interpretation. Eric C. E. Todd writes, however, that “[i]t has never been suggested that there was a common law right to compensation” in Canada. The leading case on the issue, *Sisters of Charity of Rockingham v. The King*, unambiguously states that “[c]ompensation claims are statutory and depend on statutory provisions. No owner of lands expropriated by statute for public purposes is entitled to compensation . . . unless he can establish a statutory right.”

The first step in the analysis of a compensation claim, then, is to determine whether the government action is undertaken pursuant to a relevant statute authorizing the expropriation of property with compensation. If the statute does not provide for compensation, however, *Sisters of Charity* does not stand for the proposition that no compensation is due because the statute failed to provide explicitly for compensation. The rule of statutory interpretation that “unless the words of the statute clearly so demand, a statute is not to be construed so as to take away the property of a subject without compensation” reflects the British tradition that “[t]he Legislature cannot fairly be supposed to intend, in the absence of clear words shewing [sic] such intention, that one man’s property shall be confiscated for the benefit of others, or of the public.” Nevertheless, the government always has the prerogative “to override or disregard this ordinary principle” through clearly expressed intent. For example, the British Columbia Local Government

---

351. See *Todd*, supra note 63, at 35.
352. [1922] 2 A.C. 315 (P.C.) (appeal taken from Can.).
353. Id. at 322. In one case, however, the court discussed a principle it referred to as “the common law right to compensation for interference with a subject’s property.” France Fenwick & Co. v. The King, [1927] 1 K.B. 458, 467 (U.K.).
354. Apparently, a statute may “clearly so demand” that no compensation be afforded without expressly stating that proposition. For example, in *Cream Silver Mines*, Justice Southin rejected turning a “rule of statutory construction” into a common law rule concerning compensation, where the comprehensive Park Act had few express provisions for compensation. *Cream Silver Mines*, [1993] 99 D.L.R.4th at 208.
356. Id.
357. Id.
Act\textsuperscript{359} provides that “[c]ompensation is not payable to any person for any reduction in the value of that person’s interest in land, or for any loss or damages that result from the adoption of an official community plan or a bylaw under this Division or the issue of a [development] permit under Division 9 of this part.”\textsuperscript{360} When the intent of the legislative body is expressed this clearly, there is no room for interpretation.

The determination that compensation is available under a relevant statutory authority if the government expropriates property still leaves unanswered the questions of whether the interest allegedly expropriated is “property” and of whether a regulation that does not involve the government actually taking possession or occupying land will constitute a taking or de facto expropriation of land or an interest in land. Because most claims for compensation for land use regulations are brought under land expropriation acts,\textsuperscript{361} an initial issue often is whether the property claimed to be taken is “land” or an interest in land within the meaning of those acts.\textsuperscript{362} Expropriations acts are considered remedial statutes, and common law rules of interpretation dictate that such statutes must be given broad and liberal interpretation consistent with their purpose and be strictly construed in favor of the parties who rights are affected.\textsuperscript{363} Neither zoning of land uses in a community or “regulation of specific activity on certain land,” however, are considered to affect an interest of an owner in the land.\textsuperscript{364} In \textit{Belfast Corp. v. O.D. Cars Ltd.},\textsuperscript{365} which is favorably cited on a regular basis by Canadian courts, Viscount Simonds

360. Id. § 914(1). A number of other Canadian jurisdictions also have zoning or planning legislation containing such express limitations on compensation. See, e.g., Planning Act, R.S.N.S., ch. 346 (1989) (N.S.); The Planning Act, R.S.A., ch. P-9 (1980) (Alta.); Agricultural Land Commission Act, R.S.B.C., ch. 36 (1979) (B.C.); The Planning and Development Act, R.S.S., ch. P-13 (1978) (Sask.); The Planning Act, C.C.S.M., ch. P-80 (1975) (Man.).
361. See supra text accompanying notes 65–67.
362. Canadian expropriation acts generally also provide compensation for “injurious affection” of land. Although property owners often make broader claims of the meaning the term, the term can only be applied as defined within the relevant act. In general, compensation for injurious affection only applies to damage to retained land when there has been a formal expropriation of land. See TODD, supra note 63, at 25.
363. See Dell Holdings Ltd. v. Toronto Area Transit Operating Auth., [1997] 1 S.C.R. 32, 44–45 (Can.). \textit{But see TODD, supra note 63, at 25–26 (“As Lord Pearson pointed out in Rugby Water Board v. Shaw Fox there are no common law principles in the law of expropriation. . . . [C]ompulsory acquisition and compensation for it are entirely creatures of statute.”}).
contested that “the right to use property in a particular way was itself property,” but even if the word “property” is given wide import,” he stated that “any one of those rights which in the aggregate constituted ownership of property could [not] itself and by itself aptly be called ‘property.’” This has been found to be true in Canada, even in cases where land development is blocked or frozen. Justice Estey’s statement in Tener that “compensation does not follow zoning either up or down” has become a virtual mantra for Canadian courts.

Closely related is the well-established principle in Canadian jurisprudence that even significant and dramatic loss of value of the land by zoning, by refusal to rezone or to grant development or subdivision approval, or by freezing development pursuant to planning powers alone is not enough to establish a de facto expropriation. “It is well settled that owners may be compelled to surrender some value or future value of their land to the local authority and no price has to be paid.” Professor Hogg points out that:

[m]ost forms of regulation impose costs on those who are regulated, and it would be intolerably costly to compensate them. Moreover, much regulation has a redistributive purpose, it is designed to limit the rights of one group . . . and increase the rights of another . . . . A com-

366. Id. at 517.
367. Id. This statement can be compared to the U.S. Penn Central case where the Supreme Court held that a diminution in the value of property must be determined in relation to the nature and extent of the interference with the rights in the parcel as a whole. Penn Central Transp. Co., 438 U.S. at 130.
371. See TODD, supra note 63, at 22–23.
372. See ROGERS, supra note 370, at 124.
373. This sentiment was seconded by Justice Holmes in Mahon where he stated that “[g]overnment hardly could go on if to some extent values incident to property could not be diminished without paying for every such change in the general law.” Mahon, 260 U.S. at 413.
pensation regime would work at cross-purposes to the purpose of the regulation. Nevertheless, there is an indistinct boundary between regulating and taking.\textsuperscript{374}

In \textit{Mariner Real Estate},\textsuperscript{375} Justice Cromwell of the Nova Scotia Court of Appeal noted that “the loss of economic value of land is not the loss of an interest in land within the meaning of the Expropriation Act.”\textsuperscript{376} Although \textit{Tener} is sometimes characterized as standing for the proposition that a taking occurs when the regulation leaves the land with virtually no economic value, Justice Cromwell more aptly characterized the case as requiring a “virtual extinction of an identifiable interest in land.”\textsuperscript{377} The court opined that although severe loss of economic value might be evidence of the loss of “virtually all the normal incidents of ownership, . . . [i]t is not, however, the decline in market value that constitutes the loss of an interest in land, but the taking away of the incidents of ownership reflected in that decline.”\textsuperscript{378} This interpretation is consistent with the Canadian Supreme Court’s statement in \textit{Tener} that “[e]xpropriation or compulsory taking occurs if the Crown or a public authority acquires from the owner an interest in property.”\textsuperscript{379}

In \textit{Harvard Investments Ltd. v. Winnipeg (City)},\textsuperscript{380} Justice Twaddle of the Manitoba Court of Appeal described \textit{Tener} as establishing “two elements to a [regulatory] taking: (i) the acquisition of an asset by the authority involved or its designate, and (ii) the complete extinguishment of the asset’s value to the owner.”\textsuperscript{381} This test incorporates the final tenet of Canadian takings jurisprudence, as enunciated in \textit{Steer Holdings}:

For there to be a statutory taking which gives rise to a claim for compensation, not only must the owner be deprived of the benefit in its property, there must also be a resulting enhancement or improvement conferred upon whatever entity the Legislature intended to benefit. Something must not only be taken away, it must be taken over.\textsuperscript{382}

\textsuperscript{374} {Hogg, supra note 166, § 28.5(d).}
\textsuperscript{376} Id. at 724.
\textsuperscript{377} Id. at 728; see also 64933 Manitoba Ltd. v. Manitoba, [2002] 214 D.L.R.4th 37, 41 (Man. C.A.) (holding that “a ‘de facto’ expropriation made by government can give rise to an entitlement to compensation, but only if the effect of the government’s action is to essentially extinguish the claimant’s interest in land or property”).
\textsuperscript{378} Mariner Real Estate Ltd., 177 D.L.R.4th at 727.
\textsuperscript{379} Tener, [1985] 1 S.C.R. at 556.
\textsuperscript{381} Id. at 122.
Other cases have also used language referring to the requirement that an expropriation requires the "enhancement or benefit" of public property. The nature of the enhancement or benefit the government must acquire is, however, far from clear. In Tener, Justice Estey seemed to distinguish a confiscatory taking from usual land use regulation by stating that "[t]he imposition of zoning regulation and the regulation of activities on lands . . . add[s] nothing to the value of public property." But the context of the statement was to distinguish the purpose of the denial of Tener's permit, i.e., "the action taken . . . was to enhance the value of the public park," from the purposes of a legitimate land use regulation. Significantly, the Supreme Court in Tener also found that a legal interest—a right of access that was part of the mineral grant—had been recovered by the government.

In Mariner Real Estate, the property owner argued that "where regulation enhances the value of public land, the regulation constitu tes the acquisition of an interest in land." There the private property had been designated as "beach" and protected under the Beaches Act. Permits for even single-family dwellings had been denied by the Minister. The property owners asserted that the protection of beaches and dunes for environmental and recreational purposes created a public benefit and enhanced the value of public beaches seaward of the high-water line. Relying on Tener and favorably citing the Australian High Court in Newcrest Mining, Justice Cromwell of the Nova Scotia Court of Appeal concluded that "for there to be a taking, there must be, in effect, . . . an acquisition of an interest in land and that enhanced value is not such an interest."

Similarly, in Steer Holdings, Justice Kroft of the Manitoba Court of Appeals granted the likelihood that blocking the development in that case "enhanced" a nearby public park and that preventing interference with a creek by construction served the "public interest." He asserted, however, that "[t]his acknowledgment is not . . . tantamount to a finding that there has been the kind of confiscation and transferring of

383. See, e.g., 64933 Manitoba Ltd., [2002] 214 D.L.R.4th at 41–42 (“For there to be a statutory taking which gives rise to a claim for compensation, not only must the owner be deprived of the benefit in its property, there must also be a resulting enhancement or improvement conferred upon whatever entity the legislature intended to benefit.”).
385. Id.
386. Id. at 563–64. But see Hogg, supra note 166, § 28.5(d) (asserting that “[i]n Tener, the Crown acquired nothing”).
387. Mariner Real Estate Ltd., 177 D.L.R.4th at 730.
388. R.S.N.S., ch. 32 (1989) (Can.).
390. Id. at 732.
interest or benefit of the kind found by the Supreme Court in either *Manitoba Fisheries* or *Tener* and found no legally recognized benefit had been transferred to the government or its beneficiary.

The Supreme Court of Canada most recently addressed de facto taking in *Canadian Pacific Railway v. Vancouver (City)*. The unanimously adopted decision reinforced the approach of the provincial courts in assessing whether land use regulation would generally require compensation for diminution of economic value, but added little to the jurisprudence of regulatory taking in Canada. The City of Vancouver had adopted an official development plan (ODP) which designated a former railway corridor owned by Canadian Pacific Railway Co. (CPR), the “only such intact corridor existing in Vancouver,” as a public thoroughfare for transportation. The OPR effectively froze the redevelopment potential of the land, limiting the land to uses such as greenways, cycle paths, and nature trails. CPR argued that all reasonable use of its land had been taken and that it is presumed that the Legislature does not intend to take property without compensation.

The Supreme Court stated that two requirements must be met to establish a de facto taking: “(1) an acquisition of a beneficial interest in the property or flowing from it, and (2) removal of all reasonable uses of the property.” With little analysis, the court found that neither requirement had been made out by CPR. The court acknowledged that CPR did not have to establish that the city had accomplished a forced transfer of property, if the city acquired a “beneficial interest related to the property.” CPR argued that the city had gained a de facto park because of the limitations on the use. The court concluded, however, that “[t]he City . . . gained nothing more than some assurance that the land will be used or developed in accordance with its vision, without even precluding the historical or current use of the land, . . . not the sort of benefit that can be construed as a taking.” Further, the court found that all reasonable uses of the land had not be removed by the law. CPR could continue to use the land for a railway, could lease the land, or enter into pub-

---

394. Id. at *8.
395. Id. at *19–20.
396. Id. at *20.
397. Id.
398. Id.
399. Id. at *21.
lic/private partnerships. The court even characterized the OPR as “expanding[ing] upon the only use the land has known in recent history.”

The Supreme Court of Canada was equally dismissive of CPR’s argument concerning the presumption in favor of compensation and the application of the provincial Expropriation Act. In granting planning and zoning authority to the city, the provincial legislature anticipated effects on property value and in section 569 of the Vancouver Charter provided that in such cases the property shall be deemed not to have been taken and “no compensation shall be payable.” The court found that even if the facts supported a finding of a de facto taking in the case, the provisions of the Charter conclusively negated any inference of a compensable taking.

The province’s Expropriation Act provides, however, that “[i]f an expropriating authority proposes to expropriate land, th[e] Act applies to [require compensation] and, if there is an inconsistency between any of the provisions of th[e] Act and any other enactment respecting the expropriation, the provisions of [the Expropriation Act] apply.” The court explained that the provisions the Vancouver Charter did not amount to an inconsistency, because the Expropriation Act does not apply unless there has been a taking. By statutorily deeming adverse affects on property by zoning and planning regulations not to constitute a taking or expropriation, the Charter removes any possibility of inconsistency and application of the Expropriation Act.

V. CONCLUSION

The cases attributed with introducing the concept of regulatory taking to land use law in the United States, Australia, and Canada were all cases involving mineral rights that had been extinguished. The laws involved did not explicitly acquire an interest in the property and therefore fall generally within the scope of what we refer to as regulatory taking, but these cases do differ from typical land use regulation and perhaps provide poor examples of regulation that simply devalues or restricts use of land. By denying access to the minerals in those cases, the governments

401. Id.
402. Id. at *15.
403. Id. at *22–24.
405. Id. § 2(1).
took away the only thing of value\textsuperscript{407} to the owners of the mineral rights—the ability to take possession of the minerals. U.S. cases have often referred to the right to exclude as an essential or fundamental property right,\textsuperscript{408} but in the case of tangible property, the right to possess seems to be an even more indispensable characteristic of property. Whether these mineral rights cases are viewed as totally diminishing the use or value of the property interest or denying the owner the fundamental right of possession, they present circumstances that are relatively rare in the context of planning, zoning, and environmental regulation.

The case law provides virtually no evidence that Australian or Canadian courts have accepted the concept of regulatory taking in the general context of land use and environmental law. In both \textit{Newcrest Mining} and \textit{Tener}, in addition to finding a complete diminution of value or limitation on use by the owners, the courts found an acquisition of a property interest by the government, arguably removing these cases from realm of pure regulatory taking.\textsuperscript{409} In \textit{Tener}, the Canadian Supreme Court carefully distinguished the case from land use planning and zoning,\textsuperscript{410} and no Canadian case has found a compensable taking in the application of general land use planning or developments controls. In Australia, the \textit{Tasmanian Dam Case} provides the most relevant precedent to apply to land use regulation, and that case is generally cited for proposition that mere regul-

\textsuperscript{407} This statement is based on Justice Holmes' view of the factual situation in \textit{Mahon}. Justice Holmes characterized the Kohler Act in \textit{Mahon} as resulting in a total diminution of the value of the coal company’s affected mineral rights or support estate, and not a partial diminution of value of all the coal company’s interests. See supra text accompanying notes 19–22. This kind of conceptual severance has been rejected in later U.S. taking jurisprudence. For example, in \textit{Keystone Bituminous Coal Ass’n v. DeBenedictus}, 480 U.S. 470 (1987), a Pennsylvania statute similar to the one in \textit{Mahon} was found not to constitute a taking. The coal that could not be mined to avoid subsidence could not be considered separately, but only in the context of the company’s entire coal holdings. The coal that had to be left in place amounted to only a small percentage of the overall holdings, and the diminution of value was not sufficient to constitute a taking. \textit{Id.} at 495–96; \textit{see also Penn Central Transp. Co.}, 436 U.S. at 130–31. In the end, \textit{Mahon}’s ongoing influence is in Justice Holmes’ powerful rhetoric rather than in its providing a strong precedent based on the facts of the case.

\textsuperscript{408} See \textit{Kaiser Aetna v. United States}, 444 U.S. 164, 176 (1979) (The right to exclude is “one of the most essential sticks in the bundle of rights that are commonly characterized as property.”); \textit{see also Nollan v. California Coastal Comm’n}, 483 U.S. 825, 831 (1987); \textit{Loretto v. Teleprompter Manhattan CATV Corp.}, 458 U.S. 419, 433 (1982).

\textsuperscript{409} \textit{Newcrest Mining}, (1997) 190 C.L.R. at 634; \textit{Tener}, [1985] 1 S.C.R. at 565; \textit{see also Evans, Constitutional Property Rights, supra} note 300, at 200 (“The [High] Court has not accepted the American regulatory takings doctrine.”).

\textsuperscript{410} \textit{Tener}, [1985] 1 S.C.R. at 563.
lation of land, even regulation that requires the property to be maintained in its natural state, does not require compensation.\textsuperscript{411}

The emphasis in Australia and Canada on compensation depending upon an acquisition of property or a beneficial interest in property by government reflects British, rather than American, tradition. The ultimate reliance of Australian and Canadian courts in hinging their holdings upon a finding of whether the government receives a benefit or acquires a property interest, however, leaves a back door open for finding a compensable regulatory taking. The words “acquire” and “property” are peculiarly indeterminate in a legal sense. They are broad and malleable terms that may provide a surrogate for balancing public and private interests when the court adjudges, as in cases like \textit{Newcrest Mining} and \textit{Tener}, that the regulation “goes too far.”

The balancing inherent in regulatory taking analysis is not just a balancing of private property rights against the public interest. It is also a balancing of the role of courts with the role of legislatures. Canada’s lack of a constitutional mandate for compensation in the case of compulsory taking of property and its doctrine of parliamentary sovereignty make the role of the court clear and narrow. In \textit{Mariner Real Estate}, Justice Cromwell of the Nova Scotia Court of Appeal explained as follows:

De facto expropriation is conceptually difficult given the narrow parameters of the court’s authority.\ldots While de facto expropriation is concerned with whether the “rights” of ownership have been taken away, those rights are defined only by reference to lawful uses of land which may, by law, be severely restricted. In short, the bundle of rights associated with ownership carries with it the possibility of stringent land use regulation.\ldots Canadian courts have no\ldots broad mandate to review and vary legislative judgments about the appropriate distribution of burdens and benefits flowing from environmental or other land use controls. In Canada, the courts’ task is to determine whether the regulation in question entitles the respondents to compensation under the Expropriation Act, not to pass judgment on the way the Legislature apportions the burdens flowing from land use regulation.\textsuperscript{412}

Australia’s courts are similarly constrained in reviewing land use regulations enacted by state parliaments.\textsuperscript{413} At the federal level in Australia and at both the state and federal levels in the United States, however, the courts find themselves in different role when reviewing the effects of legislation on constitutionally guaranteed rights. But it seems that these

\begin{itemize}
\item \textsuperscript{411} Tasmanian Dam Case, 158 C.L.R. at 145–46.
\item \textsuperscript{412} Mariner Real Estate Ltd., 177 D.L.R.4th at 712–13.
\item \textsuperscript{413} See supra text accompanying note 309.
\end{itemize}
roles are distinctly different in Australia and the United States. The Tasmanian Dam Case demonstrates that the High Court’s review will generally be limited to the questions of whether the government acquired land or an interest in land through its authority under section 51(33iii)\textsuperscript{414} and, if so, whether the legislation provided just terms.\textsuperscript{415} In applying the analysis outlined by the Supreme Court in Penn Central,\textsuperscript{416} U.S. courts, on the other hand, find themselves re-evaluating the legislature’s balancing of private property rights and the public interest served by a land use regulation, a role inconsistent with the British legal tradition inherited by Australia and Canada. Even U.S. courts, however, will not apply a heightened substantive due process analysis reserved for regulation affecting fundamental rights.

In Village of Euclid v. Amber Realty Co.,\textsuperscript{417} the U.S. Supreme Court stated that “while the meaning of constitutional guaranties never varies, the scope of their application must expand or contract to meet the new and different conditions which are constantly coming within the field of their operation. In a changing world, it is impossible that it should be otherwise.”\textsuperscript{418} Even if property is not considered a fundamental right, the constitutionalization of property has made it difficult for U.S. courts to evolve a view of property that conforms to the current conditions in a now densely developed country where external environmental impacts are better understood and quality of life and ecological integrity are often difficult to preserve without government intervention. With no constitutional restraint on Canada’s regulators, however, their courts can demonstrate considerable pragmatism concerning land use regulation as illustrated in Mariner Real Estate, where Justice Cromwell stated:

Considerations of a claim of de facto expropriation must recognize that the effect of the particular regulation must be compared with reason-

\textsuperscript{414} It should be noted, however, that Australia’s High Court may never get to this step of its analysis if the acquisition of property is found to be outside the scope of section 51(33iii). In many ways, the analysis of this “complex and contested” issue substitutes for the balancing analysis applied by U.S. courts in regulatory taking analysis both in terms of the relevant substantive principles and its theoretical underpinnings. See Evans, Constitutional Property Rights, supra note 300, at 202–07. In fact, Professor Evans has proposed a test for determining whether an acquisition of property is within the scope of the section 51(33iii) requirement of just terms that explicitly balances the effect of regulation on private property against the legitimate public interests served, a test that rings of the U.S. Supreme Court’s Penn Central analysis. See Evans, Acquisition of Property, supra note 282, at 203.

\textsuperscript{415} See supra text accompanying notes 292–99.


\textsuperscript{417} 272 U.S. 365 (1926).

\textsuperscript{418} Id. at 387.
able use of the lands in modern Canada, not with their use as if they were in some imaginary state of nature unconstrained by regulation. In modern Canada, extensive land use regulation is the norm and it should not be assumed that ownership carries with it any exemption from such regulation.\footnote{Mariner Real Estate Ltd., 177 D.L.R.4th at 717.}

Although Australia’s Constitution does provide protection for property at the federal level, the requirement that the government actually acquire a property interest or benefit leaves the government substantial freedom to regulate property up to the point where there is “an effective sterilisation of the rights constituting the property.”\footnote{Newcrest Mining, (1997) 190 C.L.R. at 635.} Thus, for loss of value of land through “mere” regulation of land use, property owners find little protection in Australia’s Constitution.\footnote{See Evans, Constitutional Property Rights, supra note 300, at 200.}

Finally, whether property is protected by a common law presumption of compensation, a statutory scheme, or a constitutional guarantee, all three countries continue to recognize the importance of private property in their societies.\footnote{It is beyond the scope of this Article to discuss whether property is better protected in these countries through constitutional judicial review or through political institutions. For an excellent discussion of this issue, see Daniel H. Cole, Political Institutions, Judicial Review, and Private Property: A Comparative Institutional Analysis, 15 Sup. Ct. Econ. Rev. (forthcoming 2007).} In all three countries, compensation is generally available when land is physically acquired. And in all three countries, there are vigorous proponents of expanded protection of property rights. But while the concept of regulatory taking is, at most, extremely limited in Australian and Canadian jurisprudence, it is also not so fully realized in the United States that limited effects of regulation on the use or value of land must be compensated. In all three countries, when regulation falls short of actually acquiring or “sterilizing” property rights in land, it is the legislatures, rather than the courts, who bear primary responsibility for balancing society’s interests with property rights.
CONCEPTUAL COMPARISONS: TOWARDS A COHERENT METHODOLOGY OF COMPARATIVE LEGAL STUDIES

Oliver Brand*

ABSTRACT

Functionalism is still the dominant method of comparative legal studies. This, however, is not the case because functional analysis is particularly well suited for the needs of comparatists, but because of a lack of alternatives. Comparative Law and Economics and various “postmodern” approaches have failed to provide more viable solutions. The first part of this Article examines the virtues and flaws of the respective methods. Under the heading of Conceptual Comparisons, the second part introduces a new approach to comparative law. It follows the lead of other comparative sciences, which have abandoned functionalism some time ago and have replaced it with typological considerations. Conceptual Comparisons adapts these considerations to the particular needs of legal research, thus opening new avenues for the perception of law and its role in different legal systems.

TABLE OF CONTENTS

ABSTRACT ....................................................................................405
I. INTRODUCTION ..........................................................................407
II. DOMINANT APPROACH—FUNCTIONALISM ..............................409
   A. Operation ............................................................................ 409
   B. Development ....................................................................... 411
   C. Putative Problems with the Functional Method ................. 412
       1. Particularism .............................................................. 412
       2. Externalism ............................................................... 414
       3. Ethnocentricity ........................................................... 414

* LL.M. (Cambridge); Dr. iur. (Münster); Lecturer, University of Münster. Unless otherwise indicated, translations are the author’s own. An earlier version of this Article won the 2005 C.J. Hamson Prize in Comparative Law at the University of Cambridge, United Kingdom. The author would like to thank Professor John Bell of Pembroke College, Cambridge, for his “little sparks” of enlightenment and his caring and tireless “guidance through the dark.” A further note of gratitude for their insightful remarks is owed to Ms. Elisabeth A. Passmore and Mr. Nicholas O. Wallach of Peterhouse, Cambridge, and Mr. Matt Dyson of Downing College, Cambridge. Ms. Tyra Saechao has handled the manuscript with more than remarkable skill and dedication. Remaining mistakes are all the author’s own.
D. Real Problems with the Functional Method .......................... 415
   1. Axiomatic Problems ..................................................... 415
      a. First Premise: Law as a Solution of Problems .......... 415
      b. Second Premise: Similarity of Problems ................. 417
      c. Third Premise: Problems Are Solved in a Similar Way ................................................. 418
   2. Operational Problems .................................................. 419
      a. Pseudo-Factuality ................................................... 419
      b. Contemporality ...................................................... 419

E. Conclusion ..................................................................... 420

III. ALTERNATIVE APPROACHES ........................................... 421
A. Mattei—Comparative Law and Economics ...................... 421
   1. Operation .................................................................... 421
   2. Development ............................................................ 423
   3. False Trails of Criticism ............................................. 424
   4. Real Problems ........................................................... 425
      a. Ambiguity ............................................................. 425
      b. Non-Neutrality ...................................................... 426
      c. Distortedness ....................................................... 428
B. Legrand—Comparative Law as Hermeneutic Exercise ....... 428
   1. Operation .................................................................... 429
   2. Development ............................................................ 430
   3. Problems .................................................................... 431
C. Frankenberg—Critical Comparative Law .......................... 432
   1. Operation .................................................................... 432
   2. Development ............................................................ 433
   3. Problems .................................................................... 434
   4. Conclusion .................................................................. 435

IV. OWN APPROACH—CONCEPTUAL COMPARISONS ................ 435
A. Preliminary Remarks ..................................................... 435
   1. Presuppositions .......................................................... 435
   2. Conceptualisation ....................................................... 436
   3. Overview .................................................................... 438
B. Operation ........................................................................ 439
   1. First Phase: Conceptual Orientation ......................... 439
      a. Qualitative Analysis ............................................... 441
         i. Defining Properties ............................................. 443
         ii. Gradation ........................................................ 445
      b. Quantitative Analysis .............................................. 445
I. INTRODUCTION

“Oh eerie tis to roam the fen,”¹ shivers the Münsterian poet Annette von Droste-Hülshoff—and boggy indeed are the contemporaneous fields of comparative law. Peril seems to lurk under every footstep because comparatists supposedly try to reach dry ground without the guidance of serious thoughts on methodology. Methodological “avoidance,”² “agnosticism,”³ or even “anarchism”⁴ are said to be prevalent. On closer reflec-

---

4. TP van Reenen, Major Theoretical Problems of Modern Comparative Legal Methodology (I): The Nature and Role of the Tertium Comparationis, 28 COMP. & INT’L
tion, however, the methodological malaise of comparative law seems to be incoherence rather than a lack of efforts. From the mid-1990s onward, scholars sailing under the somewhat enigmatic banner of “postmodernism” have sparked a lively debate on comparative methodology. Comparatists nonetheless still “start from different points and proceed in different directions with different goals,” as Merryman complained as early as 1974. Meaningful results are obscured because the different schools of methodological thought do not engage in constructive discourse. Thus, the central question of whether we can afford a pluralism of methods in comparative law or whether we have to encourage consensus has become a Gordian knot.

This Article will try to sever the knot by devising a methodological approach to comparative studies that allows scholars to work in their different traditions, but come to coherent results. My argument consists of three parts. First, I will present the dominant approach to comparative studies, functionalism, and analyze it critically (Part II). Subsequently, newer trends that challenge functionalism are introduced and assessed (Part III). In these parts it will be maintained that none of the existing methods suitably fulfils the needs of the comparatist, either to serve as a platform for dialogue or to demand supremacy over the other approaches. Accordingly, Part IV is dedicated to the proposal, explanation, and illustration of a new method of comparative legal studies: Conceptual Comparisons.


II. DOMINANT APPROACH—FUNCTIONALISM

In the history of comparative law, periods of integrative comparison have continuously exchanged with those of contrastive comparison. Today, the hallmark of the former, the so-called “functional method,” has risen to a position of dominance: functionalists author the major treatises on comparative law, fill the editorial boards of comparative journals, and preside over societies dedicated to the study of the subject.

A. Operation

Functionalism is so centrally relevant to contemporary comparative law because of its orientation towards the practical. It is particularly concerned with how to compare the law’s consequences across legal systems and therefore allows rules and concepts to be appreciated for what they do, rather than for what they say. Functionalists believe that the “function” of a rule, its social purpose, is the common denominator (tertium comparationis) that permits comparison.

Functional comparisons rest on three central premises. The first premise relates to the realist conception of the law as an instrument for channeling human behavior and claims that the law answers to social needs or interests. This premise establishes the “problem–solution approach” that functionalists champion. They begin their comparisons by choosing a particular practical problem. Then, they present legal systems with regard to how they resolve this problem. In a third step, similarities and differences between the solutions are listed, explained, and evaluated.

The second premise of functionalism addresses the problem that the actual function of legal institutions is a matter of sociological concern. To avoid large-scale empirical investigations, functionalists presuppose that the problems that the law is asked to resolve are similar or even identical across different legal systems. “If law is seen functionally as a regulator of social facts, the legal problems of all countries are similar. Every legal system . . . is open to the same questions and subject to the

12. KONRAD ZWEIGERT & HEIN KÖTZ, INTRODUCTION TO COMPARATIVE LAW 43–44 (Tony Weir trans., 3d ed. 1998); Husa, supra note 7, at 425.
same standards, even in countries of different social structures or different stages of development.”

The third methodological premise of functionalism, the *praesumptio similitudinis*, maintains that legal systems tend to resolve practical questions in the same way. “[D]ifferent legal systems give the same or very similar solutions, even as to detail, to the same problems of life, despite the great differences in their historical development, conceptual structure, and style of operation.”14 Two reasons explain the existence of this rather counter-intuitive presumption. First, it enables the comparatist to scrutinize social problems and their solutions within the familiar legal framework, rather than having to venture into sociological research. Second, the presumption of similarity can be used as a means of testing the results of a comparison:

the comparatist can rest content if his researches . . . lead to the conclusion that the systems he has compared reach the same or similar practical results, but if he finds that there are great differences or indeed diametrically opposite results, he should be warned and go back to check again whether the terms in which he posed his original question were indeed purely functional, and whether he has spread the net of his researches quite wide enough.15

The *praesumptio* suggests that comparative research is not complete until it has been demonstrated that the legal systems under consideration reach similar results in similar circumstances. This highlights that functional studies are out for the grand similarities of legal systems, not for differences in detail.

Grounded on these three premises, functionalists have been most interested in explaining how norms are similar or different from one jurisdiction to another, how such norms are borrowed or transplanted, and how they are expressed in differing or similar kinds of rules. Normatively, they have fostered the production of uniform law, most suggestively with

---


their attempt to delineate a common core of legal institutions. They believe that they can study legal systems neutrally. The choice of functionality as a tertium comparationis is partially an expression of the desire to avoid seeing foreign legal systems through the mind-set of one’s own legal system. The functional method disregards differences in technical-juridical construction and legal concept so as to deconstruct the local dimension of rules: “the solutions we find in the different jurisdictions must be cut loose from their conceptual context and stripped of their national doctrinal overtones so that they may be seen purely in the light of their function, as an attempt to satisfy a particular legal need.”

B. Development

From the late nineteenth century onwards, functionalism permeated all social sciences. In comparative law, it supplanted previous formalism in the 1920s. Conventionally, Ernst Rabel is its proclaimed father. In the 1950s and ‘60s, functionalism in comparative law was cross-fertilized with anthropological and sociological functionalism, particularly due to the influence of Luhmann. This “embeddedness” of comparative legal functionalism in the network of the social sciences raises first doubts over its future: from the 1970s onward, objections to functionalism convinced the other social sciences to abandon it. That begs the question of whether law is the “happy match” for functionalism or whether similar objections that uprooted functionalism in other social sciences also necessitate its abandonment for comparative law.

18. Z. Weigert & Kötz, supra note 12, at 44; Reitz, supra note 3, at 621–22.
19. Z. Weigert & Kötz, supra note 12, at 61; Ulrich Drobnig, Die Geburt der modernen Rechtsvergleichung, 14 Zeitschrift für Europäisches Privatrecht 821 (2005) (F.R.G); Peters & Schwenke, supra note 5, at 808. In fact, Rabel’s method merely rephrases earlier studies of Max Solomon. See generally Max Salomon, Grundlegung zur Rechtsphilosophie 26–39 (1925) (F.R.G). To focus on Rabel, however, is defensible on the ground that he and his disciples became leading comparatists in the United States and Germany, and thus guarantors for the success of the method.
Another issue is that in the beginning, functionalism—as devised by Rabel—served only two very specific ends. First, it was meant to solve the characterization problem in the field of conflict of laws, which was of “great scholarly concern during the first decades of the twentieth century.” Secondly, efforts were undertaken at the same time to unify commercial law in a new and unstable socio-economic environment. In particular, the International Institute for the Unification of Private Law (UNIDROIT) in Rome asked a number of leading comparative scholars—among them Rabel—to draft a uniform law for the international sale of goods. Rabel hypothesized that, if the legal constructions and characterizations particular to each legal system were ignored and if attention were instead directed exclusively to the law’s actual consequences, then a common or best solution would emerge naturally and directly from the comparison.

Rheinstein, one of Rabel’s disciples, first suggested that the functional method could be generalized and applied beyond the contexts of conflict of laws and unification of law to the entire comparative process. This generalization is related to theories of sociological jurisprudence, especially those of Pound and von Jhering (Interessenjurisprudenz). However, that functionalism originated as a specialized instrument to deal with specific problems casts doubt over the validity of this generalization.

C. Putative Problems with the Functional Method

Accordingly, the functional method has not escaped criticism. Some of the objections, however, miss the point: while aiming at the method as such, they in fact hit instances of incorrect exercise of functionalism.

1. Particularism

Gerber has accused the functional method of producing results that are “particularist,” i.e., unrelated to the socio-economic and historical cir-

cumstances that dictated them.\textsuperscript{27} In another version of this critique, the functional method is criticized for being formalistic or “legocentric.”\textsuperscript{28} As an essentially empirical-inductive method, it is said to rely too heavily on positive legal phenomena embodied in rule-texts and to pay too little attention to “law in action,” i.e., the law in its practical application. Indeed, when turning to the leading textbooks, the numerous facts deliberately left out of the picture because they are considered disruptive to the operation of the method is surprising. Schlesinger, for example, fails to mention history, mores, and ethics.\textsuperscript{29} While it may be true that many functionalists did not always attend to all the factors that play a part in the production of the problem being examined, more recent studies have shown that the functional method as such is capable of doing so. Modern functional comparisons generally accept that comparatists should study not “law in books,” but “law in action”\textsuperscript{30}—though they might not always adhere to this advice in practice. Additionally, they routinely study rules and institutions as part of a larger socio-legal and political context and assess customs and other social practices as devices for solving problems, as, for example, Cappelletti proved in his piece on questions of civil procedure.\textsuperscript{31} However, even the most contextual study can only partially address the extra-legal interdependencies of law. Legal comparisons remain “essayistic.”\textsuperscript{32} This is not a particular flaw of functionalism, but a truth that comparatists will have to accept independently from the method with which they are working.

\textsuperscript{27} Gerber, supra note 17, at 204; see also Markesinis, supra note 23, at 39; Grazia-dei, supra note 10, at 109.
\textsuperscript{28} Frankenberg, Critical Comparisons, supra note 20, at 438; Lawrence M. Friedman, Some Thoughts on Comparative Legal Culture, in COMPARATIVE AND PRIVATE INTERNATIONAL LAW: ESSAYS IN HONOR OF JOHN HENRY MERRYMAN ON HIS SEVENTIETH BIRTHDAY 49, 52 (David S. Clark ed., 1990); see also Pierre Legrand, How to Compare Now, 16 LEGAL STUD. 232, 235 (1996) [hereinafter Legrand, How to Compare Now].
\textsuperscript{31} MAURO CAPPELLETTI & BRYANT GARTH, A World Survey, in 1 ACCESS TO JUSTICE 7–8 (1978); see also Husa, supra note 7, at 423.
2. Externalism

More serious is the accusation that functional studies suffer from taking a purely external view upon the legal systems under comparison. Adopting such a perspective is said to lead to a lack of “immersion,” i.e., a failure in understanding the ideas that lie behind foreign legal systems from the inside.\(^{33}\) Externalism, however, is not necessarily disadvantageous at all, as elucidated by the search for hidden assumptions of different legal systems begun in the 1990s.\(^{34}\) Such assumptions are difficult to detect by lawyers within a particular system, but are more easily understood by foreign lawyers looking from the outside. It is usually the latter who unveil the explicatory potential of unconscious legal assumptions that are so obvious that “culturally immersed” lawyers are barely aware of them. Furthermore, it is an illusion to believe that the comparatist will ever be able to “immerse” in a foreign legal culture—however closely related this culture might be. Kohler and Großfeld are right: the comparatist will always remain bound by his or her preconceptions and cultural disposition; the comparatist will stay “one of his [or her] own people.”\(^{35}\)

3. Ethnocentricity

Another supposed flaw of the functional method is ethnocentricity.\(^{36}\) For the last several decades, U.S.-based émigrés from Europe and their students have led the functional community. These scholars have maintained close ties with their European counterparts. Not surprisingly, therefore, functional studies focus almost exclusively on the comparison of American and European legal systems. That certainly neglects problems and solutions of more remote societies. However, it is probably more telling about the attitudes and agendas of contemporary compara-

---


D. Real Problems with the Functional Method

However, some problems are inherent to the functional method and are not only issues of implementation. They can be divided into two categories: (1) axiomatic ones that originate from the three presuppositions that underpin the functional method; and (2) shortcomings in its operation.

1. Axiomatic Problems

a. First Premise: Law as a Solution of Problems

The first basic assumption of functionalism, that law is a rationally developed entity fulfilling a specific purpose, is a weak starting point. Too many factors that in practice obscure the effectiveness of legal rules are left out of the picture. Saying that law solves problems, for example, presupposes also that it is capable of doing so. That is not always the case. There are bodies of law that are dysfunctional in one of four ways. Firstly, there might be situations where law is enacted for purely symbolic reasons. For instance, a legislator may want to be seen as doing something rather than actually being committed to tackling a problem, as occurred with the German legislation on combat dogs in 2000.

Secondly, norms that would usefully address social problems may be absent in a particular system. Especially strong ideologies might hinder the law in answering social problems effectively, as socialism did with the right of workers to strike. Thirdly, a legal institution may serve ends or obtain results that were neither foreseen nor desired by its framers (“unintended functions”). The main instance of dysfunctionality, however, is that a legal institution might have lost its particular function altogether so that its existence can only be explained historically. Alan Watson has examined the evolution of the rules of private law in various “civil law” countries.

---

37. See supra Part II.A.
39. See CONSTANTINESCO, TOMÉ III, supra note 36, at 65.
40. ALAN WATSON, LEGAL TRANSPLANTS: AN APPROACH TO COMPARATIVE LAW 107–18 (2d ed. 1993); Alan Watson, Legal Culture v. Legal Tradition, in EPISTEMOLOGY AND METHODOLOGY OF COMPARATIVE LAW, supra note 3, at 1, 3; see also JOHN BELL, FRENCH LEGAL CULTURES 17 (2001); LEONTIN-JEAN CONSTANTINESCO, TRAITÉ DE DROIT
cal circumstances have changed, entire bodies of law, mostly from the *Corpus Iuris Civilis*, have been transplanted, essentially unaltered, from society to society. Any interesting connection between the social context of those countries and the rules of their private law remained elusive; legal borrowings often proved inappropriate. One has to conclude that the notion of law as mirror of society is, in fact, just a “mirage.” Not all legal norms and doctrines are functionally related to social life because they run counter to any conceivable need or interest.

A reversed image of the dysfunctionality problem is the multifunctionality dilemma. Functional studies tend to regard the function of law as a monolithic, independent entity. They focus on “the” function of legal institutions. However, a specific legal institution can have simultaneously diverse functions. There might be universal social requirements that need answering. These wants, however, are often accompanied and even eclipsed by needs that are specific to a particular society. Correspondingly, law has not only generic functions, but also at the same time national or regional functions. A contract, for example, universally secures the parties’ expectations of performance. In a market-based economy, it also allows the parties to administer autonomously their economic relations; whereas in a command economy, a contract is a tool to fulfill the goals set out by the plan. Is that really functionally equivalent? A study by Folke Schmidt has revealed that collective labor contracts have as many as five distinct functions. These functions are furthermore dependent on the specific institutional setting in which they are performed. Institutions and structures of a society reflect its unique historical experience and its established ways of working. Bell rightly ex-

---


41. See supra Part II.A; Zweigert & Kötz, supra note 12, at 44.


44. See Brand, supra note 42, at 1087.

explained the distinctiveness of constitutional provisions in Europe by the unique problems of the past that the respective jurisdictions wanted to deal with, e.g. Germany with regard to the failure of the Weimar Republic.\textsuperscript{46}

Concentrating on one generic function as the delimiting trait of the “social problem” under scrutiny consequently ignores alternative conceptualizations that could be established when choosing another function as the \textit{tertium comparationis}. Functionalism therefore runs the risk of miscomprehending or even overlooking the institution’s cumulative contribution to the respective legal system.

\textit{b. Second Premise: Similarity of Problems}

As our look at the roots of functionalism in the early twentieth century has shown, the method was not designed as a basis for all comparative studies. Especially its second presupposition, that problems are similar across legal systems, imposes severe operational limitations upon it.\textsuperscript{47} The implied universalism of this premise confines comparatists to dealing with problems defined in similar practical terms. As soon as one system attributes a different social significance to a particular problem, the similarity of function (and the comparability of “solutions”) ends.

Acknowledging this, functionalists frankly admit that there are “blind spots,” areas of the law, e.g., the law of wills or same-sex marriages, which are “system conditioned” to an extent so that they are beyond the reach of their method.\textsuperscript{48} Even if you accept this proviso, the functional method is deficit because it neither offers a precise definition of the term “system conditioned,” nor shows a way in which the comparatist may distinguish between “system conditioned” problems and those that are not.\textsuperscript{49} This shortcoming has a wider implication. If functionalism is not capable of dealing with individual “system conditioned” institutions, how can it investigate the socio-economic relativity of the legal system as a whole?

Another “blind spot” of functionalism is its lack of causal explanation. By definition, the method reverses the usual order of cause and (social) effect by explaining things in terms of what happens afterward, not what came before. This prevents functionalists, for example, from examining filial relationships between legal systems and institutions properly, be-

\textsuperscript{46} Bell, \textit{supra} note 43, at 241–42.
\textsuperscript{47} See Peter de Cruz, \textit{Comparative Law in a Changing World} 230–33 (2d ed. 1999); Husa, \textit{supra} note 7, at 424; van Reenen, \textit{supra} note 4, at 188–89.
\textsuperscript{48} Zweigert & Kötz, \textit{supra} note 12, at 39.
cause, as Alan Watson has demonstrated, cause/effect relationships between a transplant and its foreign antecedent are fundamental for studies in the migration of legal ideas.\textsuperscript{50} The general lack of structural-causal explanations forces functionalists either to reduce the explanatory claims of their theory or to deny all non-technical attributes of the law, neither of which they want.

Even in those areas in which functionalism can work, its second premise severely limits its operation by fostering reductionism.\textsuperscript{51} The functionalists’ reluctance to properly establish the comparability of the problem sociologically restricts their comparisons to grand, superficial similarities. Indeed, it is hard to believe that many legal problems are the same in two societies except on a technical level. To assume the opposite seemed “utter dilettantism” to Kohler.\textsuperscript{52} The underlying political, moral, and social values in different systems simply vary too much. Functionalists do not seem to realize this because they generally fail to discuss how one establishes “likeness” and “sameness” as a starting point for comparison. Further, they do not propose a method for finding and evaluating differences, however small, among “like” phenomena. Partly as a result of this, the functional approach is unable to solve the problem of apparently similar social and economic conditions producing radically different legal solutions, or even no solutions at all.

c. Third Premise: Problems Are Solved in a Similar Way

The heuristic principle of the praesumptio similitudinis, finally, is a further incentive to concentrate uncritically on similarities and thereby deepen the reductionist tendency of functionalism. It seduces them into neglecting the cultural-historical specificity of legal systems as long as, generally, their solutions to “problems” coincide. The assumption of similarity works reasonably well within the same cultural sphere. If comparisons, however, take place between culturally more remote systems, it becomes increasingly pointless, which the existence of institutes such as ordre public in the field of conflict of laws suggest.\textsuperscript{53} Further exemptions are necessitated by the problem, that there are institutions for

\textsuperscript{50} Alan Watson, \textit{From Legal Transplants to Legal Formants}, 43 AM. J. COMP. L. 469, 469–70 (1995).

\textsuperscript{51} From a historical perspective, see Watson, supra note 40, at 4–5; from a culturalist perspective, see Vivian Grosswald Curran, \textit{Cultural Immersion, Difference and Categories in U.S. Comparative Law}, 46 AM. J. COMP. L. 43, 61 (1998); from a critical legal studies perspective, see Frankenberg, \textit{Critical Comparisons}, supra note 20, at 436; van Reenen, supra note 4, at 191.

\textsuperscript{52} Kohler, supra note 35, at 275.

\textsuperscript{53} Graziadei, supra note 10, at 102.
which equivalents are not found in all systems, e.g., polygamous marriages. Some problems, such as marrying a second cousin, pose a problem for some legal systems, but not for others. That is why functionalists tend to apply the \textit{praesumptio} only to legal subjects that are relatively “apolitical” and unimpressed “by moral views or values,”\textsuperscript{54} such as contract law. But can we really assume that there is such a thing as “apolitical law” in an age where policy issues such as consumer protection permeate all fields of the law, including the law of contracts? Is it not that we have to accept that law is nothing else but successful politics?

Finally, the presumption of similarity questionably encourages the view that legal and extra-legal regulation are essentially the same, as long as they fulfill the same function.\textsuperscript{55} Yet, in legal theory, it makes a difference whether individuals are free to discover their obligations for themselves or whether their obligations are imposed by law. Ignoring this hides a vital issue of comparison: the way legal systems allocate regulation between law and custom.

2. Operational Problems

\textit{a. Pseudo-Factuality}

On a non-axiomatic level, functionalism is in deficit, too, by maintaining a “factual approach”\textsuperscript{56} to comparative law, while in fact not doing so. Functional studies begin by defining a social problem.\textsuperscript{57} A social problem, however, is a factual situation plus the value judgment that this situation causes consequences that need to be remedied. This value judgment is contingent. The answer to what makes a factual situation a problem can be different from one legal system to another. It may depend, for example, upon who finds a particular factual situation “problematic.” It makes a difference whether a particular situation in one legal system is considered problematic by a lobby group, in another by academics, and in a third by a large majority of the public. Functionalism does not care for this contingency. Therefore, its claim of neutrality, which hinges on the claim of being a factual approach, cannot be upheld.

\textit{b. Contemporality}

Another operational shortcoming of the functionalist method is that it is nearly exclusively occupied with studying contemporal legal problems

\textsuperscript{54} Zweigert & Kötz, supra note 12, at 40.
\textsuperscript{55} Id. at 38.
\textsuperscript{56} Peters & Schwenke, supra note 5, at 808.
\textsuperscript{57} See supra Part II.A.
It neither creates incentives to look at how a problem was solved in the past, nor does it care to compare legal systems or institutions that are remote from each other in time (vertical or diachronous comparisons). Though difficult and not universally exercisable, diachronous comparisons are possible as long as the institutions or legal systems under scrutiny have enough characteristics in common to validate a comparison. Other comparative disciplines, such as comparative politics, make frequent use of such comparisons. Studies like Buckland and McNair’s *Roman Law and Common Law* elucidate that diachronous comparisons also yield valuable knowledge for lawyers.

Not to examine how a particular legal system addressed a certain problem in the past is cumbersome, where this jurisdiction adopted a specific solution first and opted for another solution later. In such a case, functionalists will often neglect the former solution because it is discontinuous. Even where an institution stands the test of time, the functionalist will regularly overlook that it may have had additional or alternative functions in the past. Nearly every institution serves functions at some time in a given society that have been ignored in other times. It has, however, the potential to perform each of these functions even if in the given concrete contemporal case, it does not do so. This potential is important for understanding the significance of a legal institution, which functionalists will regularly fail to notice.

**E. Conclusion**

The pragmatically motivated functional method has been a useful guide for establishing comparative law as a discipline. It still has its virtues and valuable applications today. As a model for all comparative studies, however, functionalism is no longer a good fit because it is too limited in its application by its premises and operational problems. Merely abandoning the *praesumptio similitudinis*, as arguably its most

---

58. Constantinesco, Tome III, supra note 36, at 42.
59. Gerber, supra note 17, at 206.
controversial presumption, cannot save the method. The objections to the remaining presuppositions are too serious.

III. ALTERNATIVE APPROACHES

Towards the end of the twentieth century, the growing dissatisfaction with functionalism enabled new approaches to comparative law to challenge its claim for methodological dominance. Three of these trends are worth mentioning: Comparative Law and Economics, Comparative Law and Culture, and Critical Comparative Law. They shall be examined in turn to see whether they can present a viable alternative to functionalism. I will not present historical comparative law as a separate method because the historical dimension of law is, in fact, an integral part of every meaningful comparative study.

A. Mattei—Comparative Law and Economics

While comparative lawyers increasingly employ disciplines such as history, anthropology, and sociology, until recently economic theory has been largely ignored. Only in the mid-1990s did Law and Economics claim a share in comparative studies. It promises to benefit comparative research by providing a degree of measurement to its statements: economic efficiency. Comparative Law and Economics scholars strive to distinguish themselves from mainstream law and economics insofar as they claim a more neutral role for the discipline as an analytical, rather than a normative tool. This, however, is merely an ill-concealed lip service. As their studies reveal, in fact, they aim to operate both at the levels of descriptive and normative analysis.

1. Operation

Analytically, Comparative Law and Economics seeks to begin the comparison from a “neutral scale” that can be validated by observable data: economic efficiency. The term “efficiency,” in its comparative sense, is defined as “whatever legal arrangement ‘they’ have that ‘we’ wish to have because by having it they are better off”—in terms of lesser waste, lower transaction costs, better resource allocation, or greater

63. Husa, supra note 7, at 424.
64. ZWEIGERT & KÖTZ, supra note 12, at 8. See Örüçü, supra note 15, at 63–65.
67. Mattei, supra note 65, at 145.
freedom for individuals to interact. Mattei’s archetypal methodology begins by building a model of what he hypothesizes is an efficient legal institution. This blueprint, according to him, needs to be abstracted from the pool of solutions offered by existing legal systems under the presumption that all of the relevant empirical data for the assessment of efficiency are available.

In a second analytical step, law and economics comparatists then compare their model to the real-world alternatives (i.e., substantive rules) of different legal systems. When faced with departures from the efficient model, which will frequently be found, they seek to explain why this inefficiency occurs. A proper analysis is very complex because law and economics comparatists accept that efficiency is context-dependent. Even where market structure and consumer preferences are similar, the same legal rule may be efficient or inefficient depending on the institutional and cultural background to which it refers. Less efficient solutions can be justified by offering some non-distributional benefit that outweighs the gains that the efficient solution would have generated. In judging the overall efficiency of a legal system, the comparatist furthermore has to watch out for institutions that can work as efficiency restoring substitutes, where inefficiency is diagnosed. In this way, comparative scholars can isolate and evaluate variables that contribute to or detract from the relative efficiency of the systems under comparison.

The third step of Comparative Law and Economics transcends from the analytical to the normative level. After having identified and explained the deviations from the efficient model, the comparatist is supposed to define the conditions for policy changes in order to get closer to the model in those instances where the reasons for distance do not appear justified. Such conditions are determined by the transaction costs of changing a given historically prefixed routine. For example, while certain aspects of trust law, as developed in the common law tradition, have efficiency advantages in the case of bankruptcy over their counterparts available in civil law, the costs of changing the general structure of prop-

70. Mattei, supra note 65, at 182.
71. See Krimphove, supra note 68, at 191; Mattei & Cafaggi, supra note 66, at 347.
72. Mattei & Cafaggi, supra note 66, at 347.
erty rights to accommodate trusts might outweigh the benefits of such an introduction.  

As we can see by this example, Comparative Law and Economics focuses on developments in the relationship between legal systems, i.e., convergence, divergence, and the occurrence of legal transplants. It seeks to explain convergence and divergence as a result of competition between legal systems. Legal systems are believed to function as markets for the supply of different solutions for a specific problem. If transaction costs were zero, then law would be freely transplantable (free movement of legal rules) and would evolve naturally toward the most efficient rule. Legal diversity (transplantation resistance)—according to Comparative Law and Economics—results from the transaction costs of tradition, culture, and ideology.

2. Development

Law and Economics originated in the late 1950s in the United States, and dominated the legal discourse there during the 1960s and ’70s under Coase and, later, Posner. Progressively it found its way into other countries. Its appeal “at home,” however, began to wither when Ugo Mattei and Dieter Krimphove paved the path for its application to comparative studies in the 1990s. For a proper evaluation of Comparative Law and Economics, it is important to note that the intellectual history of this variant of law and economics is curiously longer than that of its parent. Technically, Comparative Law and Economics, in its search for the “efficient solution of a given problem,” is a narrowed and specified version


76. SMITS, supra note 69, at 69.

77. Mattei, supra note 65, at 121.


79. See, e.g., Mattei, supra note 65; Krimphove, supra note 68.
of functionalism.\textsuperscript{80} It “radicalizes” this method by focusing on one particular function only: the rule’s or institution’s efficiency.\textsuperscript{81}

The second source of inspiration for Comparative Law and Economics is the work of Gustav Radbruch. It was Radbruch who first maintained that when comparing two \textit{comparanda}, this could only be by reference to a third, constant element.\textsuperscript{82} According to Radbruch, this common point of reference has to be a supra-national legal system, an objective, “higher” or “natural” law ("\textit{richtiges Recht}"). For adherents of Comparative Law and Economics, the “efficient” rule is this higher or natural law.

### 3. False Trails of Criticism

Naturally, not everybody agrees on the existence of such a kind of “higher law.” A frequent reservation against Comparative Law and Economics is that it relies on simplistic presuppositions drawn from neoclassical economics.\textsuperscript{83} This criticism is worthwhile, but it only disqualifies Comparative Law and Economics based on neoclassical models and not the method as such. The neo-institutional movement in economics allows for observing the institutional settings of human interaction and their impact on transaction costs.\textsuperscript{84} Behavioral Economics employs social sciences to rectify the inaccurate assumptions in traditional “Law and Economics” models by adopting a more realistic idea of man.\textsuperscript{85} Consequently, there are viable alternative theoretical foundations for Comparative Law and Economics.

\begin{itemize}
\item \textsuperscript{80} Ewald, \textit{Posner’s Economic Approach}, supra note 30, at 383; Peters & Schwenke, \textit{supra} note 5, at 808.
\item \textsuperscript{82} Gustav Radbruch, \textit{Über die Methode der Rechtsvergleichung}, 2 \textit{Monatsschrift für Kriminalpsychologie und Strafrechtsreform} 422, 423 (1906) (F.R.G).
\item \textsuperscript{83} Catherine A. Rogers, \textit{Gulliver’s Troubled Travels, or the Conundrum of Comparative Law}, 67 \textit{Geo. Wash. L. Rev.} 149, 186 (1998); see also Mackaay, \textit{supra} note 78, at 86.
\item \textsuperscript{84} Mattei & Cafaggi, \textit{supra} note 66, at 347.
\end{itemize}
4. Real Problems

There are, however, limits to Comparative Law and Economics, which disqualify it as a central method for comparative studies. The most fundamental objection is that—as its development has shown—it is a mere variant of functionalism. That makes Comparative Law and Economics vulnerable to the same objections put forth against functionalism in its broader sense. Furthermore, its focus on efficiency as the sole function of law invites additional criticism.

Mattei’s idea of a neutral model as a tertium comparationis is appealing, not because a defined, objective point of reference is regarded as a logical necessity of comparative methodology as Radbruch (erroneously) believed, but because it “purges” the comparative process from preconceptual bias. In order to be neutral and have such a purifying effect, the comparatum must not be loaded with preconceptions itself. However, it is doubtful whether that is true for “efficiency.”

a. Ambiguity

A first problem with this term is its ambiguity. Efficiency can refer to “partial equilibrium solutions,” i.e., pursuing an efficient outcome for a particular problem in a particular market; but it can also mean “general equilibria,” i.e., efficient solutions for an entire economic system. The comparatist is not told which of these scenarios to rely upon when building “efficient models.” The reason for this might be that both equilibria are indeterminate within themselves. Efficient solutions in partial equilibrium situations cannot be defined unambiguously for logical reasons because they are path-dependent and may become inefficient once one takes into account third parties or collateral effects to other sectors. To seek general equilibrium efficiencies is technically impracticable for the

86. Brand, supra note 42, at 1087; Michaels, supra note 7, at 108; Peters & Schwenke, supra note 5, at 809.


88. Rogers, supra note 83, at 185.


comparatist because of the enormous information requirements that it
would place on him or her.91

b. Non-Neutrality

Even worse, efficiency as a criterion is not only ambiguous, but also
partisan. Efficiency analysis is essentially a dynamic cost-benefit analy-
sis. It does not look statically at a certain provision and asks whether this
rule is “efficient.” Instead, it examines collective decisions (e.g., a
change in legal rules) and asks, under the predominant Kaldor-Hicks
test,92 whether they generate sufficient gains to their beneficiaries so as
to hypothetically compensate the losers and render the latter fully indif-
ferent to the change but still have some gains left over for themselves.
This test contains the seed of non-neutrality in the form of the so-called
offer-asking problem. Kelman was the first to observe that people gener-
ally have a greater concern for and attachment to things as they are com-
pared to things as they could be.93 As a result, people will ask for a
higher price when they have to give up something (asking price) than
what they would be willing to pay when bound to acquire the same good
(offer price). This difference in price matters to efficiency because it ren-
ders the determination of the respective gains and losses of winners and
losers dependent on hidden value judgments. The offer-asking problem
even creates room to argue about who is a winner and who is a loser of a
proposed change. At first blush, the application of the offer-asking prob-
lem to efficiency’s test—that winners must be able to bribe losers—
would measure the bribe at its offer price and the loss at its asking
price.94 Other law and economics scholars, however, have interpreted
win and loss just the other way around.95 Neither position can be said to
be wrong. The choice to measure win and loss at the offer price or the
asking price is a question of perspective and correspondingly a question
of policy: the winners from change are the losers from no-change. If effi-

91. Duncan Kennedy, Cost-Benefit Analysis of Entitlement Problems: A Critique, 33
STAN. L. REV. 387, 395 (1981); Rizzo, supra note 90, at 641–42.
92. For the test and the rival Pareto test, see MICHAEL J. TREBILCOCK, THE LIMITS OF
93. Mark Kelman, Consumption Theory, Production Theory, and Ideology in the
94. Thomas C. Heller, The Importance of Normative Decision-Making: The Limita-
tions of Legal Economics as a Basis for a Liberal Jurisprudence—As Illustrated by the
Misallocation: A Checklist for Micro-Economic Policy Analysis, 28 STAN. L. REV. 1, 2–3,
n.4 (1976) (measuring the winners’ gains at their asking price and the losers’ losses at
their offer price).
ciency therefore inevitably involves value judgments, it cannot serve as the neutral tertium comparationis that Comparative Law and Economics requires it to be.

Another expression of the inherent dependency of efficiency on value judgments is the problem of multiple optima. Comparative Law and Economics suggest an efficient model as comparatum against which to assess real-world legal institutions. Such a model cannot be built with reference to efficiency alone. If transaction costs are low, i.e., almost any agreement that is to the mutual benefit of the parties concerned is made, any assignment of rights will lead to an efficient outcome. Winners and losers of a proposed change in the law would always bargain for the efficient solution.96 This means that there is a set of efficient solutions, rather than a single efficient outcome. The choice of any of them is a matter of value judgment unrelated to efficiency. The situation is no different when transaction costs are numerous and/or high. Here, it is the lawyer-economist’s task to manipulate entitlements and redistribute units of factors until the allocation of resources resembles the efficient solution found in a world with no/low transaction costs.97 This is impossible to achieve with reference to efficiency alone, because, as we have just seen, even in a world with no transaction costs there is no single efficient solution. Therefore, under no circumstances, can law reformers use efficiency alone to support any particular program of rules.

Even proponents of a law and economics approach to comparative law are reluctantly realizing how political and value-laden efficiency in fact is. Mattei, for example, complains about a severe “American-centric” provincialism in the discipline.98 Indeed, law and economics scholars mostly work under the rather uncritical assumption of the American institutional background; they seem to assume that “there is, whether by conscious choice or by social necessity, a strong tendency for the [American] common law to adopt efficient rules.”99 Mattei tries to tackle the problem by establishing the context-dependency of efficiency. If efficiency, however, is context-dependent, there is no way to work with it as a neutral tertium comparationis.

96. DAVID D. FRIEDMAN, LAW’S ORDER 39 (2000); Kennedy, supra note 91, at 395, 441–42.
97. Kennedy, supra note 91, at 444.
98. MATTEI, supra note 65, at xii, 69–70.
99. Id. at 125.
**c. Distortedness**

Efficiency is also unsuitable as a scale for comparative studies because it distorts the perspective of the comparatist in two ways. First, efficiency is a transient phenomenon. Even slight changes in the economic circumstances invalidate findings because what is considered to be efficient in today’s economic environment might be inefficient in that of tomorrow. Therefore, efficiency is oriented toward short-term results. While this might be acceptable in a (reactive) common law jurisdiction, it does not agree with legal systems based on codification with their higher need for legal certainty. Secondly, efficiency analysis is purely concerned with resource allocation. Distibutive issues, i.e., all non-efficiency considerations, are ignored or marginalized. Such a focus on an exclusive criterion under which to evaluate laws is a decidedly reductionist view of the law and its role in society. It breathes new life into Gerber’s fear that functional comparisons of any kind produce particularist knowledge. Opting for Comparative Law and Economics instead of functionalism, would therefore change things from bad to worse.

**B. Legrand—Comparative Law as Hermeneutic Exercise**

With Comparative Law and Economics apparently unable to overcome the objections against functionalism, we have to turn our eyes to other contenders. Since the mid-1990s, conventional comparative law has been challenged by members of the culturalist movement. Their thinking orbits around the term “legal culture,” which means “those historically conditioned, deeply rooted attitudes about the nature of law and about the proper structure and operation of a legal system that are at large in the society.” From this starting point, culturalists essentially contend that legal rules are embedded in local dimensions of the law. Each legal culture is a unique, culturally contingent product, which is incommensurable and untranslatable except through a deep understanding of the surrounding social context. Pierre Legrand is the prime proponent of this movement in comparative law.

---


1. Operation

While functionalism is concentrated on finding similarities and convergences, Legrand’s basic experience is that of plurality and difference. He argues that comparative law is not a search for function, but a hermeneutic exercise (démarche herméneutique). Functionalism, according to his view, was only partially successful in penetrating the façade of language. What a rule does “functionally,” he believes, is yet another layer of the surface appearance of law. The task of the comparatist is to delve beyond that technical surface and to uncover what the rule signifies in terms of its political, social, economic, and ideological context.

For Legrand, the specificity of legal traditions and cultures is central. He sees their structures and values as contained in language, which is not always translatable, and builds therefore an insurmountable framework of contingent ways of legal reasoning. Comparative law is about “finding what is significantly different” abroad. Accordingly, it can only be descriptive, not normative. Foreign law may only be used as a tool to encourage criticism of the presuppositions of one’s own tradition, but not to provide a model for its reform because each legal tradition is requisite and irreplaceable. Methodologically, the explanation of the deep cultural and mental structures expressed by legal texts becomes the main aim of comparative law. This “hermeneutic exercise” is, in fact, a search for the cultural, moral, and linguistic relativism of law. How that search is to be conducted, however, is merely hinted at. After having established a linguistic framework that at least allows a proper perception of foreign law, the comparatist has to carefully lift in his or her description the latter’s cultural veils one by one, paying special attention to decision making structures, i.e., legal actors and the way they interact and reason.

Legrand can only find limited applications for comparative law because of his focus on perspective. He is “anti-foundationalist,” i.e., rejects the belief that there is a universal truth. Instead, he claims human knowledge can, at best, be partial in nature and only validated within a specific context; the comparatist, in other words, cannot escape from his

or her cultural framework. This belief, that not only law, but also all knowledge is culturally and historically contingent, poses an immense methodological hurdle for comparative studies.

2. Development

A look at the roots of Legrand’s approach provides an explanation. Historically, it is based on forethinkers, who were biased against comparative law, and philosophically, it is rooted on movements that are weak in methodology-building.

When Legrand regards legal cultures as unique “spiritual creations” of the community, he reaches back to Montesquieu, who proclaimed the dependency of law on local conditions as early as 1748 with his seminal *De l’Esprit des Lois*. Legrand can also be seen as an unintentional successor of the German historical school of law that considered law to be the manifestation of the people’s national spirit (“Volksgeist”) and thereby particular to every nation—an organic product of society which has to be watched for and discovered, rather than made or tampered with. Even though the contribution of its founder, Savigny, to comparative law is underestimated, this school was more than ambivalent towards comparative legal studies.

Theoretically, Legrand’s works are based on two different schools, culturalism and deconstruction. The banner of culturalism was raised in the early 1980s by Stuart Hall as a countermovement to structuralism and universalism. The main objective of culturalism was to allow societies to be evaluated in their uniqueness and their complex social and political context. The dark side of this perspective is that societies, like individuals, all of a sudden seemed incommensurable if not incompatible.

Accordingly, the diversity of cultures became sacrosanct. Research therefore had to diversify as well, often to such an extent that a fruitful dialogue was hardly viable anymore.

Culturalism found a natural ally in deconstruction, which in turn stands on the shoulders of hermeneutics. The latter is concerned with human understanding and the interpretation of texts. Gadamer, one of Heidegger’s disciples, first applied hermeneutics to legal texts in order to understand them as signifiers of something deeper. The French poststructuralist philosopher Derrida, whom Legrand refers to frequently, also borrowed from Heidegger and transformed his form of textual analysis (“Dekonstruktion”) into a tool of “destructive criticism.” Central to deconstruction is the French neologism “différance.” In simple terms, this means that rather than seeking commonality, simplicity, and unifying principles, deconstruction emphasizes difference, complexity, and non-self-identity. A deconstructive interpretation of law aims to unveil hidden inconsistencies and biases by demonstrating how a seemingly unitary concept contains different or opposing meanings. According to Derrida, however, deconstruction never condensed to either a school or a method, but is merely an occurrence within the text itself.

3. Problems

The upshot of Legrand’s “method” is total incomparability across history and culture. He simply negates the mere existence of things such as legal transplants or convergence of legal systems. Any cross-cultural comparisons are labeled “superficial.” One wonders whether that is just over-conscious thinking or, in fact, the expression of an agenda to pose as many obstacles to the normative use of comparative law as possible. Cultures are, as Legrand admits, not hermetic, closed, or immutable entities. They influence and infiltrate each other (Americanization!), because unlike individuals, they do not have readily determinable boundaries. If boundaries between cultures are blurry, then those of their

\begin{itemize}
  \item 114. See id. at 68.
  \item 117. J.M. Balkin, Deconstruction, in A COMPANION TO PHILOSOPHY OF LAW AND LEGAL THEORY 367, 368–69 (Dennis Patterson ed., 1996).
  \item 118. Id. at 367.
  \item 119. Legrand, The Impossibility of “Legal Transplants,” supra note 103, at 111–12.
  \item 120. Pierre Legrand, European Legal Systems Are Not Converging, 45 INT’L & COMP. L.Q. 52, 52 (1996); Legrand, How to Compare Now, supra note 28, at 295.
  \item 121. Legrand, How to Compare Now, supra note 28, at 238, n.26.
\end{itemize}
epistemic and socio-economic background are too. That makes a fruitful dialogue between members of a different cultural background possible.

Certainly, Legrand and other culturalists have commendably elucidated shortcomings of functionalist studies in paying due respect to the socio-cultural context of law, drawing “destructive” strength from their philosophical roots. However, this potential might be exhausted. As pointed out above, functionalists today accept that the comparative study of law needs to be contextual. Because they also can accommodate this consideration within their theoretical framework, it seems as if all that can be done with Legrand’s approach meaningfully, is to “deconstruct the ambiguities and indeterminacies” within individual comparisons.\(^\text{122}\) Otherwise, Legrand’s comparative thoughts have little to offer for actual research-practice in comparative law. “Methodologically,” he wants to contextualize the objects of comparison and thereby capture their essence as a unique manifestation of the community. This, however, is more a goal than a method. One cannot avail oneself of the idea that Legrand, like most of the law-as-culture scholars, is still working on a “proto-methodological” level. He is preoccupied with contemplating the legitimacy and the aims of comparative law. In his “*Droit comparé,*” for example, Legrand concentrates on philosophic, epistemological problems besetting the discipline. He devotes no space to the history, accomplishments, or methods of the subject. The question of “how to compare now” dawns, but it is still superimposed by the question of “why to compare.” It is not unlikely that Legrand will never emerge from this state. His theoretical sources, culturalism, and deconstruction hardly provide him with workable tools for method-building.

**C. Frankenberg—Critical Comparative Law**

A similar fate seems to await Guenter Frankenberg, one of the few scholars associated with Critical Legal Studies who pays closer attention to questions of comparative methodology.

1. **Operation**

   For Frankenberg, the question of method encompasses not only the “how,” but also the “why” of comparative studies.\(^\text{123}\) He negates functionalism’s presumptions of the necessity, functionality, and universality of law and aims to transform conventional comparative law into a tool for the critique of law by using an analysis of abortion decisions to illus-

\(^{122}\) Husa, *supra* note 7, at 441–42; Peters & Schwenke, *supra* note 5, at 812; Riles, *supra* note 20, at 248.

\(^{123}\) Frankenberg, *Critical Comparisons, supra* note 20, at 416, 445.
trate his approach. This approach divides the comparative process into three steps.\(^{124}\) In the first step, the comparatist is required to scrutinize carefully what happens when the multiple facets of a factual situation are taken out of their social context and fitted into a legal framework. A second step is to elucidate this legal framework by critically analyzing its structure (especially the public/private distinctions) and to delve into the processes of legal decision-making. The latter opens the door to the political dimension of law, which Frankenberg sees primarily as a theoretical instrument for the purpose of gaining, cementing, and justifying the exercise of power.\(^{125}\) He thinks that comparatists need to move from traditional conceptions of legal discourse, such as rights and duties, to the politics of the subject studied. For him that is the only way to insure that comparative law does not obey hidden political agendas of hegemony and domination. In a third step, finally, the comparatist has to reintroduce the socio-cultural context that has been lost by “legalizing” a problem.

To avoid the bias of functionalism, Frankenberg wants comparatists to understand their studies as “learning experience[s]” that require “a greater sensitivity to the relationship between the self and the other,” and “tolerance for ambiguity.”\(^{126}\) They are encouraged to avoid concentrating on similarities and to allow differences, especially in the political context, to emerge. To cope with pre-conceptual bias, Frankenberg wants to unravel the ties that bind the comparatist to his domestic legal regime. Comparative research undertaken so far, he believes, has to be reassessed by taking into account the scholars’ motives, interests, and perspectives, i.e., the scholars’ legal education, exposure to different legal cultures, networks, etc.\(^{127}\) Self-criticism as the prime virtue of the researcher is meant to insure that future comparative law produces valuable knowledge. Only then can the comparatist unearth the sub-structural, often unarticulated, categorizations and silent assumptions of law.\(^{128}\)

2. Development

Frankenberg’s insights have two sources of inspiration. First, like LeGrand, he is driven by the idea of difference that in Frankenberg’s case is loosely inspired by feminist studies and the philosophical theory of literature. The theme of difference harmonizes well with Frankenberg’s second source, the Critical Legal Studies movement. This rather hetero-

\(^{124}\) Id. at 451–52.


\(^{126}\) Frankenberg, *Critical Comparisons*, supra note 20, at 441.

\(^{127}\) Id. at 443.

\(^{128}\) Id.
geneous school, whose crystallization point was a conference at the University of Wisconsin-Madison in 1977, applies ideas of Marx, Marcuse, and Adorno to the study of law. Of central importance is the idea that legal rules and institutions are tilted towards the preservation of entrenched interests with wealth and power and therefore are biased against the poor and oppressed, especially the working class, women, and people of color. Critical Legal Studies flourished in the United States in the 1980s, but its influence began to fade in the early 1990s.

3. Problems

Some scholars sense a “thirst for a more aggressive and dynamic comparative law,” such as Frankenberg’s approach. However, his insights have so far not helped to establish a more convincing methodological approach to comparative studies than the heavily criticized functionalism. Like Legrand, Frankenberg is good at analyzing weaknesses of conventional comparative studies and at further reducing their reach and explanatory power. In addition, his elaboration on the political dimension of law and the need for self-reflection provides helpful guidance for the comparatist—even though it may divert the focus of research from the comparison of laws to the history, epistemology, and politics of comparative research itself. Again, like Legrand however, Frankenberg does not muster the strength to come up with a fully developed counter-theory to functionalism. His “three step approach” remains “patchwork”—mainly because the theoretical framework he is working in, Critical Legal Studies, does not provide the necessary tools for theorizing. It confines its adherents to the critical analysis of pre-existing institutions, rules, and theories. Correspondingly, Frankenberg leaves us in a kind of Socratian aporia: our belief in functionalism is shattered, but no replacement is offered.

130. Husa, supra note 7, at 420.
4. Conclusion

Parts II and III have drawn a rather gloomy picture of comparative legal methodology as a garden filled with withering or infertile flowers. It seems that functionalism is still dominant because of a lack of alternatives, not because it is particularly well-suited. Apparently, there is need for a fresh sapling. But what qualities would it have to bring? The analysis of functionalism has demonstrated negatively that whatever methodology we want to adopt for comparative studies must not work under the three presumptions that: law answers to social needs; problems are at least similar across legal systems; and these problems tend to be resolved in the same way. Positively, Comparative Law and Economics suggested that model building might be a promising alternative—“efficiency” is just not the neutral scale required. Cultural Deconstruction and Critical Legal Studies, finally, did not offer much methodologically, but provided at least hints for avoiding pitfalls by emphasizing the need for a contextual approach.

IV. OWN APPROACH—CONCEPTUAL COMPARISONS

This Part tries to develop a method for comparative studies that serves the needs of the comparatist without being subject to the criticism outlined above. Section A is devoted to preliminary observations, section B outlines the operation of the method, and section C suggests some applications. In sections D and E, I will discuss advantages of the proposed method and anticipate likely objections to it. As the method uses models or concepts as comparata, I shall call it “Conceptual Comparisons.” Wherever possible, the terminology used in describing my approach is adapted to the usage in other social sciences that work comparatively with concepts to enable and foster an interdisciplinary dialogue. To illustrate how Conceptual Comparisons works, I will discuss the second-tier protection of inventions in Europe as an example.

A. Preliminary Remarks

1. Presuppositions

Conceptual Comparisons rejects the presumptions upon which functionalism works. However, it cannot operate without three presumptions of its own. Firstly, law, as the object of comparison, is understood as a normative system that consists of principles, rules, institutions, and other institutionally defined instruments. The conceptual structure employed

133. A similar presumption is used by Banakas, supra note 49, at 306.
by legal systems is important in ordering legal understanding and ensuring that like cases are decided alike. Its effects must not be marginalized, as functionalism does and as culturalists and critical scholars do to an even greater extent.

Secondly, Conceptual Comparisons acknowledges that impurities in the comparative act are inevitable, mainly because comparatists cannot escape the preconceptions of their own legal culture and education. However, while it is true that the perspective on law and accordingly the work of individual scholars are inevitably subjective, the process by which legal institutions are compared is not necessarily subjective, too. Conceptual Comparisons shares the conviction of Comparative Law and Economics that a neutral tertium comparationis as an “expurgatory tool” is not only desirable, but also constructible. Objectivity is envisaged to derive from combined efforts of comparatists using a common system of reference for mutual scrutiny and criticism of each other’s work. Unlike Comparative Law and Economics, however, Conceptual Comparisons endeavors to use this reference system purely analytically.

Thirdly, Conceptual Comparisons presumes that it is possible to formulate a neutral reference system in the form of concepts. By concepts, I mean abstract models derived in an inductive process from specific instances of real-existing law. Culturalists and critical legal scholars, however, believe that all types of categories and classifications are suspicious because they are culturally contingent. Outsiders are said to be unable to understand foreign classifications and translate them into their own. Indeed, domestic categories and classifications vary in different legal systems, at different times. They are neither universal nor neutral. That, however, only out-rules real law as a neutral tertium comparationis, but not purely theoretical, abstract models that dwell “in between” the existing legal systems.

2. Conceptualisation

Abstract models successfully serve as neutral analytical tools in the natural sciences. Fortunately, conceptualization is firmly established in the social sciences, too, and accepted by analytical jurisprudence as

136. Curran, supra note 51, at 48; Frankenberg, Stranger than Paradise, supra note 81, at 267.
137. See KARL ENGISCH, DIE IDEE DER KONKRETISIERUNG IN RECHT UND RECHTSWISSENSCHAFT UNSERER ZEIT 238–39 (1953) (F.R.G); see, e.g., Arthur L. Kalleberg, The Logic of Comparison: A Methodological Note on the Comparative Study of
crucial to our perception and understanding of law. It is the basic tool for lawyers to communicate with each other and to transfer knowledge from one area of the law to another. However, an analogy to the concepts and categories of the natural sciences is only possible to a limited extent. The latter rely on properties that can be verified empirically. Accordingly, the categories of natural sciences can be construed so that they are mutually exclusive: an animal is either a vertebrate or an invertebrate. In social sciences, the properties used for concept construction have their source in theoretical discourse, rather than in observable matter; concepts can be cumulative. A particular legal institution can belong to more than one category at the same time. A compulsory license, for example, can be classified (by substantive matter) as a limitation of an intellectual property right or (procedurally) as a defense in an infringement action. That is why concepts in social sciences have to be more fluid, flexible types, rather than schematic terms. To diminish the “lack of observability” in law, the comparatist is well advised to scrutinize legal institutions not in the abstract, but based on factual situations.

Comparative law is not unaware of conceptualization. Classifications dominate macro-comparisons. In micro-comparisons, models have been used occasionally in order to give the comparatist some point of reference, according to which he can position legal phenomena in relation to one another. Moreover, nineteenth century comparative law...
dreamed of an ideal system, a network of universal archetypes (“universalities”) that all existing legal system should approximate. Closer scrutiny of the operation of legal systems and their extra-legal connotations, however, has shattered this dream. Today we accept that there is no single set of archetypes that can accommodate the plenty of legal ideas across the globe. Assuming the opposite necessarily leads to bias and misconception.

3. Overview

Conceptual Comparisons seeks to establish an approach to conceptualization that allows expression of the variety of conceptualizations in different legal systems. The method draws inspiration from typology—especially typological comparisons—comparative methods of other social sciences, and plant taxonomy. It operates in two phases. In the first phase (conceptual orientation), the researcher construes certain elements of legal reality in logically precise, abstract, and unambiguous models (comparative concepts). In the second phase (systematic comparison), real-world institutions and rules can be matched and assessed against these concepts. The ultimate, admittedly Herculean, goal of Conceptual Comparisons is to establish a comprehensive network of concepts covering all legal institutions from all jurisdictions and to assess how these different concepts complement each other or conflict.

The two phases of Conceptual Comparisons are two separate comparative processes. Conceptual orientation requires—as we will see—considerable preliminary comparative investigation. The following systematic comparison between the concept and the rules or legal institutions from the chosen legal systems is not only supposed to provide a


144. See Raymond Saleilles’ droit commun idéal in R. Saleilles, La Fonction juridique du Droit comparé [The Juridical Function of Comparative Law], in JURISTISCHE FESTGABE DES AUSLANDES ZU JOSEF KOHLERS 60. GEBURTSTAG 164 (Fritz Berolzheimer ed., 1909) (Fr.) and Édouard Lambert’s “droit commun legislative” in ÉDOUARD LAMBERT, LA FONCTION DU DROIT CIVIL COMPARÉ [THE FUNCTION OF CIVIL COMPARATIVE LAW] 922 (1903) (Fr.), which are the most prominent versions of this approach; see also Kohler, supra note 35, at 275; Radbruch, supra note 82, at 423.

145. See CONSTANTINESCU, TOME II, supra note 40, at 66; GEORGES SAUSER-HALL, FONCTION ET MÉTHODE DU DROIT COMPARÉ [FUNCTION AND METHOD OF COMPARATIVE LAW] 54–64 (1913) (Fr.); WATSON, supra note 40, at 12–13; Großfeld, supra note 35, at 337; Michaels, supra note 7, at 100–01.

detailed analysis of communalities and differences, but shall also lay bare the underlying determinants of the legal phenomena under scrutiny, including their historical and cultural dimensions. The two processes, however, cannot be conducted strictly one after another. They have to be mutually adjusted, necessitating the comparatist to shift back and forth in a “hermeneutical circle”—writing up the conceptual orientation to verify and refine it later when he or she has produced the systematic comparison and vice versa. For the sake of simplicity, I will refer, in the following sections, only to the comparison of legal institutions. The comparative process as such applies *mutatis mutandis* to rules and principles as well.

B. Operation

The conceptual comparatist begins his or her study by examining a factual situation in a particular legal system. Unlike the functionalist, the conceptual comparatist disregards, at this stage, to what extent these facts amount to a social problem. The situation can be hypothetical, but it should be for reasons of verifiability, preferably one that has arisen in practice. The research question put to the fact is, which institutions address the facts under scrutiny? Beginning with one legal system, the respective institutions are “carved up” into their properties, i.e., their structure and consequences according to the methods of the particular legal system. The merit of this approach is that it honors the dogmatic individuality of the legal systems under scrutiny.

1. First Phase: Conceptual Orientation

After having broken down the initial institution into its components, the comparative process begins. The main objective of the first phase of Conceptual Comparisons, the conceptual orientation, is to establish a standard for comparability in the form of a concept. Conceptualization renders phenomena comparable by putting them in a common context. The key question is how to establish a comparative concept in a meaningful way—regardless of whether the comparatist is working with pre-existing concepts or is developing new ones. Pre-existing concepts must never be passively accepted. They have to be reassessed critically—and potentially reformulated—in the same way the researcher would have constructed a new concept: the resulting comparative concepts must fulfill six criteria in order to serve as a neutral *tertium comparationis*:

---

147. *See also* Chodosh, *supra* note 2, at 1091–93.
1. They have to be appropriate to the theoretical questions posed by the comparatist.
2. They must be effective, i.e., address the factual situations chosen for comparison. Accordingly, concepts must not be characterized by disorganized or trivial properties.
3. The properties used for concept construction must not be context-dependent.
4. They have to be named unambiguously.
5. They should be construed so that they vary as little as possible over time.
6. Finally, their construction has to be falsifiable. A concept is useless when no statistical or other evidence can be obtained in order to review it.

In order to fulfill these requirements, it is not advisable to design concepts a priori and apply them afterwards in a deductive process to real-world legal institutions. In that case, a concept might be construed too ideally and lack any real-world application. Political scientists, such as Theodore Becker for example, once defined the concept of “court” for a comparative study abstractly by seven characteristics.¹⁴⁸ Shapiro rightly criticized this approach because he could hardly find anything real to subsume under the entirety of these criteria.¹⁴⁹ Accordingly, concepts have to be construed by abstracting common elements from observed phenomena in a number of given legal systems. This inductive process could look like this: the comparatist should apply the initial facts, tested against the rules of one legal system, to another legal system and determine again the institution(s)/rule(s) that address the situation. After having done this with a number of legal systems, the search for communalities in the properties of the institutions/rules invoked begins. These communalities form the bones of the comparative concept.

To gain meaningful properties for the construction of the concept, the abstraction process has to be exercised on two different levels, a qualitative and a quantitative one. This two-fold analysis is necessary to unveil appropriately the legal strategies of different legal systems to deal with a particular factual situation. The qualitative analysis of a concept tells

¹⁴⁹. MARTIN SHAPIRO, COURTS: A COMPARATIVE AND POLITICAL ANALYSIS 1 (1981); see also Chodosh, supra note 2, at 1107–08.
which requirements a particular strategy has whereas the quantitative analysis reveals which factual situations a strategy can address. In this way, Conceptual Comparisons becomes an analysis of capability. Like owners of a Swiss Army knife who seldom if ever use the tools of their knives in their entirety, not all jurisdictions employing a certain strategy might be aware of which potential is inherent in the tools they use. Only a comparison with other legal systems that employ similar tools can disclose this potential.

Both the qualitative and the quantitative analysis look at the structure and the consequences of a given institution or rule. The latter’s “function” and history, at this stage, are irrelevant. They are the factors that render legal institutions context-dependent and therefore cannot be used in formulating a tertium comparationis. Due respect to them will be paid in the systematic comparison.

a. Qualitative Analysis

The qualitative analysis establishes the intension of a concept: the key attributes that define it, delimit the concept from other concepts, and determine membership. It is an empirical search for common, cross-sectional qualities in several legal orders. Each concept is characterized qualitatively by the fact that its members share at least one property that is not found in another concept. Properties of legal institutions or rules can be differentiated into typical, defining ones (a, b, . . .) and accompanying properties that are peculiar to individual institutions (α, β, . . .). For the construction of concepts, only defining properties are used; accompanying ones are disregarded. A comparative concept established in this way is analogue to a species in plant taxonomy\(^\text{150}\) and to a type in social sciences—what Engisch called an average type (“Durchschnittstyp”),\(^\text{151}\) based on the characteristic properties of an average member of a group, rather than an ideal type (“Idealtyp”) of the kind Weber regarded as the basic comparative unit.\(^\text{152}\) Ideal types such as the models of “presidentialism” and “parliamentarism,” formerly used in comparative public law, are in decline.\(^\text{153}\)

\(^\text{150}.\) V.H. HEYWOOD, PLANT TAXONOMY 14 (2d ed. 1976).

\(^\text{151}.\) See ENGISCH, supra note 137, at 240–41.

\(^\text{152}.\) This type describes the perfect condition of an institution or rule that is only partially approximated in reality. See NEIL J. SMELSER, COMPARATIVE METHODS IN THE SOCIAL SCIENCES 116 (1976). Critically on Weber’s approach, see generally Ahmed A. White, Max Weber and the Uncertainties of Categorical Comparative Law, in RETHINKING THE MASTERS OF COMPARATIVE LAW, supra note 17, at 40, 53.


were subsumable under these headings to allow a meaningful comparison. The following diagram illustrates the composition of comparative concepts:

**Diagram 1: Properties Used for Concept Formation**

![Diagram](attachment:diagram.png)

As shown in Diagram 1, defining properties and accompanying properties should be further differentiated into two sub-categories each. Defining properties can be distinguished into characteristics and indicatives. Characteristics are “compulsory” defining properties, i.e., such ones that in their sum are necessary and sufficient to identify an institution as a member of the concept. A comparative concept, being a type, should always consist of a multitude of characteristics. This decreases the risk of choosing a distinguishing property judgmentally, adds depth to the concept, and enhances its explanatory utility. To find out which properties are characteristics, ideally all legal systems have to be searched for legal institutions that share characteristics with the one of interest. Mini-

---

154. The terms “defining properties” and “accompanying properties” are taken from Giovanni Sartori, *Guidelines for Concept Analysis*, in *Social Science Concepts: A Systematic Analysis* 15, 33 (Giovanni Sartori ed., 1984). These subdivisions are original.

nally, all legal systems that are part of the comparative research project have to be surveyed.

Indicatives, or "sometimes differentiating properties," are optional defining properties. They are individually and in their sum not capable of defining an institution as a member of a specific concept, but they indicate membership because many members of a concept share these properties. To include indicatives in the definition of a concept makes sense because they illustrate its dynamics. In the case of emerging concepts, indicatives refer to properties that might become characteristics once the concept has materialized fully. In the case of aging concepts, they indicate properties that once were characteristics. The latter is important to tackle the problem of conceptual stretching. This problem refers to the distortion that occurs when an aging concept does not fit new cases. The usual result in such a case is that an otherwise useful concept is malformed or abandoned. This rather unwelcome outcome can be avoided when means are found not to depend on the assumption that members of a concept share a full set of defining properties. Indicatives are such a means.

Accompanying characteristics are those that do not serve to distinguish concepts from one another. They fall into the categories of dissimilars and idiosyncratics. Idiosyncratics are properties that are unique to a particular institution, but not characteristic enough to distinguish from other members of the same concept. The consideration doctrine of the common law, for example, would be such an idiosyncratic property within the concept of "voluntary agreements." Dissimilar properties of a legal institution have a complement in other members of the same concept. This complement, however, is of a different kind, but again, not characteristically different. Compulsory licenses, for example, are administered in some countries by public authorities, in others by courts. This dissimilar difference is not substantial enough to subsume the two systems under different concepts.

i. Defining Properties

To determine defining properties properly and to distinguish them from accompanying ones is probably the most important operation of the

157. Heyde, supra note 155, at 239.
158. Collier & Mahon, supra note 156, at 852.
159. Reitz, supra note 3, at 621.
method. For practical reasons, only a part of the properties of an institution can be used for the purpose of its classification. A relevant choice has to be made because the classification will vary depending upon which element is used as the focal point.

Gerring has recommended two key criteria for good concept formation: internal coherence and external differentiation.\textsuperscript{160} Internal coherence means that the defining properties should not merely coincide in space and time, but have an inner relationship to one another. External differentiation means that the defining properties are chosen so that they clearly define the borders of the concept by delimiting it against other concepts. The defining properties should not only say what the concept is, but also what it is not. Consequentially, concepts should not overlap each other. Though there might be hybrid real-world legal institutions insofar as they might fall in between two legal concepts, there are no hybrid concepts. That would reduce the utility of each of the concepts and might increase the danger of the appearance of memberless concepts.

Otherwise, the six criteria set out for the formation of comparative concepts guide the selection of defining properties. To preserve the validity of a concept, its defining properties must not be of wide variation, must not be easily modified by extra legal factors, and must not change readily.\textsuperscript{161} Furthermore, to avoid bias, characteristics should be devised as neutral as possible. Although, this goal will not always be achievable. “Genocide” is hard to define without recourse to pejorative attributes and “human rights” without valorizing ones.\textsuperscript{162} However, characteristics can at least be named neutrally: the comparatist must not rely on the idiosyncrasies in taxonomy and terminology of any particular jurisdiction because these were created as explicatory mechanisms for particular legal phenomena. Especially, homonyms in different legal systems present traps.\textsuperscript{163} The conceptual comparatist needs to distinguish carefully between the legal terms he or she finds (which are irrelevant for the comparative process) and the legal institutions they represent (which are the “bricks” that he or she wants for the construction of his or her concept).

Characteristics have to represent both the structure and the consequences of the legal institutions that are members of a concept. If either structure or consequences differ substantially, then the institutions in question belong to different concepts. “Morals crimes” (prostitution, dis-

\textsuperscript{160} Gerring, supra note 155, at 373–76.
\textsuperscript{161} Heywood, supra note 150, at 33.
\textsuperscript{162} Gerring, supra note 155, at 385.
\textsuperscript{163} Sartori, supra note 154, at 35–39; Watson, supra note 40, at 11; Bartels, supra note 146, at 92; Otto Kahn-Freund, Comparative Law as an Academic Subject, 82 L.Q. REV. 50, 52–53 (1966); see, e.g., Ainsworth, supra note 94, at 20–21.
tribution of child pornographic material, etc.) may, for example, fulfill the structural criteria of statutory sanctions in three different legal systems. In system A, though, the behavior is treated as a misdemeanor, in system B as a felony, and in system C as a mere infringement of administrative regulations. In this case, the legal institutions of systems A and B can be attributed to the same comparative concept because the consequences—though not identical—are sufficiently similar (both systems grade the factual circumstances as criminal offenses). The legal institution of system C forms part of a separate concept (administrative offense).

ii. Gradation

The difference in consequences between legal systems A and B lead to another feature of Conceptual Comparisons: gradation. Concepts have to be defined and named in such a broad way that they cover even heterogeneous legal institutions, as long as they share overriding defining properties. Accordingly, the characteristics employed in concept formation have to be considerably wide. As in the example of “morals crimes,” some members of the concept will have more of a characteristic than others. This varying grade is highlighted in Diagram 1 as superscript numerals. The higher this number, the more typical is the particular institution. We measure gradation for two reasons. First, it helps concepts to weather the storm of scholarly debate. As empirical research of psychologists suggests, concepts that are construed as rigid, logically bound domains defined exclusively by all-or-none criteria are unlikely to maintain their hold on the discourse. This is because scholars naturally disagree upon how well an institution represents a concept or one of its characteristics. Secondly, gradation has an inherent analytical value. It allows for singling out institutions that are more central to the concept than others (prototypes) because they express its characteristics to an exceptionally pure degree.

b. Quantitative Analysis

In its quantitative analysis, the conceptual orientation looks at the extension of a particular concept, the entities in the world to which it refers.

165. JOHN CYRIL SMITH & BRIAN HOGAN, CRIMINAL LAW 26 (10th ed. 2002).
166. See Eleanor H. Rosch, Natural Categories, 4 COGNITIVE PSYCHOL. 328, 328–50 (1973).
These entities are the various factual situations that are covered by the legal institutions that form part of the concept (case lines). The underlying question is, which other factual situation than the one we started with do the institutions that are members of the concept cover? These factual situations can be considered the quantitative characteristics of the comparative concept. Again, similar to the qualitative analysis, ideally all legal systems are browsed for case lines, which members of a certain concept address. All factual situations discovered in this way form the quantitative dimension of the concept, though only seldom will a real-world institution cover the full set of situations. In its quantitative dimension, therefore, the comparative concept is a Weberian ideal type, or “moulded type” (“Ausprägungstypus”) in Engisch’s words, an ideal model for a group of existing legal institutions in which the determining qualities of these institutions are represented exceptionally purely. The reason why the quantitative analysis is looking for an ideal type, as opposed to the real type that the qualitative analysis established, is the former’s significance for the capability analysis at which the conceptual orientation aims. Only an ideal type allows for assessing the relative potential of a particular institution later in the systematic comparison.

In construing the quantitative dimension of the concept, one should remember the analogy drawn between comparative law and a dictionary. Just as a dictionary is full of obsolete and archaic, unused and common words, the quantitative dimension of the taxonomic concept is timeless. It includes all case lines that have been addressed by the concept, past and present. Should more than one institution address one factual situation in a particular legal system, like, for example, an accident that can give rise to delictual as well as contractual liability, it is quantitatively relevant for all of these institutions.

c. Example: Conceptualization of Second-tier Protection for Inventions

In the following section, I will give an example of how a conceptualization under Conceptual Comparisons might look, and how it might make comparative studies more meaningful. The example is taken from the field of intellectual property law. Second-tier protection for inventions (second-tier protection) has become a hotbed for comparatists since 1995 when the European Commission announced its intention to harmonize national laws of the EC Member States in this regard. Second-tier

167. See ENGISCH, supra note 137, at 246.
protection is conventionally defined as all forms of intellectual property that provide protection for minor technical inventions, which do not comply with the requirements of patentability. Such a kind of intellectual property right is believed to be particularly suited for small and medium-sized enterprises that make modest workshop improvements to existing products and lack the resources and the necessary market information to engage in full-scale patenting.

i. Pre-existing Conceptualization

Applying Conceptual Comparisons to second-tier protection is an exercise in re-conceptualization. Comparatists have divided European jurisdictions already into three or four different groups of regimes that they see as instances of second-tier protection. The criteria used for differentiation, so far, are the extent of subject matter protected and the test of necessary advance over the prior art. According to this conceptualization, a first group, including Italy, Portugal, and Spain, has retained the classic “utility model” regime of the late nineteenth century. Second-tier protection, here, mainly fills a particular gap in the law of technical design by protecting functional shapes (“spatial forms”) of hand tools and similar implements, which are neither covered effectively by patent law nor by trade secret law. Jurisdictions of this group distinguish second-tier protection from patent protection by lowering the threshold of admis-


172. Newton, supra note 170, at 446.


sibility compared with the inventive step requirement of patent law on the one hand, and by confining the protectable subject matter to “products in a spatial form” on the other hand. A second group, including Germany, Denmark, and Austria, has moved away from the classic utility model regimes by dropping the spatial form requirement while retaining the “soft” inventive step standard. A third group of countries, including France, Belgium, and the Netherlands, likewise refrains from the spatial form requirement, but unlike the second group, uses the full inventiveness test of patent law.

In addition to these three groups, some scholars recognize a fourth group of European jurisdictions, notably the United Kingdom and Luxembourg, where protection for minor inventions allegedly is provided by “functional equivalents” to second-tier protection, such as trade secret law, unregistered design rights, or the lowering of the general inventiveness test in patent law.

ii. Mis-conceptualization

This conceptualization of second-tier protection—especially the inclusion of the fourth group—is devised by scholars working in the functional tradition. True to his or her method, the conceptual comparatist has to reassess the pre-existing classification critically. Upon his or her presumption, that the structure of a particular legal institution is important in ordering legal understanding, it will strike him or her as questionable, whether both the third and the fourth group of countries can be subsumed under the heading “second-tier protection.” The French “certificat d’utilité” of the third group, for example, runs parallel with the patent law system with the only difference being that inventions merely get a short-term protection of six years under a certificat d’utilité, which is granted without a prior search report. Inventions of a lower level of inventiveness do not receive any particular attention. Accordingly, “second-tier protection” in France is, in practice, used to a much lesser extent


The conceptual comparatist will also reject subsuming “functional equivalents” that supposedly exist in legal systems under the concept of second-tier protection. Some of these “equivalents,” such as trade secret law, are not even equivalent because second-tier protection was explicitly devised in the nineteenth century to compensate their deficits (in the case of trade secret law deficits in protecting minor inventions).\footnote{182 Reichman, supra note 174, at 2458–59.} Others, like the unregistered design right, afford short-term protection to original shapes and configurations. They are nonetheless dissimilar to second-tier protection in the form of a utility model right because the kind of protection they grant is different.\footnote{183 SUTHERSANEN, supra note 179, at 19–046; LLEWELYN, supra note 170, at 80–82; Andrew Parkes, Short-Term Patents in Ireland, 25 INT’L REV. INDUS. PROP. & COPYRIGHT L. 204, 208 (1994).} Most importantly, the unregistered design right does provide its holder with a lower level of exclusive protection\footnote{184 On the various levels of exclusivity in intellectual property law, see Oliver Brand, Die Ketten des Prometheus—Grenzen der Ausschließlichkeit im Immateriagüterrecht, in JAHRBUCH JUNGER ZIVILRECHTSGESELLSCHAFTLER 2005 77, 78–83 (Axel Halfmeier et al. eds., 2006) (F.R.G).} than second-tier protection of inventions. A violation of the registered design will only be recognized by courts when its holder can prove “reproduction” by copying, whereas a second-tier protection regime will always grant exclusive protection regardless of a proof of copying. Furthermore, the law of unregistered designs works with a different standard for protection (“creativity” instead of “inventiveness”) and does not warrant priority for the inventor because no registration takes place. Like the protection of minor inventions, within the framework of patent law, by lowering the latter’s requirements for inventiveness, protection by unregistered design seems to be an alternative form of protection, rather than an equivalent to second-tier protection.\footnote{185 LLEWELYN, supra note 170, at 33.}
iii. Re-conceptualization

Accordingly, the conceptual comparatist has to try to re-conceptualize second-tier protection. The conceptual comparatist will begin by breaking down legal institutions from jurisdictions, which fall into one of the uncontroversial first two groups, into their components to abstract the defining properties that he or she needs to establish the qualitative dimension of the concept “second-tier protection.” Most likely, the conceptual comparatist will come up with three characteristics. Second-tier protection covers any institution:

1. whose requirements for acquiring exclusive protection are less stringent than those for patents, because the tests of “inventive step” and/or “non-obviousness” are lower or absent altogether;
2. whose term of protection is shorter than that for patents (usually between seven and ten years without the possibility of extension or renewal); and
3. that is not subject to substantive examination prior to the grant.

In addition, the conceptual comparatist will find one indicative, showing the rise of a new characteristic of second-tier protection. An increasing number of legal systems that award some form of second-tier protection follow the German lead in affording for the applicant a grace period. The conceptual comparatist’s (or his or her predecessor in title’s) written publications or public use within six months before the priority date of the invention do not constitute prior art. On the other hand, unique features, such as a limitation of the number of claims available under second-tier protection in Australian innovation patent law, will be disregarded as accompanying properties.

Gradation allows us to identify German, Spanish, Danish, and Irish institutions as members of the concept “second-tier protection,” though the threshold of novelty is surmounted at a different level. While Danish and Irish laws apply novelty for second-tier protection at a universal level, like in patent cases, German examiners take into account a more restricted state of the art (written publications worldwide, but only if they are in public use within Germany before the priority date of the invention). In Spain, the standard is even more restricted. Here, novelty for

186. A full list of the properties that need to be assessed can be found in Tunnell, supra note 180, at 15.
187. Janis, supra note 175, at 172.
second-tier protection is determined according to the national state of the art only.\footnote{188}

The defining properties, as abstracted above, would enable the conceptual comparatist to approach second-tier protection with a refined definition of the concept: it refers to all registration-based forms of intellectual property that provide protection for technical inventions with a lower level of inventiveness without prior examination and under less onerous conditions than patent law. Its qualitative dimension would accordingly only encompass the first two groups of the initial conceptualization and might, in the context of neighboring concepts, look like this:

**Diagram 2: Second-Tier Protection Re-conceptualized**

<table>
<thead>
<tr>
<th>Design Law</th>
<th>Second-Tier Protection of Inventions</th>
<th>Patent Law</th>
</tr>
</thead>
<tbody>
<tr>
<td>United Kingdom</td>
<td>Spain</td>
<td>Germany</td>
</tr>
<tr>
<td>Italy</td>
<td>Austria</td>
<td>Belgium</td>
</tr>
</tbody>
</table>

Portugal

The quantitative dimension of second-tier protection—as abstracted from the legal institutions that contain at least all of its characteristics—is provided in Diagram 3. The “case lines” addressed by this concept are the different instances of protectable subject matter that the second-tier regimes cover. Diagram 4 shows the quantitative “share” of the concept “second-tier protection” that a real-world institution—the German “Gebräuchsmuster”—has. The darker color indicates case lines that German law does not cover.

Diagram 3: Quantitative Dimension of the Concept

The Concept of "Second Tier Protection for Inventions"

Diagram 4: Quantitative Dimension of a Legal Institution

Second Tier Protection for Inventions in Germany - "Gerrauchsmuster"

2. Second Phase: Systematic Comparison

We have now seen how a conceptualization in phase one of Conceptual Comparisons might look. In the second phase, the systematic comparison, the concept is utilized by comparing it to real-world institutions. The aim is to determine and explain the extent to which these institutions conform to the concept or deviate from it and how they deviate from one another. The systematic comparison has three stages. As with the two phases of the comparative study themselves, these three stages of the second phase are not always distinctly separated from each other, nor are they always dealt with in the same order: they may be intermingled in the same discussion. The three stages are all analytical in nature. Comparatists will disagree, as discussed in Part III, upon whether it is permissible to exploit the knowledge derived from these stages normatively. Therefore, this optional fourth stage of the systematic comparison is left out of the picture here.

a. Descriptive Stage

The first stage describes the legal phenomena under scrutiny. A proper description has to be objective, i.e., free from critical evaluation, as well as comprehensive. The latter requires the comparatist to examine the comparanda from at least four different perspectives. It is crucial to begin with an “internal description” of the legal institution that uncovers its precise construction and consequences because this shows whether the material used for construing the defining properties of the concept has been assessed properly. The internal description has to take into account all sources of law that the legal system under comparison regards as such, including the written and unwritten authoritative material as well as affirming and derogative usage. 190 These sources have to be presented as reflected in the intentions and procedure of the legislature, the jurisprudence of the courts, and the position of the doctrine. Additionally, it might be helpful to analyze whether the latter is divided and whether it is concordant or discordant with judicial positions.

The second perspective is the conceptual one. In describing the legal institutions, the comparatist must make clear how they relate to the concept established in phase one of the comparative process. Qualitatively, the comparatist has to demonstrate that the institution shows the defining properties of the concept. Quantitatively, it has to become clear which factual situations of the ideal set contained in the concept that the real-world institutions address.

190. See Zweigert & Kötz, supra note 12, at 35–36.
The third perspective—the systematic one—will place the topic under study in the context of the entire legal system, for the same institution may be central for one legal system and of only marginal relevance for another. The systematic description will ask, furthermore, how an institution conforms to general principles, and whether it depends on rules or institutions in other areas of the law such as procedural rules, constitutional provisions, or the requirements of supra-national law (e.g., EC directives).

The fourth perspective, finally, is meta-legal. It requires a description of the socio-economic factors of the systems in question; their policy considerations; their philosophical premises, such as the role of law; and their social values. This perspective is extremely important for comparisons, which, like Conceptual Comparisons, are based on typology. If the socio-economic context of the comparanda is left unexamined, then variations in the level of power that they confer (effective law versus symbolic law) remain hidden, which might impair the comparative process. The meta-legal perspective also seeks to identify the actors that have played a role in shaping the objects of comparison because law must be understood as a consequence of political decisions and power structures.

b. Identification Stage

The second stage identifies the differences and similarities between the systems under comparison (identification stage). At this stage, the comparatist has to establish the extent to which the respective institutions deviate from the concept and from one another. In determining the deviation from the concept, the question of gradation has to be addressed. Furthermore, the accompanying properties of particular institutions have to be brought out and distinguished from the defining ones that place them within the concept.


c. Explanatory Stage

The third stage accounts for divergences and resemblances—especially with regard to the concept—and can therefore be called the “explanatory stage.” Historical analysis, functional analysis, and actor analysis exercise especially strong explanatory power. Historical analysis puts institutions to a “diachronic test” by revealing whether they are genuine or borrowed from another legal system: legal institutions may be similar for three different reasons. They may have common ancestry, i.e., derive from the same (now discontinuous) legal institution in the past (e.g., Roman law), they may have developed in parallel, or they may have converged. Parallel development and convergence can be distinguished by the fact that in the former case, legal institutions developed similar features independently, while in the latter case, they did so through direct contact or through mediation of a third legal institution. The variation of legal institutions, on the other hand, can be explained with regard to three other factors. The extra-legal environment might have modified them in a different way or they might have diverged either through the influence of a third legal system or due to innovative doctrinal reconstruction in the domestic law.

Alongside historical inquiry, functional analysis can be used at the explanatory stage. Here, this sort of analysis is valid because it is only used as a means of differentiating legal institutions from one another with regard to an abstract concept that has no particular function because of the way its characteristics are defined. Neither similarity nor difference in the function of the institutions under comparison is presupposed. Their function shall be determined exclusively in their respective domestic contexts. Deviations from the concept can be caused by the fact that the institutions might address different cases or that the cases addressed are problematic to a different extent in diverse legal systems. As part of a functional explanation, “hidden rules,” such as the political norms of the qian guize in China, can be detected and economic considerations can be put forward as well. Economic issues, however, again have to be analyzed strictly contextually. Otherwise, the problem of multiple optima might rear its troublesome head again, as we can see beautifully in our example of second-tier protection. Here, economic analysis in Germany, Australia, and the United States have recommended rather

194. See von Sengen, supra note 192, at 56–58.
different arrangements for second-tier protection, arguably because the respective local needs, perceptions, and institutional settings are different, too.

Actor analysis, finally, takes into account that legal processes are complex configurations of institutional and non-institutional legal actors. They give legal institutions their characteristic shape by applying and interpreting the standards of legal norms and doctrine. In this respect, the occurrence of “pre-eminent lawgivers” (überragende Nomostheten) demands particular attention. A pre-eminent lawgiver is a person who coins institutional actions (i.e., court decisions or enactments) to such an extent that these actions appear his or her own, rather than actions of the institution in which he or she is embedded. At the same time, the influence of actors cannot be measured in terms of their impact on positive law only. Otherwise, the influence of legal scholars like Rheinstein, Ehrenzweig, or Kelsen, who “never directly influenced enactments or court decisions,” would be overlooked. As the influence of a particular actor is determined by his role within a given legal culture, the central questions to be addressed by the comparatist regarding actor analysis are: Who is responsible for the development of the law? Which interest groups support change or adherence to the law? Are there changes in the institutional structure of an actor (affirmative action, regime change, etc.)? What are the motives and perspectives of these groups? Which are their formal and informal roles in a given society and how do they interact with other actors? In our example of second-tier protection, answering questions such as these can highlight, for example, why some legal systems have not adopted such a form of intellectual property: in the Netherlands, local industry lobby groups prevented an enactment of second-tier protection because they feared its enactment might favor foreign competitors.

197. Janis, supra, note 175, at 189–90.
198. Bell, supra note 40, at 12–13, 19; Chodosh, supra note 2, at 1112.
199. On pre-eminent lawgivers, see Oliver Brand, Language as a Barrier to Comparative Law (unpublished article) (on file with author).
d. Contextuality

In all three stages, the legal institutions under scrutiny must be viewed in the socio-economic and cultural context in which they thrive. This context is vital for a proper understanding and accurate delineation of the law at the descriptive stage, for the precise identification of differences and similarities at the identification stage, and for a valid evaluation at the explanatory stage as well. As established above in the critique of functionalism, a mere study of texts and formal rules will give an incomplete and distorted picture. Some socio-economic and cultural background is needed in every comparative study. The amount of contextual discussion necessary depends on the socio-cultural proximity of the legal systems chosen for comparison. The greater the proximity between the comparanda, the less detail of the general social context needs to be examined. In intra-cultural comparisons—e.g., comparing German and Austrian law—the “socio-cultural context” may be restricted to the immediately relevant aspects of the social and economic environment. In comparisons of legal systems that belong to distinct societies (cross-cultural comparisons)—e.g., English law and Maori law—a detailed discussion of the social structure and organization is essential to properly assess factors such as social differentiation. To highlight the socio-economic background, methods of sociological research might have to be employed, such as statistical evidence, questionnaires, and interviews. Documents from legal practice such as standard business terms, register forms, etc., furthermore help to fill the abstract rules of foreign law with life.

C. Applications

Conceptual Comparisons aims to permit comparisons among the entire range of legal systems, even between systems from dissimilar socio-economic environments and between “radically different cultures.” Naturally, however, the same set of legal systems is not equally relevant for century Britain, see Brad Sherman & Lionel Bently, The United Kingdom’s Forgotten Utility Model: The Utility Designs Act 1843, 3 INTELL. PROP. Q. 265, 273–75 (1997).


205. See Drobnig, supra note 30, at 499–03.
every comparative study. The question of which legal systems the comparatist should choose to compare must be answered on a case-by-case basis. The concentration on “parent systems”\textsuperscript{206} is not even good as a working rule. Rather, the choice of the comparanda depends on the kind of study that the comparatist wants to undertake. In the following sections, the range of studies that Conceptual Comparisons allows is explored. In the diagrams, included for reasons of clearness, squares symbolize concepts, ellipses stand for individual legal institutions, and arrows indicate the comparative studies that may be undertaken.

1. Deviant Case Studies

![Diagram for Deviant Case Studies]

The first interesting application of Conceptual Comparisons is the “deviant case study.” After having established a concept, this kind of study takes one of the ordinary members of the concept and compares it to another member of the same concept that possesses either a remarkable quantitative dimension—be it exceptionally large or small—or a notable number of accompanying properties (deviant case). Deviant case studies make use of the critical potential of comparative law. They reveal the tolerance of a concept towards the moral and political values imposed by legal systems upon the form and substance of legal institutions. In the example of second-tier protection, Austrian law would qualify as a deviant case within the concept because, quantitatively, it is the only jurisdiction that uses this kind of intellectual property right to provide protection for software.\textsuperscript{207}

2. Contrastive Comparisons

![Diagram for Contrastive Comparisons]

Like deviant case studies, contrastive comparisons aim to highlight the diversity of law. There are two versions of this application. Intra-conceptual contrastive comparisons concentrate on “polar types,” members of the same concept that differ maximally either in illustrative and accompanying properties or in the underlying socio-economic implications. These comparisons

\textsuperscript{206} ZWEIGERT & KÖTZ, supra note 12, at 41.

use concepts as “benchmarks” against which to establish the unique features of the real-world systems under scrutiny. Their heuristic value is to map the width of the concepts. In our example of second-tier protection, the comparison of a (narrow) classical utility model system, such as the Spanish one, and a much wider second-tier system, such as the Irish “petty patent” or the Austrian “Gebrauchsmuster” that protect any patentable subject-matter, would be an instance worthy of a contrastive study. Intra-conceptual contrastive comparisons also allow for comparisons in highly value-laden areas of law. One might compare, for example, the approach of the “antipodes of the Muslim world,” Indonesia,\textsuperscript{208} and Morocco,\textsuperscript{209} towards the discrimination of women under the common roof of the shari’a principles.

Inter-conceptual comparisons juxtapose a member of one concept with a member of another concept, while both institutions address a similar set of factual situations. To compare members of different concepts is valid because the applicability of concepts as a tertium comparationis does not depend on the actual presence or absence of the relevant characteristics in the legal institutions compared, but rather on the capability of these institutions to exhibit that characteristic.\textsuperscript{210} That is the reason why it is possible to apply foreign concepts to one’s own legal system, as it has been done, for example, in a study applying American theories of evidence to Dutch criminal justice.\textsuperscript{211} A contrastive comparison with an institution outside the concept of second-tier protection would probably ask how the United States or the United Kingdom addresses the case lines covered by second-tier protection, highlighting problems with substituting second-tier protection with trade secret law and a reduced inventiveness standard in patent law.

\textsuperscript{208} See, e.g., SHARI’A AND POLITICS IN MODERN INDONESIA 123 (Arskal Salim & Azyumardi Azra eds., 2003).


\textsuperscript{210} E. Gene DeFelice, Comparison Misconceived, 13 COMP. POL. 119, 123 (1980).

3. Developmental and Hybrid Studies

Conceptual Comparisons can also analyze developments in the relationships between legal systems. Such an analysis has to be exercised in two steps. In a first step, the comparatist has to explore tendencies of convergence or divergence between legal systems (in terms of “real” differences). In a second step, a comparative developmental study would explain and evaluate such tendencies: Why do systems converge or diverge? Is convergence desirable or undesirable? In an increasingly integrated world, convergence may, for example, be required under international (e.g., World Trade Organization [WTO]) or supra-national (e.g., European Union [EU]) law. Consequently, “hybrid” legal institutions at the “intersections” of comparative concepts may emerge. The existence of hybrids can also be explained by the fact that characteristics cannot always be formulated in a “binary code:” real-world legal institutions do not fit into characteristics in a “yes-or-no” way, or a “more-or-less” way, if you take gradation into account. They might fit in only “partially,”212 as the French “certificat d’utilité” does, for example, in the system of second-tier protection as well as in the system of patent protection.

4. Taxonomic Comparisons

A fourth possible application for Conceptual Comparisons is taxonomic comparison. Concepts can be differentiated into equivalent, horizontal categories and hierarchical, vertical ones. The latter can be organized in the form of taxonomies similar to the botanical ordering of genus, species, and sub-species. A taxonomy of concepts—while certainly not an end in itself—might enhance the heuristic value of comparative law as an organizational discipline.213 It simplifies and systematizes data collection and makes intra- and inter-group comparisons easier by highlighting the interrelation of concepts.

Concepts, which share the most characters in common, can be placed into larger, more inclusive classes (genera); these in turn are assembled into even more inclusive groups called families as outlined in Diagram 5. The concept, “agreement of sale,” for example, belongs to the genus “voluntary agreements” that forms part of the family “obligations.” Each

212. See generally Tate, supra note 148, at 19–20 (listing examples).
vertical class can be understood as a different level of abstraction. We climb the “ladder of abstraction by reducing (in number) the characteristics of a concept [and] descend a ladder . . . by augmenting (in number) [those] characteristics.” It is advisable, however, to climb the ladder of abstraction with care. A concept described by only very few characteristics can become analytically insignificant. When the conceptual comparatist climbs the ladder of abstraction, he or she has to be careful. The quantity of objects that a concept refers to says nothing about its analytical utility. Therefore, even concepts that have only one member can be built (monotypic concepts). That might be useful and necessary for including very remote legal systems in a comparative study.

Diagram 5: Taxonomic Organization of Concepts

At this point, it should be reiterated that the concepts, which Conceptual Comparisons forms, are not necessarily alternative, i.e., mutually exclusive. They can also be cumulative, i.e., concepts in which institutions can be classed that also belong to another category. This is unproblematic as long as comparatists are aware that the taxonomies they create with Conceptual Comparisons constitute a number of interrelated taxonomic “trees” that comparatively describe the law, and not a single, unitary one.

214. Sartori, supra note 154, at 44.
215. Barthels, supra note 146, at 76.
216. Cf. Smelser, supra note 152, at 176.
217. The diagram is an adaptation of Heywood, supra note 150, at 14.
218. See supra Part IV.B.1.b.
5. Prototype Studies

Concepts will be frequently derived from a specific system of geopolitical and economic significance to the comparatist’s audience. That might lead to an implicit equivalence of a parochial “prototype” with the concept. It might be interesting to compare the prototype with other members of the concept. This could elucidate the way in which the concept and its members develop. Prototype studies could reveal, for example, that concepts are becoming more condensed because of internal tendencies for further convergence. As it might happen in the case of second-tier protection, the intended harmonization of the law of EC Member States is closely based on the “prototypical” German approach. On the other hand, it might be shown that concepts are becoming weaker because benchmark jurisdictions that have provided models for shaping a particular area of the law lose their appeal, as it now happens with German company law.

6. Diachronous Comparisons

A sixth application for Conceptual Comparisons is the diachronous study. These studies involve either the comparison of subsequent legal institutions in one legal system that is a member of a specific concept, or the comparison of different members of a concept that existed at different periods of time. Such comparisons establish the historical connections within a concept and allow the study of legal development. In the case of second-tier protection, it might be interesting, for example, to compare the German approach prior to 1990, when it still had a spatial form requirement, with the present Spanish or Italian law, which currently maintains this requirement.

219. Chodosh, supra note 2, at 1107.
220. Sutheesanen, supra note 179, at 19–048.
7. Case Studies

Case studies, the final application of Conceptual Comparisons, are concentrated on one particular factual situation. In a particular case, we might ask for example: How do different legal systems deal with minor improvements in the manufacture of hand-tools? From that starting point, the comparatist has to find out which legal institutions are invoked by the legal systems that he or she wants to study to deal with the situation. These institutions, in turn, have to be grouped under the various concepts to which they belong. Only then is a direct comparison between the different legal institutions possible.

D. Virtues of the Method

The applications of Conceptual Comparisons indicate some advantages of the method. First, it allows for studies not viable under functionalism or any of the other methods discussed in Part III. Prototype studies, taxonomic comparisons, and diachronous comparisons, are unique to Conceptual Comparisons. This method also responds to a rising call for expanding the scope of comparative law to purely domestic contexts (one-country studies).222 In areas where legal transplants are frequent, as in company law, transplanted and domestic institutions often co-exist and compete with each other in the same legal system. Conceptual Comparisons might provide a proper yardstick to analyze such situations because it does not require comparanda from different legal systems to be operational.

Secondly, over time, Conceptual Comparisons will establish a common reference system in the form of the concepts that it develops. Thereby, it might serve as a “research cycle” that allows a more fruitful dialogue of scholars working in different traditions. Critical legal scholars and culturalists can continue their work within the framework of “deviant case studies” and “contrastive comparisons.” They can, furthermore, monitor whether the explanatory stage of phase II pays due respect to the extra-legal context of the rules examined. Functionalists—in the broader and the narrower sense—can also carry on contributing to comparative studies. Conceptual Comparisons probably makes their insights even more

222. Chodosh, supra note 2, at 1084.
valuable. It frees them from objections against the premises upon which they work by acknowledging that the function of law is context-dependent and by consequently limiting the role of functionalism to that of an explanatory tool in phase II of the comparative process.

Thirdly, Conceptual Comparisons might make the application of knowledge derived from comparative studies more likely. The works of comparative lawyers, especially those from civil law jurisdictions, are often disregarded by their more dogmatically oriented colleagues because of their functional or otherwise un-dogmatic approach. Conceptual Comparisons has the capacity to alter this by providing a framework that pays more attention to the constructional details of the law and might therefore dissolve or mitigate the antagonism between comparatists and dogmatic scholars.

Fourthly, Conceptual Comparisons makes the choice of legal systems for comparison more transparent and rational. The process advocated so far by functionalism is dependent on implicit choices necessitated by the functionalists’ general avoidance to discuss questions of “likeness” and on the operational limitations that the three presuppositions of functionalism force upon the comparatist. In contrast, the conceptualization undertaken in phase I of Conceptual Comparisons provides a criterion that requires the comparatist to make an explicit and rational choice. As the sample applications sketched above have shown, the comparatist will have to limit research to those legal systems that fit the kind of study he or she wants to undertake. That can still be “radically different cultures” in the case of a contrastive comparison or members of the same concept in a prototype or deviant case study.

Finally, Conceptual Comparisons can assist the study of legal transplants in numerous ways. Conceptualization has predictive value. It can predict legal transplants and legal change by answering questions such as: Why and how do legal systems change? Which factors are more likely to resist legal change by imitation? How does the structure of a recipient legal system affect and modify a received legal institution? The conceptualization undertaken in phase I of the method also helps to understand the transplanting process itself better. It provides a degree of measurement against which it can be established, whether legal change takes place in the form of a gradual process of “cross-fertilisation” or whether entire tracts of law move from one system to another. Thus,

224. This term is used by John Bell, Mechanisms for Cross-fertilisation of Administrative Law in Europe, in New Directions in European Public Law 147 (Jack Beatson & Takis Tridimas eds., 1998).
regulative competition between legal systems becomes more transparent. Conceptual Comparisons furthermore allows us to distinguish between two kinds of transplants: intra- and inter-conceptual transplants. The former are more likely to happen and more likely to be successful because the receiving and the donating legal system can rely on a sufficiently similar structure in addressing a certain factual situation.

E. Possible Objections to the Method

In conclusion of this study, there are four likely objections to Conceptual Comparisons that I would like to pre-empt.

First, it might be said that Conceptual Comparisons has little value for practical comparative law. Practitioners, especially judges, will not base their decisions on artificial average types or ideal types like the concepts established by the proposed method. That is certainly true. It is maintained, however, that “practically applicable comparative law” is an illusion anyway. The restraints on time and resources—especially of courts—are too pressing to allow meaningful comparative work on a broad scale. To make comparative law practicably relevant, it has to be “filtered” through the “lenses” of comparative scholarship that in turn can be consulted by practitioners. Conceptual Comparisons, in this regard, makes sure that the distorting effect of the lenses is mineralized.

Secondly, comparatists might fear that Conceptual Comparisons leads them back into the formalism of nineteenth century comparative law and Saleilles’ dream of a droit commun idéal. That fear is unfounded. Conceptual Comparisons acknowledges that social facts are so diverse that they do not lend themselves to a single scheme of concepts. Respectively, it does not claim that there is a single set of mutually exclusive concepts under which all legal institutions of all legal systems can be subsumed. As shown in the description of taxonomic comparisons, the proposed method rather sees concepts on a horizontal level as competing

225. See Sandrock, supra note 87, at 48.
ways of addressing a factual situation, allowing for a variety of alternative and cumulative concepts that can be vertically integrated in different ways. As long as the comparatist explains the construction of his concepts in a way that allows other comparatists to verify them, there is also no need to fear\footnote{Bartels, supra note 146, at 76.} that value judgments of the individual comparatists may become intermingled inseparably with seemingly neutral concepts.

Thirdly, one might suggest that the proposed method makes case studies, which form the bulk of the comparative work, exceedingly difficult. It is conceded that Conceptual Comparisons makes them more complex. That, however, is only appropriate. Case studies are inherently dangerous because they only ask how the law addresses one particular factual situation, without regard to other situations with which the institution may have to deal. Conceptual Comparisons forces the comparatist to take into account that not only the law, but also the factual situation it addresses, must be seen as part of a wider context.

Finally, scholars might object that functionalism may not be displaced as the core method of comparative law because of its analytical value for private international law. In fact, there are various approaches to conflict of laws that use “functional” or “teleological” methods in characterizing foreign legal institutions or rules.\footnote{For Continental European approaches, see Jan Kropholler, Internationales Privatrecht 122 (4th ed. 2001). For American approaches, see Eugene F. Scholes et al., Conflict of Laws 43–58 (3d ed. 2000) (discussing the functional approach developed to analyze conflict of laws).} However, these approaches have to be distinguished methodologically from functionalism as the basis of comparative law. That becomes evident already by the fact that functional analysis in the field of private international law routinely works in areas that comparative functionalism admits to be beyond its reach, e.g., family law.\footnote{Bernd von Hoffmann & Karsten Thorn, Internationales Privatrecht: einschließlich der Grundzüge des Internationalen Zivilverfahrensrechtes § 6, Nos. 5–8 (7th ed. 2002) (F.R.G).} The demise of functionalism as the basis of comparative studies and its replacement with Conceptual Comparisons would therefore not interfere with the mechanisms of private international law.
CORPORATE GOVERNANCE: THE LAW’S RESPONSE TO AGENCY COSTS IN NIGERIA

Dr. Ige Omotayo Bolodeoku*

I. INTRODUCTION

Corporate failures are hardly a new story. Indeed, states and economies have failed in the past, and many more are failing. However, the magnitude of corporate failures and the extensiveness of their impact have made them part of the several issues hotly debated in contemporary corporate and political governance scholarship. This also helps to explain why in the United States, a country with a federalized corporate law system, the federal government, following major corporate failures in 2001, introduced certain corporate governance safeguards that were once overlooked or considered unnecessary. What is interesting is that

* Senior Lecturer, Dept. of Commercial & Industrial Law, University of Lagos. LL.B. (Hons) (Lagos State University, Ojo); LL.M (University of Lagos, Akoka); D.Jur. (Osgoode Hall Law School York University, Toronto, Canada). I dedicate this Article to my wife, Karlene, whose support has been invaluable, and to my children Olumayowa and Olorunfemi for all the fun they provided me in the course of working on this Article. The author may be contacted at ibolodeoku@gmail.com.


2. See Jean Cartier-Bresson, The Causes and Consequences of Corruption: Economic Analyses and Lessons Learnt, in ORG. FOR ECON. CO-OPERATION AND DEV., NO LONGER BUSINESS AS USUAL: FIGHTING BRIBERY AND CORRUPTION 11, 12 (2000) (discussing agency problems in the political governance context, the author said: “Opportunities for corruption depend on the size of the rents in the control of public agents, the discretion they have in allocating them, and their lack of a sense of accountability to society.”).

failures, be they of corporations or states, are traceable, in part, to a common source, that is, activities of the respective agents representing the institutions.4

Every representative system must confront the problem of agency costs, whether as an internal matter or one that informs policy considerations by the appropriate regulatory bodies of governance structures.5 To be sure, agency costs are inevitable in a representative system, be it a business corporation,6 government, or any other form of association, whose activities are, in the nature of things, conducted by human beings.7 In other words, agency problems are associated with the delegation of power, management, and control, and thus highlight the transaction costs of specialization, which undoubtedly create value and enhance efficiency.8 On a more positive note, however, agency costs may be conceptualized as the transaction costs which a representative system must incur in ensuring efficiency in the system. In this regard, the system’s success may be measured both in terms of the magnitude of the costs

4. To be sure, the agency relationship is pivotal to the existence and operations of business corporations. See Eric W. Orts, Shirking and Sharking: A Legal Theory of the Firm, 16 YALE L. & POL’Y REV. 265, 274 (1998) (“Legal agency relationships are one foundation of business organizations. Without legal agency, business firms of any complexity are impossible.”). Note also that both law and reality supports the use of agency for the company’s activities because of the company’s peculiar nature as a legal entity. For example, sections 63 and 64 of the Companies and Allied Matters Act (CAMA) authorize a company to act through both institutional agents, such as the board of directors and the members in general meetings, and any agent or officer appointed by those two organs. Companies and Allied Matters Act (CAMA), (1990) Cap. C20, §§ 63, 64 (Nigeria). Even at the board level, the law permits a board of directors to delegate all its powers to an individual, called the managing director. See id. § 64(b).


6. It is important to note Orts’ caution regarding the nexus between agency costs and the firm. See Orts, supra note 4, at 278 (“Although agency costs help to explain the economic dynamics of firms, they are not sufficient to explain the existence of firms and their boundaries.”).

7. Orts argued that, “agency costs provide one important theoretical explanation for limitations on the size of firms.” See Orts, supra note 4, at 275. In fact, the primacy of agency relationships to societal development may be better understood from the perspective of division of labor. For works describing the value of agency to development, see Adam Smith, An Inquiry into the Nature and Causes of the Wealth of Nations, (Kathryn Sutherland ed., Oxford Univ. Press 1993) (1776) [hereinafter Smith, Wealth of Nations]; Émile Durkheim, The Division of Labor in Society, 11–30 (1997); Michael Jensen & William H. Meckling, Specific and General Knowledge, and Organizational Structure, in CONTRACT ECONOMICS 251 (Lars Werin & Hans Wijkander eds., 1992) [hereinafter Jensen & Meckling, Specific and General Knowledge].

8. See discussion supra note 7.
incurred relative to the benefits produced and by how well the costs are controlled or managed. On the whole, however, the inevitability of agency costs reinforces the need to design effective strategies for their management.9

In a corporate law context, the complexity of agency problems and costs is shaped, among other things, by both organizational size and ownership structure.10 A big organization with a large membership size (or diffused membership) will likely face graver agency problems requiring special contrivances for control than one that is small and has manageable and concentrated membership.11 While agency costs usually emanate directly from the conduct of those entrusted with the management responsibility of a given system, their persistence may depend on, and sometimes be exacerbated by, certain exogenous factors such as culture,12 perception, and the existence or lack of alternative control or intervention systems. Effective agency costs management will thus require that control systems are designed around the internal setting of the system; however the possibility that internal structures may weaken internal-oriented control mechanisms should serve as a reminder of the need to have alternative intervention systems.13

This Article will examine whether the response of the 1990 Companies and Allied Matters Act (CAMA or Act) and other allied legislation to


10. See James S. Ang, Rebel A. Cole & James Wuh Lin, Agency Costs and Ownership Structure, 55 J. Fin. 81, 87 (2000) (finding that agency costs (1) are higher when an outsider rather than an insider manages the firm, and (2) are inversely related to the manager’s ownership share).

11. One needs only to remember the analysis of Berle and Means about shareholders’ size and its impact on management behavior and agency costs. See Adolf A. Berle, Jr. & Gardiner C. Means, The Modern Corporation and Private Property (1933).


13. In the context of this Article, the provisions of CAMA on investigation by the Corporate Affairs Commission are treated as the alternative control mechanism that can substitute for the failures of the internal control mechanisms. See Companies and Allied Matters Act (CAMA), (1990) Cap. C20, §§ 314–15.
agency problems are effective enough to guarantee good and accountable corporate governance in Nigeria. To be sure, this Article recognizes that other institutions beyond the law matter in measuring the effectiveness of the law itself. However, examination of these other institutions is not within the scope of this Article. The Article gives a pass mark to the CAMA, while noting areas that may need to be re-examined for reform.

Analysis in this Article begins in Part II with an explanation of the concept of agency costs, pointing out how agency costs may present different challenges in relation to public companies in two separate situations—public companies with disperse ownership structure and those that combine both dispersed and concentrated ownership. The latter type is of particular importance to Nigeria, as most companies in Nigeria are in this category. Part II further defines agency costs and examines their relationship with the concept of separation of ownership and control. Part III focuses on the theoretical and normative analysis of how agency costs may be effectively dealt with in both types of public companies already noted in this Article. Part IV examines the CAMA provisions in order to assess the Act’s response to agency problems. This Part examines some of the provisions of CAMA in light of the normative or theoretical analysis guidelines discussed in Part III.

II. THE CONCEPT OF AGENCY COSTS EXPLAINED

A. Foreground

The organization of resources for productive purposes through the corporate form has become virtually complete, in the sense that it is dominant. A country will usually have a mix of private and public companies, with the former used mainly where organization of business through the partnership form would have been ideal; while the latter used not only where the requirement of capital is extensive, but also where it is in-

14. Note, for instance, that the judicial system is always pivotal not only to enforcement of rights in general, but also in ensuring the reduction of agency costs, in terms of prompt response to litigations in the enforcement of fiduciary duties. One may also note that the search for global capital, which forces domestic companies to either seek credit facilities abroad or list their securities in foreign stock exchanges, may require more disclosures on their part, thus enhancing the production of information and monitoring. See Rafael La Porta, Florencio Lopez de-Silanes & Andrei Shleifer, Legal Determinants of External Finance, 52 J. Fin. 1131 (1997); Rafael La Porta, Florencio Lopez de-Silanes, Andrei Shleifer & Robert Vishny, Investor Protection and Corporate Governance (1999), available at http://ssrn.com/abstract=183908 [hereinafter La Porta et al., Investor Protection] (“When investor rights such as voting rights of the shareholders and the reorganization and liquidation rights of creditors are extensive and well enforced by regulators or courts, investors are willing to finance firms.”).
tended to pool resources from a wide range of individuals sharing no other relationship than the mere coincidental desire to earn income from the resources they lack the capability or time to invest. Investment is not costless. It requires not only the capability and time to invest, but also the acumen to make resource allocation decisions which many people do not have.

While the membership of a typical public company is large and diffused, and the individual stake of its members (investors) is small relative to the company’s overall capitalization, it is not uncommon to have public companies, characterized by both concentrated and widespread share ownership. In their seminal book, Adolf Berle and Gardiner Means alerted the public to the consequences of the emerging modern corporation in which ownership of shares, due to its diffusion, is separated from control. “The corporate system appears,” they observed, “only when this type of private or ‘close’ corporation has given way to an essentially different form, the quasi-public corporation: a corporation in which a large measure of separation of ownership and control has taken place through the multiplication of owners.” Indeed, this separation can only be found in the United States and United Kingdom. In many other Anglo-American jurisdictions, share ownership of companies is largely concentrated. Often, in such companies, owners are also managers.

We, however, need not look further than Nigeria to get the sense of the other type of public company that combines concentration and diffusion of share ownership. Virtually all public companies in the financial and non-financial sectors have shareholders holding significant proportions of the issued shares, with some even holding a clear majority. Yet,

16. See Jensen & Meckling, Specific and General Knowledge, supra note 7, at 253.
17. BERLE & MEANS, supra note 11, at 4 (theorizing that the quasi-public corporations were dominating corporate America).
18. For a discussion of the ownership structure, see Boniface Ahuwan, Corporate Governance in Nigeria, 37 J. BUS. ETHICS 269, 271–73 (2002). One should note that Ahuwan’s description of bank structures as widespread may mask the reality, because most banks, at least before the 2005 reform, had both combined concentrated and widespread ownership. Recent developments in the banking sector, which this Article addresses, have the potential to produce more diffuse share ownership of banks.
19. Recently, the Central Bank of Nigeria (CBN) implemented a policy that requires all banks to increase their paid up share capital to twenty-five billion naira by December 2005 at the latest. For further description of this policy, see CENTRAL BANK OF NIGERIA, GUIDELINES AND INCENTIVES ON CONSOLIDATION IN THE NIGERIAN BANKING INDUSTRY (Aug. 5, 2004), available at http://www.cenbank.org/our/publications/bsd/2004/consolidationold.pdf. This policy will definitely occasion widespread share ownership of
there is, to some extent, widespread public ownership of the remaining equity. Two separate and unrelated developments in the Nigerian corporate landscape are increasingly helping to shape the ownership structure of registered companies. First, the privatization of state-owned enterprises (SOEs) in Nigeria, which favors the sale of SOEs to strategic or core investors, is bound to turn out a good number of public companies with concentrated share ownership, and also ensure widespread participation by the public in equity of the affected enterprises. While the government is to retain 40% of equity and cede 40% to strategic/core investors in the partially privatized industries, the underlying policy is to cede the management of privatized enterprises to the strategic/core investors. On the one hand, this arrangement produces investors who are not only substantial shareholders, but also retain managerial powers. On the other hand, the arrangement brings about widespread share-ownership in the public.

The second development is the decision by the Central Bank of Nigeria (CBN) to raise the shareholders’ funds of Banks to twenty-five billion naira. By forcing existing banks to either merge or approach the capital market in order to raise the requisite funds, the policy will invariably change the ownership structure of existing banks in Nigeria. It is worth noting that out of the eighty-nine banks that existed before the CBN introduced this new policy, only twenty-nine banks survived the exercise.

banks in Nigeria, which may begin a path to the separation of ownership and control in line with the Berle and Means construct.

20. For analysis of some of the factors that historically shaped the ownership structure of Nigerian companies, see Ahunwan, supra note 18, at 270–71.


Additionally, paragraph 12.3 of the Guidelines on Privatization provides that “no individual shall be allowed to acquire more than 1% equity in any enterprise whose shares are offered for sale” under the privatization program. The guidelines reinforce the philosophy of widespread ownership of the Public Enterprises (Privatization and Commercialization) Act (PEA) 1999. See PEA, paras. 5(2), 5(3).


24. See id., para. 4.
Additionally, the CBN indicated that domestic banks would be allowed to manage part of Nigeria’s external reserves if only they could meet a threshold of equity capital. This statement induced some banks to approach the capital market for more equity funds.

These developments create two possibilities. First, the share ownership structure of banks that emerged from mergers is likely to be both concentrated and dispersed. Second, banks that had to depend on the capital market to raise the required equity capital and those still canvassing that market for more funds in order to qualify are likely to have more dispersed ownership. The degree of concentration of share ownership of banks in this category will depend on how much equity they need to raise from the capital market, the spread of share ownership, and how often the company returns to the market for further funds.

It is worth noting that the corporate landscape in Nigeria is dominated by companies in which corporate managers are also significant shareholders. Here, residual claimants are also the decision agents, while minority shareholders, that is, individuals with small equity interests, have no decision-making power or authority, nor could they exert any tangible influence on decision makers. This minority group is often invisible in corporate governance due mainly to rational apathy. As scholars Cubbin and Leech noted, “if the controlling group is management operating through ownership, then increasing dispersion will have the effect of making management’s position more secure and the firm’s behavior more ‘managerial.’”

B. Agency Cost Defined

In an agency relationship, three basic propositions are crucial to a proper understanding of the concept of agency costs. First, there is a general assumption that an agent, like all human beings, is a utility maximizer. Second, it is not economically expedient for an agent to capture...

---

26. John Cubbin & Dennis Leech, The Effect of Shareholding Dispersion on the Degree of Control in British Companies: Theory and Measurement, 93 Econ. J. 351, 354 (1983). As the authors noted, management has more discretion to pursue its own objectives at the expense of shareholders than in cases where there is no controlling shareholders. Id.
27. This is defined “as a contract under which one or more persons (the principal(s)) engage another person (the agent) to perform some service on their behalf which involves delegating some decision making authority to the agents.” See Jensen & Meckling, Theory of the Firm, supra note 9, at 308.
28. For another view criticizing the classical assumptions about agents from a behavioral perspectives, see Daniel Levinthal, A Survey of Agency Models of Organisations, 9
the total value added to his or her principal’s business. Were an agent able to do so, it would be hard to imagine how the principals could make gains by engaging an agent. Third, agency from an economic perspective not only explicates specializations of tasks, but also helps to increase efficiency. These propositions accentuate the functionality of the agency concept just as they underscore the need to properly meter the agent’s performance and ensure that appropriate reward systems are provided to motivate the agent to good performance.

Michael Jensen and William Meckling defined agency costs as the sum of the monitoring expenditures by the principal, the bonding expenditures by the agent, and the residual loss.\(^29\) They further argued that, even if both the principal and agent incur monitoring and bonding costs (non-pecuniary as well as pecuniary), there would still be some divergence between the agent’s decisions and those decisions which would maximize the welfare of the principal.\(^30\) It is the dollar equivalent of such reduction in welfare experienced by the principal that Jensen and Meckling referred to as the “residual loss.”\(^31\) Jensen and Clifford Smith characterized this definition “as the sum of the out-of-pocket costs of structuring, administering, and enforcing contracts (both formal and informal) plus the residual loss.”\(^32\) As they explained, enforcement costs “include both monitoring and bonding costs, that is, resources expended by the principal and agent, respectively to ensure contract enforcement,” whereas, residual loss “represents the opportunity loss remaining when contracts are optimally but imperfectly enforced.”\(^33\) Agency costs “include all the costs frequently referred to as contracting costs, transactions costs, moral-hazard costs, and information costs.”\(^34\)

Despite its wide acceptance in the literature, the Jensen and Meckling definition requires some qualification. Bonding expenditures on the part of an agent do not exacerbate agency costs; rather, they attenuate them. Whether agents will incur bonding expenditures depends on the agents’ appreciation of their importance (in terms of their unique or peculiar

\(^{29}\) See Jensen & Meckling, *Theory of the Firm*, supra note 9, at 308.

\(^{30}\) Id.

\(^{31}\) Id.


\(^{33}\) Id.

\(^{34}\) Id.
skills or expertise and their marketability) to the principals’ businesses, the competitiveness of their positions, and the viability of a principal’s agent replacement options. More importantly, however, the pecuniary value of bonding expenditures increases the principal’s gain rather than derogates from it. Indubitably, bonding expenditures are costs to the agents; but, they are costs from which both the principal and the agent are expected to benefit.

Additionally, the above definitions are somewhat inadequate for omitting the cost effects of the various incentive programs that a principal may implement with a view to aligning the agent’s interest with his own. While incentives may take several forms, they usually have monetary value. Strictly, the cost component of an incentive (or reward) system seems to fit uneasily into the monitoring taxonomy, although both monitoring and incentive systems are directed toward the same goal: reducing profusion and shirking by the target agent. However, if inappropriately designed, an incentive or a reward system may further exacerbate agency costs. Enron, WorldCom, and Nortel provide useful explications of this point.35

C. The Dynamics Between Agency Costs and the Separation of Ownership and Control

To the extent that the reduction of agency costs depends, in part, on the ability of a principal to monitor the agent’s activities, the separation of ownership and control will remain a factor in the overall design of control mechanisms needed to reduce agency costs. But as noted above, a modern analysis of the agency problem must take into account the various shades or degrees of separation of ownership and control, since dif-

35. See Jeffrey N. Gordon, What Enron Means for the Management and Control of the Modern Business Corporation: Some Initial Reflections, 69 U. CHI. L. REV. 1233, 1242 (2002) [hereinafter Gordon, What Enron Means for the Management] (discussing stock-based director compensation as a double-edged sword: it “may enhance the board’s vigor as a shareholder agent but also increase its ambivalence about uncovering embarrassing facts that will reduce the share price”); see also PERMANENT SUBCOMM. ON INVESTIGATIONS OF THE COMM. ON GOVERNMENTAL AFFAIRS, THE ROLE OF THE BOARD OF DIRECTORS IN ENRON’S COLLAPSE, S. REP. NO. 107-70 (2d. Sess. 2002) [hereinafter ENRON REPORT] (detailing how top executives at Enron Corporation abused stock options). Recent revelations on the “financial reporting mess” at Nortel Inc. show that the company’s executives, with the acquiescence of the board of directors, adopted financial reporting methods that allowed them to report artificial profits on the basis of which they paid out bonuses that were contingent on the profits. At least twelve executives returned a total of Can$10.4 million representing the value of the bonuses they collected under the fraudulent scheme. See Tyler Hamilton, Review Rips Former CEO, TORONTO STAR, Jan. 12, 2005, at C1.
Different types of separation create peculiar delegation problems requiring special measures and attention.

The nexus between agency costs and the separation of ownership and control has received recognition for quite some time, beginning with Adam Smith. In *The Wealth of Nations*, Adam Smith highlighted the problem of agency costs by linking it with the separation of ownership and control. He stated that:

> The directors of such companies . . . being the managers of other people’s money that of their own, it cannot well be expected that they should watch over it with the same anxious vigilance with which the partners in a private copartnery frequently watch over their own. Like stewards of a rich man, they are apt to consider attention to small matters as not for their master’s honour, and very easily give themselves a dispensation form having it. Negligence and profusion, therefore, must always prevail, more or less, in the management of the affairs of such a company . . . .


Berle and Means captured the inevitability of agency problem in their book, given what they saw as the emerging pattern of ownership structure of public companies in the United States. According to them:

> The surrender of control over their wealth by investors has effectively broken the old property relationship and has raised the problem of defining these relationships anew. The direction of industry by persons other than those who have ventured their wealth has raised the question of the motive fore back of such direction and the effective distribution of the returns from business enterprise.


It is in the foregoing context that Jensen and Meckling’s observation becomes apposite.

> Since the relationship between the stockholders and the managers of a corporation fits the definition of a pure agency relationship, it should come as no surprise to discover that the issues associated with the “separation of ownership and control” in the modern diffuse ownership corporation are intimately associated with the general problem of agency.


One only needs to add, for the purposes of this Article, that it is the degree of separation of ownership and control that will determine the acuteness of the associable agency problems and the measures that should be taken to solve the problems.
III. CONTROL OF AGENCY COSTS IN PUBLIC COMPANIES

A. Nature of the Problem

A typical public company has a complex authority structure. The board is entrusted with the power of management or with the power to supervise management. As is often the case, there is a delegation of management authority to the chief executive officer (or management director) with further delegation to other top, but lower-level managers, depending on the organization’s size. In this regard, the shareholders, often far removed from operational details of a typical public company, are poorly positioned to police top management or the board of directors, because they lack material information that is key to monitoring. However, we have another genre of public companies in which share ownership is both concentrated and diffused. Most publicly listed companies in Nigeria fall into this category. The two sections following will discuss the agency-costs implications of the two types of public companies and how they may, at least, in theory, be addressed.

B. Analysis in the Context of Public Companies with Dispersed Ownership

Where no single shareholder holds significant proportion of the company’s shares to be able to influence decisions or remove directors from office, it is unlikely that the shareholder group will be able to monitor the company’s management effectively. Monitoring is not costless, as it requires a great deal of information as well as the ability to analyze, process, and utilize the information as a monitoring tool. Above all, the cost of monitoring is a key factor that is likely to shape shareholders’ monitoring behaviors. The shareholders often face collective action and free

40. See Orts, supra note 4, at 316.
42. This term has been widely written about in political and social sciences and examined from different perspectives. See Olson, supra note 25, at 53–54. According to Michael Taylor, “a collective action problem exists where rational individual action can lead to a strictly Pareto-inferior outcome, that is, an outcome which is strictly less preferred by every individual than at least one other outcome.” Michael Taylor, The Possibility of Cooperation 19 (1987).
rider problems where they are dispersed and each has a small proportion of the company’s shares. Because all the shareholders of a company will share in the benefit of monitoring, while the monitoring shareholder bears the financial burden of monitoring alone, the assumption is that shareholders of a widely held company will remain rationally apathetic to monitoring.

On the other hand, management is more organized as a unit to respond to allegations of misconduct or improprieties, knowing that shareholders are unlikely to have the necessary information to substantiate claims. Information becomes more expensive and difficult to gather where disclosure of material changes are not mandated; the only option being for the shareholders or creditors to privately bargain for the right to information on key transactions. Management controls the flow of information, and has access to the corporate treasury to prosecute proxy contests. Thus the limitation of the shareholders’ monitoring capability, measured against the real possibility that corporate managers of widely held companies may shirk or become involved in unfair dealings and positional conflicts, raises the need to devise control mechanisms that will reduce agency costs. Discussed below are some of the control devices commentators have identified for addressing agency costs in the context under consideration. The discussion will serve as a prelude to the assessment of the CAMA and other allied legislation in Nigeria.

1. Separation of Decision Management and Decision Control

Open corporations or publicly held companies are usually characterized by two distinct features, both of which help explain the nature of

43. For a recent analysis of the rational apathy theory, see Ige O. Bolodeoku, Corporate Governance in the New Information and Communication Age: An Interrogation of the Rational Apathy Theory, 7 J. CORP. L. STUD. 69, 72–76 (2007) [hereinafter Bolodeoku, Rational Apathy Theory] (discussing how the possibilities created by the Internet in information gathering, analysis and dissemination may motivate shareholders, especially retail shareholders, to be more active in corporate governance).


45. Note that other studies identify a tradeoff between dispersed ownership and managerial performance. See Mike Burkart, Denis Gromb & Fausto Panunzi, Large Shareholder, Monitoring and the Value of the Firm, 112 Q.J. ECON. 693, 694 (1997) (The authors argue that, “to the extent that managerial initiative . . . contributes to firm value, there is a trade-off between the gains from monitoring and those from managerial initiative.” Dispersed ownership, they argue, induces managerial initiative, as it implies little interference with managerial discretion, while concentrated ownership reduces those initiatives and firm value, because it facilitates excessive limitation of managerial discretion or initiatives.).
agency costs and how they should be controlled. First, shares (or the residual claims) of a publicly held company are unrestricted and freely tradable, a fact that allows residual risks to be spread across many residual claimants and enables risk bearers to diversify across companies offering such claims. This feature ensures that none of the residual claimants bears enough risk to create a real need for participation in company management. Since the capital requirement of a typical public company is large, it is cost efficient to have a wide range of individuals bear the residual risk. Second, the size of such an organization and its activities are such that specific knowledge relevant for decisions is widely diffused among agents, so that efficiencies in decision making are accomplished by delegating the decision-management rights to agents with the specific knowledge valuable to those decisions.

The need for specialized risk bearing and management inherently presents agency problems that need to be controlled. As Eugene Fama and Michael Jensen observe, “control of agency problems in the decision process is important when the decision managers who initiate and implement important decisions are not the major residual claimants and therefore do not bear a major share of the wealth effects of their decisions.” They propose that effective control procedures are necessary, because such decision managers are more likely to take actions that deviate from the interests of residual claimants. In this regard, and coupled with the necessity to delegate decision rights, Fama and Jensen suggested that an effective system of control requires that “decision management” be separated from “decision control.” Without the separation, they ar-

46. Defined as “the risk of the difference between stochastic inflows of resources and promised payments to agents.” Eugene Fama & Michael Jensen, Agency Problems and Residual Claims, 26 J.L. & ECON. 327, 328 (1983) [hereinafter Fama & Jensen, Agency Problems].


48. Jensen & Meckling, Special and General Knowledge, supra note 7, at 257.


50. Defined as the initiation and implementation of decisions. Id. at 308.

51. Id. at 308. This includes the ratification and monitoring of decisions. As Fama and Jensen noted, the devices for separating management and control include: (1) hierarchal structures in which decision initiatives of lower-level agents are passed on to agents above then in the hierarchy, first for ratification and then for monitoring; (2) boards of directors that ratify and monitor the organization’s most important decisions and hire, fire, and compensate top-level decision managers; and (3) incentive structures which encourage mutual monitoring among decision agents. Management and decision control are the components of the organization’s decision process or decision system. Id.
gued that residual claimants have little protections against opportunistic actions of decision agents, just as non-separation lowers the value of unrestricted residual claims (shares). However, “diffusion and separation of decision management and control have benefits because they allow knowledge to be used at the points in the decision process where it is most relevant and they help control the agency problems of diffuse residual claims.”

While separating decision management and decision control will not eliminate agency costs, the value of such separation lies in the fact that “the benefits of diffuse residual claims and the benefits of separation of decision functions form residual risk bearing are generally greater than the agency costs they generate.”

One clear implication of the foregoing view is that the control of management, and thus of agency costs, will be undertaken by a board of directors which must, at least, be independent of the managers, and whose main duty will be to monitor and ratify managerial decisions, in order to curb managerial shirking and opportunistic behaviors. It will appear that this approach does not actually anticipate that the board will run the company’s business and affairs; rather it anticipates that the board will undertake the task of supervising the managers. However, the optimism of effective monitoring by the board of directors is, oftentimes, illusory, since the social and economic relationship between top-level managers and members of the board can, in fact, undermine the latter’s effectiveness as monitors.

In a recent work, Boot and Macey affirm, on a new

52. Id. at 309.

53. Part of the value of separating decision management and decision control is the general administration of executive compensation by independent, outside members of the board of directors, whose role, essentially, is to monitor the performance of inside board members and determine their compensation. Such a scheme creates incentives for managers to monitor other managers, particularly when there is the possibility that hard-working managers may be promoted over the lazy ones. See Jensen & Smith, supra note 32 (“Lower level managers have an incentive to monitor managers above them because of the interdependence of their productivities, as well as the direct gains from successfully stepping over less competent managers.”).


55. Contemporary corporate governance codes or projects emphasize the need to have a corporate board that can effectively monitor management. To do this, the codes have not only suggested that a majority of the board members should be independent of management, but that the position of a managing director and that of the head of the board of directors should be occupied by two persons. See Cadbury Report, supra note 1, at 57; Committee on Corporate Governance, The Combined Code: Principles of Good Governance and Code of Best Practice (May 2000); Corporate Governance Code in Nigeria, supra note 3.

theoretical basis, the doubts expressed by some commentators that the management and directors relationship is a factor likely to undermine the objectivity of proximate monitors such as the directors.\textsuperscript{57}

The orthodoxy is that a board of directors comprising mostly of independent/outside directors will be more daring to challenge management in appropriate circumstances, and yet promote a friendly working atmosphere among the executives and non-executive members of the board.\textsuperscript{58} Indeed, it has been suggested that independent directorship is a better mechanism than regulation for the monitoring of management’s integrity, efficiency, and social responsibility. While independent directorship has become a global phenomenon, Enron, WorldCom, and other major corporate failures continue to raise doubts in the minds of many and, in particular, its critics, on whether the board alone, even with a majority of independent/outside directors, can be relied upon to effectively monitor corporate managers.\textsuperscript{59} In particular, despite the clamor for the independence of directors and the effort by many companies to promote it, corporate failures under the watch of independent directors are worse in intensity than when the consensus for independence began about thirty years ago.\textsuperscript{60}

\textsuperscript{57} See Boot & Macey, supra note 41, at 369 (“In addition to psychological barriers to objectivity, proximate boards lack objectivity from an economic perspective. Board supervision generally means that the board is jointly responsible with management for the state of the firm.”).

\textsuperscript{58} THE BUSINESS ROUNDTABLE, supra note 39, at 11–13.

\textsuperscript{59} See Gordon, What Enron Means for the Management, supra note 35, at 1241–45 (discussing the fact that, notwithstanding the appearance of independence, a corporate board could still be compromised as was the case with the Enron board); see also Brudney, Independent Director, supra note 56, at 612–14. On how to address the failures in the internal monitoring process, see Michael C. Jensen & Joseph Fuller, What’s a Director to Do? (Harvard NOM Research Paper No. 02-38, Oct. 2002), available at www.http://ssrn.com/ABSTRACT=35772 [hereinafter Jensen & Fuller, Director to Do?] (criticizing these practices as responsible for the ineffectiveness of corporate boards as monitors and recommending what the board must do to fulfill their role as the top-level corporate control mechanism); John Pound, The Promise of the Governed Corporation, in HARVARD BUSINESS REVIEW ON CORPORATE GOVERNANCE 79 (2000).

\textsuperscript{60} For a view reinforcing this position, see Alton B. Harris & Andrea S. Kramer, Corporate Governance: Pre-Enron, Post Enron, in CORPORATE AFTERSHOCK: THE PUBLIC POLICY LESSONS FROM THE COLLAPSE OF ENRON AND OTHER MAJOR CORPORATIONS 49, 76–78 (Christopher L. Culp & William A. Niskanen eds., 2003).
2. The Use of a Combination of Mandatory and Suppletory Constitutive Rules

The law matters. Undoubtedly, there is a sort of correlation between investor protection and control of agency costs. As scholars have argued:

When investor protection is very good, the most the insiders can do is overpay themselves, put relatives in management, and undertake some wasteful projects. After a point, it may be better just to pay dividends. As the diversion technology becomes less efficient, the insiders expropriate less, and their private benefits of control diminish.

Melvin Eisenberg identified certain characteristics of publicly held companies, which should help determine the proper approach to dealing with agency costs that arise in company management. First, it is Eisenberg’s view that bargaining in a publicly held company among the shareholders, or between managers and the shareholders as a body, is virtually impossible. Accordingly, he observed that most of the constitutive rules of such a company are determined not by contract, but by law or private bureaucratic rulemaking, i.e., by managerial orders; by board or committee resolutions; by board-adopted by-laws (articles of association); or by the board’s determination of governance terms in preferred stocks, stock rights, or debt instruments.

61. La Porta et al., Investor Protection, supra note 14.
62. Id.
63. First, Eisenberg defined the companies to mean those that “have a large number of shareholders, most of whom neither participate in the management of the company nor directly monitor corporate management.” Second, he observed that in a publicly held company, the number of managerial and non-managerial agents is also large, and for reasons of efficiency the business of the company is controlled and conducted by those agents. See Melvin Eisenberg, The Structure of Corporation Law, 89 COLUM. L. REV. 1461, 1471 (1989) [hereinafter Eisenberg, Corporation Law].
64. Id.
65. Eisenberg was in fact responding to the view expressed by the contractarians, such as Frank H. Easterbrook & Daniel R. Fischel, The Corporate Contract, 89 COLUM. L. REV. 1416, 1418 (1989) [hereinafter Easterbrook & Fischel, Corporate Contract] (“The corporation is a complex set of explicit and implicit contracts, and corporate law enables the participants to select the optimal arrangement for the many different sets of risks and opportunities that are available in a large economy. No one set of terms will be best for all; hence the ‘enabling’ structure of corporate law . . . .”).
66. Id. It should be noted that Eisenberg’s analysis relates to the corporate law system in the United States. Obviously, some of the powers given by statute to a board of directors are expressly dealt with by the legislature under CAMA. A board of directors of a Nigerian company lacks the competence to alter the articles of association (by-laws) without recourse to the general meeting or to vary the voting rights of the shares it issues, as these matters are regulated under CAMA in a manner that leaves no discretion to the
While recognizing that the interests of the shareholders and managers of a public company converge in many respects, Eisenberg noted that, as in any principal-agent relationship, the two interests also diverge. Basically he argued that managers of a public company are prone to shirking, outright diversion of the company’s assets to personal use through unfair self-dealing, and, more importantly, to positional conflicts. To him, while top managers of public companies might control shirking and unfair self-dealing because their self-esteem was tied to hard work and accomplishment, they might be unable to avoid positional conflicts, which, he believed, were more persistent. First, inefficient top managers are unlikely to believe themselves as inefficient. Second, they may even be unable to recognize that positional conflicts exist in the first place. Finally, they may be even more tempted to maintain or enhance their positions than they would in situations marked by shirking or unfair self-dealing.

In view of the foregoing, Eisenberg observed that “[t]he divergencies of interest between corporate agents and shareholders, and the special problem of positional conflicts, explain why some constitutive rules that govern publicly held corporations should be enabling or suppletory while other should be mandatory.” He then made the following agency costs-reducing propositions in relation to public companies:

(1) the legal rules that govern internal organization of the company and the conduct of corporate actors below the level of top managers should by and large be enabling or suppletory. The reason being that top managers have both the self-interest and the power to install constitutive rules that will efficiently determine the roles, coordination, supervision, and monitoring of lower corporate agents and constrain those agents from giving expression to their own interests in preference to the interests of the shareholders.

(2) the core fiduciary and structural rules that govern material divergences of interest of top managers in publicly held corporation should be neither determined nor subject to material variation by the action of managers or managerial organs. This is because ‘agents whose interests

67. For the definition of this term and instances of position conflict, see Eisenberg, Corporation Law, supra note 63, at 1471–72.
68. Id. at 1473.
69. Id.
70. Id.
71. Id. at 1473.
72. Eisenberg, Corporation Law, supra note 63.
may materially diverge from the interests of their principals should not have the power to unilaterally determine or materially vary the rules that govern those divergencies of interests;  

(3) the rules in (2) above should also normally not be subject to determination or material variation even with shareholder approval, because of the possibility that shareholder approval may be nominal, tainted by conflict of interest, coerced or impoverished.

In offering a rule-based approach to solving some species of agency problems, Eisenberg proposed that, “in publicly held corporations, mandatory rules should govern those core fiduciary and structural areas in which the interests of shareholders and top managers may diverge.”

For instance, he argued that, to deal with traditional conflicts of interest, mandatory rules should impose a duty of fair dealing, provide for disclosure of self-interested transactions, and establish an effective enforcement mechanism. In the case of shirking, however, he argues that mandatory rules should impose a duty of care. In essence, fiduciary duties should be mandatory at the core, and no difference should exist in this regard between closely held corporations (private companies) and publicly held corporations.

Eisenberg classified core structural rules into three main categories. One set of structural rules, he argues, should provide for the appointment and monitoring of senior executives by a governing organ (board of directors) that is elected by the shareholders for a limited term of office, a majority of which is composed of members who are independent of the senior executives. A second set of structural rules should require periodic disclosure of detailed financial data and information concerning material business and legal developments and should provide for an institutional mechanism to ensure that the information is reliable. A third set of rules should provide for approval of, and dissent from, transactions that tend to raise positional conflicts. Specifically, shareholder approval should be required for any control transaction or that which causes a change in the assets or business of the company; when a transaction involves either traditional or positional conflicts, or creates a significant potential for overreaching by control shareholders, dissenting sharehold-

73. Id. at 1473–74.
74. Id. at 1474.
75. Id. at 1480.
76. Id.
77. Eisenberg, Corporation Law, supra note 63.
78. Id.
79. Id.
80. Id.
ers should have the right to require the corporation to purchase their shares at fair value.\textsuperscript{81}

3. Markets as Agency Costs Control Mechanism

Those opposed to the rule-based approach to controlling agency costs contend that regulations overlook the contractual nature of the corporation, which requires that parties discreetly contract around agency problems, since regulation may not always produce the most efficient outcomes.\textsuperscript{82} Scholars have noted that regulations of the type contained in a corporation code were essential, as they provided standard form contracts that shareholders would likely choose. Noting that writing contracts is costly, these contractarians pointed out that a corporation statute which contained some prescriptive rules might help reduce the transaction costs of negotiations. Besides, the difficulty of anticipating all possible contingencies in long-term relationships make ex ante or default rules very helpful to lessen bargaining costs. However, they emphasize the essence of contracting or private ordering in the resolution of future disputes that could not have been anticipated and, therefore, were not specifically addressed.

While advocating an enabling structure of corporate law that enhances innovation and provides flexibility for the parties, the contractarians also make the case\textsuperscript{83} that the control of agency costs in business corporations should be left to the markets, which they argue, are most effective in aligning both the managerial and shareholders’ interests.\textsuperscript{84} However, a

\textsuperscript{81} Id.
\textsuperscript{82} See Easterbrook & Fischel, The Corporate Contract, supra note 65.
\textsuperscript{83} Part of the economic theories of the corporation see the latter as a “nexus of contract,” a perspective that not only equates the interests of other groups within the corporate family with those of the shareholders, but also challenges the view that management should be accounted to the shareholders. See Eugene Fama, Agency Problems and the Theory of the Firm, 88 J. POL. ECON. 288, 289 (1980) [hereinafter Fama, Agency Problems] (making a case for setting aside the orthodoxy that a corporation has owners, and arguing that the two functions attributed to the entrepreneur, management and risk bearing, should be treated as separate factors within the set of contracts called the firm). For critiques of this approach, see William W. Bratton, The “Nexus of Contracts” Corporation: A Critical Appraisal, 74 CORNELL L. REV. 407, 409 (1989); Victor Brudney, Corporate Governance, Agency Costs, and the Rhetoric of Contract, 85 COLUM. L. REV. 1403 (1985) [hereinafter Brudney, Rhetoric of Contract]; Melvin A. Eisenberg, The Conception that the Corporation Is a Nexus of Contracts and the Dual Nature of the Firm, 24 DEL. J. CORP. LAW 819 (1999) [hereinafter Eisenberg, Dual Nature of the Firm] (arguing that the “nexus of contracts” theory is unsatisfactory as a positive or descriptive analysis of the corporation).
\textsuperscript{84} Ralph Winter, State Law, Shareholder Protection, and the Theory of the Corporation, 6 J. LEGAL STUD. 251 (1977).
preference for the institution of the markets as ideal or better monitors means that shareholders will have no monitoring role to play. The markets that have been identified as capable of checking managerial excesses are the market for capital, the products market, the market for corporate control, and the managerial market.

The takeover market is said to provide investors with an important source of protection when the internal control mechanisms of the corporation, which operate through the board of directors, break down. According to Jensen and Ruback, the market for corporate control is the arena in which alternative management teams compete for the rights to manage corporate resources. Henry Manne argues that, besides the stock market, there is no objective standard of managerial efficiency, particularly because courts, as evidenced by the business judgment rule, are loath to second-guess business decisions or remove directors from office. To Manne, therefore, “[o]nly the take-over scheme provides

85. See Fama & Jensen, Ownership and Control, supra note 49.

86. In some cases, the market for managers and corporate control are treated as one, because the two are intertwined. See Michael Jensen, Takeovers: The Controversy and the Evidence, in WILLIAM G. KARNES SYMPOSIUM ON MERGERS AND ACQUISITIONS 27, 28 (Charles M. Linke, ed., 1986) (“The market for corporate control is best viewed as a major component of the managerial labor market. It is the arena in which alternative management teams compete for the rights to manage corporate resources.”). Granted that replacement of the chief executive office may take place without a change in control, there is still the need to separate the two analytically. As Jensen and Smith observed, “competition in the labor market tends to ensure that the manager receives only a competitive level of compensation. Reputational effects cause the value of a manager’s human capital to depend on his performance inside the firm.” Jensen & Smith, supra note 32. Since the market is likely to be aware of the role a top-level manager plays in a control transaction that occasions a change in management, to the extent that top-level managers anticipate a relocation of their human capital, they have incentives to invest in decisions that reduce divergence from value-maximizing behavior. See Fama, Agency Problems, supra note 83, at 292–93.


89. Henry Manne is credited with insight on the role of the market for corporate control in governance.

90. Manne, Market for Corporate Control, supra note 87, at 113. For the monitoring effect of the market for corporate control, see Mary S. Schranz, Takeovers Improve Firm Performance: Evidence from the Banking Industry, 101 J. POL. ECON. 299 (1993) (finding that banks, in jurisdictions where there are active takeover activities, perform better than those in jurisdictions where takeovers are restricted by legislation). But see Julian Franks & Colin Mayer, Hostile Takeovers and the Correction of Managerial Failure, 40 J. FIN. ECON. 163 (1996) (The authors studied “the disciplining function of hostile takeovers” in the United Kingdom in 1985 and 1986 and found that there was “little evidence
some assurance of competitive efficiency among corporate managers and thereby affords strong protection to the interests of vast numbers of small, non-controlling shareholders.’”

The alignment of interests occurs, and with it a possible reduction of agency costs, because “[t]he very existence of potential competition for positions of incumbent managers conditions them to think in terms of keeping stock prices as high as possible relative to other companies in the same industry.”

The capital market, it has been argued, also constrains management’s excesses and forces corporate managers to allocate the company’s assets more efficiently. Free tradability (or transferability) of the shares of publicly held companies provides an organized market, whose role is to price shares and transfer them at low costs. The stock market thus represents an external device for monitoring public companies. According to Fama and Jensen, “stock prices are visible signals that summarize the implication of internal decisions for current and future net cash flows.” Underperforming managers, the argument goes, will find it more expensive to raise capital in the market, either by way of equity capital or debt capital. Thus the competitive nature of capital and the repercussion that will attend a company’s inability to raise the needed capital are likely to force

of poor performance prior bids,” suggesting that the high board turnover did not derive from past managerial failures.

91. Manne, Market for Corporate Control, supra note 87, at 113.

92. Henry G. Manne, Cash Tender Offers for Shares—A Reply to Chairman Cohen, 1967 DUKE L. J. 231, 237 (1967). Note that some commentators have stressed the need to curb takeover activities because of its damaging effect to the company’s growth and to the economy. See Martin Lipton, Takeover Bids in the Target’s Boardroom, 30 BUS. LAW 101 (1979); Martin Lipton, Takeover Bids in the Target’s Boardroom: A Response to Professors Easterbrook and Fischel, 55 NYU L. REV. 1231, 1233 (1980) (arguing that tender offers should be left to the directors’ business judgment, as long as the economic benefits of takeovers are debatable); see also Michael C. Jensen, The Takeover Controversy: Analysis and Evidence, 4 MIDLAND CORP. FIN. J. (1986) (arguing that despite their benefits, including increased efficiency within the corporation, takeovers have created additional pressures on top-level corporate executives which may result in increased anti-takeover restrictions).


94. See Fama & Jensen, Ownership and Control, supra note 49, at 313. While this assessment may be largely correct, cases exist in which the price of a company’s stock may be simulated to the public, so that the prices do not actually reflect the true implication of internal decisions. This sort of situation happened with Enron share prices before its collapse and with the share prices of Nortel before the bubble burst. See Harris & Kramer, supra note 60, at 64–70 (detailing corporations previously falsified financial reports which were later restated).

95. See Winter, supra note 84.
corporate managers to orient a corporation’s decision process toward the interests of residual claimants.\(^96\) However, one needs to note that unless the capital market is efficient,\(^97\) it may be unable to effectively perform its monitoring role.

As noted earlier, the managerial market is linked with the market for corporate control, to the extent successful takeovers often result in the change of top management personnel. However, this market exists even outside the market for corporate control. Internally, middle level managers often aspire to displace their seniors or to gain promotion to the top, and those at the top are equally aware that they need to constantly perform well to retain their jobs. Such internal rivalry, Fama noted,\(^98\) constituted a threat to all managers, and could ultimately induce them to cut down on shirking or other value-decreasing behaviors. Additionally, the need for mobility within the industry could force managers to invest in value-maximizing behaviors, especially as the managerial market may easily access their performance records.

4. Shareholder Voting as a Monitoring Option

Few commentators concede any role to the shareholders of public companies in corporate monitoring.\(^99\) Largely, the internal monitoring discourse in corporate law focuses on and concedes to the board of directors the task of monitoring top-level management; whereby the board fails in this regard, the various markets are expected to narrow the divergence between managerial and shareholders’ interests, by serving as alternative or even concurrent monitoring mechanisms.

\(^96\) Id.


Notably, voting by the shareholders has been regulated differently, depending on the jurisdiction. The United States and the United Kingdom present contrasting approaches which also reflect differences in how policy makers in these two jurisdictions perceive the shareholders’ place in the governance scheme relative to the extent to which they think management should account to the shareholders. For instance, proxy regulation in the United States appears to reflect, over time, the lawmakers’, or the Securities and Exchange Commission’s (SEC), perception of the power shareholders have in corporate monitoring. However, since 1992, there has been increasing recognition by the SEC of the need to not only empower the shareholder in corporate governance, but also to encourage increased shareholder participation, by lessening several of the strictures which exacerbated the costs of shareholder participation. On the other hand, little regulation exists under the U.K. company legislation to limit shareholder participation and the shareholders’ right to utilize the corporate proxy machinery, as is the case under the federal proxy rules in the United States.

The good news is that not all the countries in the Anglo-Saxon jurisdictions share the U.S. position, which, in theory and practice, diminishes

100. The U.S. SEC rules on shareholder participation have developed on two fronts, namely, rules dealing with proxy solicitation and those dealing with the use by the shareholders of the corporate proxy machinery. For the historical development of the latter, see Leila N. Sadat-Keeling, Comment, The 1983 Amendments to Shareholder Proposal Rule 14a-8: A Retreat from Corporate Democracy?, 59 TUL. L. REV. 161, 166–94 (1984). Of course, the audience should note that the SEC further reviewed this rule, the aim being to simplify their interpretation and application. The latest effort at reform was the 2003 SEC proposal to obligate corporations to list shareholders nominees in the corporate proxy materials in certain circumstances. In relation to proxy solicitation, the 1992 SEC reform liberalized the concept of solicitation as to exempt certain communication by the shareholders which hitherto qualified them as solicitors and obligated the filing of proxy returns. See SEC Rules, 17 C.F.R. § 240.14a-1(l)(iv), § 240.14a-2(b)(2), § 240.14a-3(f). It was widely believed that the old expansive definition of solicitation had a chilling effect on shareholder participation. For a comparative work that put the U.S. proxy rules in perspective, see Douglas G. Smith, A Comparative Analysis of the Proxy Machinery in Germany, Japan and the United States: Implications for the Political Theory of American Corporate Finance, 58 U. PIT. L. REV. 145, 190–203 (1996).

101. Regulation of Communications Among Shareholders, SEC Release No. 34-31326, 57 FED. REG. 48276 (1992). Note that the stated purpose of the amendments contained in the Release was to facilitate communications among shareholders and to reduce the costs of complying with SEC regulations for persons engaged in proxy solicitations.


103. Abram Chayes, The Modern Corporation and the Rule of Law, in THE CORPORATION IN MODERN SOCIETY 25, 40–41 (Edward S. Mason ed., 1973) (arguing that shareholders are the least in need of protection, and that focusing on them in the corporate democracy debate is anomalous because, as a group, shareholders’ interest can be
the shareholder constituency as a monitoring body. Shareholder voting is receiving increasing attention in modern corporate governance projects, with the OECD encouraging its members to ensure that the vote matters.\textsuperscript{104} The increasing dominance of institutional investors in corporate finance has helped to renew the clamor for more of a monitoring role for the shareholders.\textsuperscript{105} For shareholder monitoring to be effective, however, the voting policy in the governing corporation code should facilitate informed decision making by ensuring that management furnishes sufficient information to the shareholders on every crucial issue considered at general meetings. In the extreme, the operative voting system should promote “vote solicitation” rather than proxy solicitation, by disengaging participation from physical presence at meetings and reducing the cost to shareholders of personally participating in general meetings. Shareholders can be allowed to participate (that is vote on issues) by electronic means, namely telephone or via the Internet. Happily, most developed countries have already undertaken reforms of their corporate laws to accommodate the use of modern communication technologies in corporate law administration and in the context of corporate governance.\textsuperscript{106}

Moreover, where, as is presently the case in the United States and in Canada,\textsuperscript{107} the use of the Internet has been integrated in the administra-
tion of both corporate and securities law to create information depository and retrieval systems (such as the EDGAR and the SEDAR) so as to allow for easy and affordable retrieval, analysis, and dissemination of corporate information, one should expect the shareholder constituency to be more active. As dissident shareholders may now use the Internet to communicate with non-proponent shareholders, coordinate monitoring activities, and transmit information, it is not unlikely that the apathy which once beset the shareholders of most public companies may wane over time, particularly with retail shareholders.108 It is expected that policymakers, especially in Nigeria, will initiate the necessary reforms so that extant communication and information technologies can become integrated into corporate governance laws relating to enforcement of shareholders’ rights as well as reforms encouraging shareholder coalitions.109

5. Control of Information Asymmetry

Information asymmetry is a critical barrier between investors and issuers of common shares.110 Bernard Black puts the issue more pointedly when he observed as follows:

The value of a company’s shares depends on the company’s future prospects. The company’s past performance is an important guide to future prospects. The company’s insiders know about both past performance and future prospects. They need to deliver this information to investors so that investors can value the company’s shares.111

Being able to value shares gives investors the chance to make rational resource allocation decisions. Given that the limited liability concept externalizes risks by shifting them to corporate creditors,112 information on material activities of public companies will assist the creditors to better

---

108. See Bolodeoku, Rational Apathy Theory, supra note 43 (discussing in detail how retail shareholders may overcome passivity by taking advantage of the possibilities of the Internet).

109. For the areas of reforms that Nigerian law may consider in new reform initiatives, see Bolodeoku, Going Virtual, supra note 107.


111. Id.

evaluate risks and to determine appropriate checks to enhance their post-bankruptcy conditions.

However, investors’ concerns are not limited to those present at the entry point. Investors should be able to evaluate the companies in which they invest to determine whether it is rational to remain as investors or relocate their investments. These concerns necessitate a regime of information disclosure that obligates corporate managers to disclose the right and sufficient information with little opportunity for distortions. As Black pointed out, “[i]nsiders have an incentive to exaggerate the issuer’s performance and prospects, and investors can’t directly verify the information that the issuer provides. This problem is especially serious for small companies and companies that are selling shares to the public for the first time.”113 It is not enough to permit corporate executives to restate profits. Sometimes, restatements come in too late, as most investors are already harmed by being prevented to bail out at the right time because they are fed with false information, as in the cases of Enron, WorldCom, and Nortel.

C. Control of Agency Costs in Public Companies with Concentrated Ownership

Dealing with agency costs or problems in public companies with concentrated ownership poses somewhat different challenges than those posed in a typical widely held public company, although the challenges are considerably similar in both. In most cases, the same group of persons undertakes the risk-bearing and management functions. Even when major shareholders do not undertake management functions, they have the power to elect the directors who will appoint top-level managers.

The fusion of risk-bearing and management functions in the same hands is efficient in many respects.114 First, if those who contribute large capital possess the technical skill required for business decisions, it will be efficient to combine residual risk-bearing and management func-

113. Black, Legal and Institutional Preconditions, supra note 110, at 786.
Allowing significant residual claimants to be decision agents may then substitute “for costly control devices to limit the discretion of decision agents,” because the fact that decision agents also bear the wealth effects of their decisions could help reduce the degree of divergence between the shareholders and management’s interests.

Where concentrated ownership is combined with weak complementary systems, such as laws that do not protect minority rights, underdeveloped capital markets, and unreliable judicial systems, there may be a significant and irremediable expropriation from the minority group. Thus, while concentrated ownership is not problematic on its own, country-specific situations may aggravate the associated agency problems.

A further characteristic of a public company with concentrated ownership is that, unlike private or closely held companies, outside residual claims may not be held by persons whose relationships with management allow control of agency problems without the separation of management and ownership.

115. One can observe that the ownership structure of privatized enterprises which the law envisages after privatization may be efficient in light of this observation.

116. See Fama & Jensen, Ownership and Control, supra note 49, at 306; see also Jeremy S.S. Edwards & Alfons J. Weichenreider, Ownership Concentration and Share Valuation, 5 German Econ. Rev. 143 (2004). First, this Article provides empirical support for the view that the extent of the conflict of interest between controlling and minority shareholders depends on the extent to which the controlling shareholder’s rights exceed her cash-flow rights. Id. at 151. Second, the authors found, while focusing on ownership structure in Germany, that while controlling shareholders, sometimes, gain private benefits at the expense of minority shareholders, “the net effects of equal increases in both the control and cash-flow rights of the largest shareholder does not harm, and, depending on the type of largest shareholder, may benefit minority shareholders.” Id. at 146. Pointing to the German corporate governance, the authors argued that it is difficult for a controlling shareholder in a German company to abuse its power, because of the admixture of the governance structure and minority rights over major corporate decisions. Id. at 147–49.

117. See Fama & Jensen, Ownership and Control, supra note 49, at 307–11; see also Ang et al., supra note 10 (confirming empirically the view of Jensen and Meckling in their 1976 seminal paper).

118. Stijn Claessens, Simeon Djankov, Joseph P.H. Fan & Larry H.P. Lang, Expropriation of Minority Shareholders: Evidence from East Asia (The World Bank, Policy Research Working Paper Series No. 2088, 1999), available at http://www.worldbank.org/html/dec/Publications/Workpapers/wps2000series/wps2088/wps2088.pdf (studying companies in East Asia, and finding that that the risk of expropriation is indeed the major principal-agent problem for large publicly traded corporations, depending on whether concentration is through family holding, state, or widely held institutions or the use of pyramidal holdings, and that expropriations ultimately depend on country-specific circumstances); Andrei Shleifer & Robert W. Vishny, A Survey of Corporate Governance, 52 J. Fin. 737, 759 (1997) (“As ownership gets beyond a certain point, the large owners gain nearly full control and . . . prefer to use firms to generate private benefits of control that are not shared by minority shareholders.”).
and control decisions. Moreover, there is no guarantee that business associates whose goodwill and advice are important to the organization are also potential candidates for holding minority residual claims. Thus, agency problems are kept alive in public companies that combine concentrated ownership and a stream of outside residual claimants, since decision agents are likely to be appointed by the residual claimants with the largest portion of claims.

The fact that decision agents may not be hundred-percent owners creates an incentive for the managers to maximize the wealth effect of their decisions at the expense of outside shareholders. One interesting point to note about this type of public company is that the existence of a shareholder or a few shareholders with high control rights does not inexorably mean that those few shareholders provide the greater portion of the capital, as control rights need not correlate with cash-flow rights, especially in countries that permit companies to issues shares with differing control rights. Another problem often associated with a public company with concentrated and dispersed ownership is the possibility that the insiders may abuse confidential information at the expense of the outside shareholders.

Jensen and Meckling observed that “[i]f a wholly owned firm is managed by the owner he will make operating decisions that maximize his utility.” And further that:

if the owner-manager sells equity claims on the corporation which are identical to his own (i.e., which share proportionately in the profits of the firm and have limited liability), agency costs will be generated by the divergence between his and those of the outside shareholders, since he will then bear only a fraction of the costs of any non-pecuniary benefits he takes out in maximizing his utility.

119. Claessens et al., supra note 118.
120. Jensen & Meckling, Theory of the Firm, supra note 9, at 313 (“As the owner-manager’s fraction of the equity falls, his fractional claim on the outcome falls and this will tend to encourage him to appropriate larger amount of the corporate resources in the form of perquisites.”).
121. See La Porta et al., Investor Protection, supra note 14.
122. Addressing insider dealing is important, at least, to provide outside investors and the public at large with some confidence and assurance of the effectiveness of the system. See Black, Legal and Institutional Preconditions, supra note 110, at 804; Brian R. Cheffins, Does Law Matter? The Separation of Ownership and Control in the United Kingdom, 30 J. LEGAL STUD. 459 (2001) (generally discussing the impact of law on the evolution of the separation of ownership and control in the United Kingdom).
123. See Jensen & Meckling, Theory of the Firm, supra note 9, at 312.
124. Id.
Jensen and Meckling further noted that as the equity of the owner-manager falls, his claims to the company free cash flows falls, the implication being that such a fall in the owner-manager share of the total profits is likely to “encourage him to appropriate large amounts of the corporate resources in the form of perquisites.”\(^{125}\) As a result, minority shareholders must expend more resources in monitoring the behavior of the managers. They must be equally ready to initiate costly derivative actions to redress possible managerial abuse of office. But if the minority group lacks the capability to monitor, other types of intervention must be employed to protect their interest.

Furthermore, when a public company combines concentrated share ownership with diffused outside equity, the shareholder-managers will usually have the votes to alter the memorandum and articles of associations. They can elect and remove directors and, in some cases, may be able to effect fundamental transactions independently of the outside shareholders. Because outside shareholders will realize that they have relatively small equity compared to the shareholder-managers, they are unlikely to have the incentives to monitor the managers.\(^{126}\)

In the Nigerian context, one of the governance problems likely to emanate from the privatized companies with concentrated ownership in core investors and scattered public ownership is that non-core investors are most likely to be rationally apathetic to monitoring. It may also be said that, if the market is efficient, in that it is able to absorb the likely tendencies of owner-managers to generate agency costs as their claim to the company free cash flow decreases, than the amount outside shareholders will pay for their shares will reflect these possible costs. But this is merely relevant at the entry point; the threat and burden of agency costs are not removed and still have to be dealt with throughout the life of the company. Above all, significant share ownership stake in a company affects its viability as a takeover target and can diminish the control function that the market for corporate control is expected to perform.\(^{127}\)

\(^{125}\) Id. It is important to note that the realization by the owner-manager that he has no total claim on the resources of the company may also induce him to shirk, rather than merely enhance his perquisites.

\(^{126}\) See Olson, supra note 25, at 50–52, 55.

\(^{127}\) See Rene M. Stulz, Managerial Control of Voting Rights: Financing Policies and the Market for Corporate Control, 20 J. FIN. ECON. 25 (1988) (showing that the fraction of share votes owned by managers is an insignificant determinant for an all-equity firm, as it affects the possibility of a tender offer and the size of premium. A higher equity stake by management, Stulz argues, would lower the probability that a hostile bid would occur.).
In this regard, internal control mechanisms, from the monitoring role of the board of directors or shareholders’ participation, may be a less effective mechanism in controlling agency costs. Information is crucial. Unless representation of the minority group on corporate boards of public companies in this category is mandated or a robust regime of information disclosure exists, major shareholders are likely to monopolize material information that the other group of shareholders may need for monitoring. Even when there is no such monopoly, it is doubtful whether outside shareholders will have the time and competence to process material information into a monitoring tool. It is pertinent to worry about how agency costs in companies that combine concentrated ownership of shares with diffused share ownership should be monitored and controlled, if the minority shareholders, most of whom may be unorganized, are not represented on corporate boards and have neither the time nor means to monitor the decisions agents.

Because of the value of information to monitoring of any sort, it is desirable to have cumulative voting through which minority shareholders can elect, at least, some directors on corporate boards. The presence of minority shareholder representatives on corporate boards could create information access and acquaint them with the activities of the shareholder-managers. The combination of concentrated and diffused ownership also requires that other forms of protection exist, which outside shareholders can take advantage of. With a good regime of information disclosure and access to information on opportunistic behaviors of the managers, outside shareholders should be able to trigger administrative investigation of the affairs of the company, and, if need be, commence derivative actions against the owner-managers or the directors loyal to him or commence a proceeding charging the directors with violating fiduciary duties. Having examined the normative issues involved in controlling agency costs in both types of public companies, the Part following will discuss how the CAMA has responded.

IV. RESPONSE OF CAMA TO AGENCY PROBLEMS

In assessing to what extent CAMA conduces to effective control of agency costs, this Part will focus on the control mechanisms that: (1) limit agency costs through either private contracting or mandatory pre-

128. As Warren Buffett once pointed out, if the controlling shareholder is an owner-manager, the board has virtually no functions. Lawrence A. Cunningham, Compilation, The Essays of Warren Buffett: Lessons for America, 19 CARDOZO L. REV. 1, 40 (1997).

129. This practice is adopted under the Canada Business Corporations Act (CBCA), Canada Business Corporations Act, R.S.C., ch. 44 (1985). The SEC in Nigeria has also recommended it in its recent corporate governance project.
scriptions; (2) reduce behavioral opportunism on the part of managers; (3) align the interests of management with those of the shareholders; (4) reduce control positional conflicts; and (5) subject managerial behaviors to the shareholders’ control.

A. Contracting Around Agency Costs and the Enabling Nature of CAMA

As noted earlier, a corporate law system that allows managers and residual claimants to contract around agency costs may, to a considerable extent, produce a more efficient result than that which dictates solutions by regulations. As a mechanism to reduce agency costs, however, private ordering is most effective when the parties’ bargaining power is equal or near equal and information cost is at its lowest.

The good news is that CAMA provides an enabling structure that members of the corporate family may utilize to reduce agency costs. However, it is impossible to bargain in every situation, cost may be prohibitive and contracting parties may have unequal bargaining power. Therefore, desirable corporate legislation should contain mandatory provisions, procedural guidelines, or managerial conduct regulation to empower managers to initiate some transactions, while retaining shareholder approval rights to mitigate the effect of conflicts of interest.

In relation to the board’s management powers and the extent to which the board’s discretion may be limited, sections 41 and 63 of CAMA provide the essential flexibility in the framework that provides the flexibility needed to craft customized solutions to agency problems. Section 63(2) of CAMA provides that, “[s]ubject to the provisions of this Decree, the respective powers of the members in general meeting and the board of directors shall be determined by the company’s articles.” This subsection is then followed by a default provision in subsection (3), which empowers the board of directors to manage the business of the company and exercise “all such powers of the company as are not by this Decree or the articles required to be exercised by the members in general meeting.”

Section 63(2) thus leaves it entirely for the company to design the intra-

130. This claim is the central focus of the contractarians. For the contractarian rendering of corporate law, see Stephen M. Bainbridge, Community and Statism: A Conservative Contractarian Critique of Progressive Corporate Law Scholarship, 82 CORNELL L. REV. 856 (1997) (book review); Jonathan R. Macey, Fiduciary Duties as Residual Claims: Obligations to Nonshareholder Constituencies from a Theory of the Firm Perspective, 84 CORNELL L. REV. 1266 (1999). For criticisms of the contract construct, see Brudney, Rhetoric of Contract, supra note 83; Eisenberg, Dual Nature of the Firm, supra note 83.


132. Id. § 63(3).
corporate power structure and to provide the desirable limits or restrictions on the powers exercisable by corporate managers.

While it is hardly expected that the shareholders, particularly of public companies, will undertake any management obligations, what is not in doubt is that the power of management may be curtailed or subjected to the shareholders’ scrutiny, even when it is conceded to the board of directors on efficiency grounds. However, until the shareholders act in that direction, the board of directors, when operating within powers conferred on it either by CAMA or the articles of association, are not bound to obey the directions and instructions of the members in general meetings, except the articles confer on the shareholders the power to override management in respect of specific matters or transactions. 133 Obviously, it will not be cost efficient for the shareholders to direct or instruct the board of directors in all cases involving the business of the company; however, the fact that directors may, sometimes, abuse their powers renders the existence of such control desirable.

For its full effect, section 63 should be read with section 41 of CAMA. While subsection (1) of the latter provision reinforces the contractual effect of the memorandum and articles of association between the company and its members and officers, and between the members and officers inter se, subsection (3) provides a legal platform for flexibility in the

133. See Companies and Allied Matters Act (CAMA), (1990) Cap. C20, § 63(4). The subsection provides as follows:

Unless the articles shall otherwise provide, the board of directors, when acting within the powers conferred upon them by this Act or the articles, shall be bound to obey the directions or instructions of the members in general meeting:
Provided that the authors acted in good faith and with due diligence.

Id. (emphasis added). It is this author’s view that the phrase “shall be bound,” as it appears in the Act, is meaningless in light of the proviso at the end of the subsection, and that perhaps the draftsman inadvertently admitted the word “not,” which would have made the phrase read “shall not be bound.” It makes no sense whatsoever to compel a board acting with the power conferred on it by legislation and articles to obey the instruction of members acting in a general meeting, when the opening phrase already indicates that the rule intended by the subsection is to operate as a default rule. This author further argues that the legislature, indeed, intended to reenact the principle in the English case of Automatic Self-Cleansing Filter Syndicated Co. v. Cunninghame, [1906] 2 Ch. 34 (C.A.), subject to the statutory “good faith and due diligence” clause. In essence, what the phrase “[u]nless the articles shall otherwise provide . . . .” implies is that the shareholders could reserve to themselves the power to instruct or direct the board in respect of the business of the company. Where such reservation has been made, it is the opinion of this author that the power so reserved shall apply even to matters in respect of which the Act or the articles have empowered the directors to act. Consequently, this provision qualifies the application of the common law principle enunciated in Automatic Self-Cleansing Filter Syndicated Co. v. Cunninghame [1906] 2 Ch. 34 (C.A.).
relationship between the company and any other party. The subsection provides that “[w]here the memorandum of articles empower any person to appoint or remove any director or other officer of the company, such power shall be enforceable by that person notwithstanding that he is not a member or officer of the company.”

Since the section refers to “any person,” it seems to provide a framework for those dealing with the company, whether in their capacity as shareholders, creditors, or officers to safeguard their interests by reserving the right to appoint directors that may protect their specific interest in the company. The directors appointed in the foregoing circumstance may be useful to bridge the information gap that may prevent the appointing party from effectively protecting his or her interest in the company.

The foregoing analysis brings into relevance management agreements contemplated under the privatization scheme in Nigeria. A management agreement may implicate corporate governance issues. Specifically, they may limit the discretion of the managers or the board of directors of the privatized companies. While there is nothing wrong with such agreements, it is important to note that, unless they are incorporated as part of the companies’ articles or memoranda, it may prove difficult to enforce the agreements. Moreover, if a term of a management agreement contradicts specific provisions of CAMA relating to the powers reserved for the general meeting, the term may become void and unenforceable.

The provisions of section 63 encourage preemptive monitoring of managerial actions, and they are best appreciated when compared with their counterpart in jurisdictions with statutory grant model. In these jurisdictions, such as the United States and Canada, the board of directors is expressly conferred with the power to manage (or to supervise the management of) the business and affairs of the corporation. Because directors enjoy a statutory grant, shareholders are correspondingly denied the right to interfere with the management powers of boards of directors. Indeed, the federal proxy rules in the United States reinforce the

134. Companies and Allied Matters Act (CAMA), (1990) Cap. C20, § 41(3).
directors’ managerial power by preventing the shareholders from utilizing the corporate proxy machinery in respect of matters that fall within the ordinary management powers of the board of directors.138

The legislative policy that prevents shareholders from interfering with the management power of corporate boards in the United States and Canada manifests in a variety of ways that may exacerbate agency costs. In the context of hostile takeovers, the statutory grant model appears to be a major factor that shapes the judicial views on the directors’ power in takeover bid transactions in these jurisdictions. By tying a board’s power of intervention in takeover bid situations to the board’s management power, the courts in these jurisdictions hold that directors can “just say no” to bids they perceive as harmful to corporate policy and effectiveness, an approach that effectively eviscerates the monitoring role that many believe the market for corporate control will perform in corporate governance.139

B. The Deregulation of Limited Liability

The limited liability rule implies that corporate managers and shareholders do not bear the cost of their decisions,140 and except when sued

138. The U.S. position contrasts sharply with the position in Nigeria, Canada, the United Kingdom, and other common law jurisdictions where shareholders are not disallowed from using the corporate proxy machinery in respect of matters falling within the ordinary powers of the board of directors. For Canada, see Canada Business Corporations Act, R.S.C., ch. C-44, § 137; for Nigeria, see Companies and Allied Matters Act (CAMA), (1990) Cap. C20, § 235; for United Kingdom, see Companies Act, 2006, c. 46, § 294.

139. The Supreme Court of the State of Delaware gives authority for the defensive tactics mounted by target boards against takeover bids, even if the shareholders would have preferred the bid. See Paramount Communications, Inc. v. Time, Inc., 571 A.2d 1140, 1154 (Del. 1989). The court held that:

Delaware law confers the management of the corporate enterprise to the stockholders’ duly elected board representatives. The fiduciary duty to manage a corporate enterprise includes the selection of a time frame for achievement of corporate goals. That duty may not be delegated to the stockholders. Directors are not obliged to abandon a deliberately conceived corporate plan for a short-term shareholder profit unless there is clearly no basis to sustain the corporate strategy.

140. This is the implication of the decision in Salomon v. Salomon & Co., [1897] A.C. 22, 35 (H.L.) (appeal taken from Eng.) (U.K); see also Manne, Our Two Corporation
Corporation may escape being held accountable for misappropriation of corporate assets or shirking. Moreover, the concept of limited liability, while shielding residual claimants from the cost of investment, externalizes the associated risks by transferring them to creditors.\footnote{systems, supra note 15, at 262 (“[L]imited liability . . . allows individuals to use small fractions of their savings for various purposes, without risking a disastrous loss if any corporation in which they have invested becomes insolvent.”).} In a way, therefore, limited liability has the potential to aggravate agency costs. The deregulation of limited liability by CAMA may, therefore, be seen as an insightful way to deal with agency problem, as those dealing with the company may force its directors to tie their fortunes to the success of the company and, by so doing, induce them to bond to good performance.\footnote{see, e.g., walkovsky v. carlton, 223 N.E.2d 6 (N.Y. 1966). For criticisms of the Salomon decision, see O. Kahn-Freund, Some Reflections on Company Law Reform, 7 Mod. L. Rev. 54 (1944); see also Jonathan M. Landers, A Unified Approach to Parent, Subsidiary, and Affiliate Questions in Bankruptcy, 42 U. Chi. L. Rev. 589 (1975) (arguing that limited liability transfers the risks of business failure from shareholders to creditors without compensation); Halpern et al., supra note 47, at 129 (“The existence of a limited liability regime provides the owners of the firm with the ability to obtain insurance, provided by the creditors, against loss on default without the necessity of the existence of a formal insurance market.”). But see Richard A. Posner, The Rights of Creditors of Affiliated Corporations, 43 U. Chi. L. Rev. 526 (1967) (contending that creditors are often compensated for the risks they bear through high interest rates, which they often charge because of the limitations imposed by the limited liability rule).} The deregulation of limited liability by CAMA is even more commendable in view of the fact that some modern corporate legislation now requires companies to pay for directors’ insurance\footnote{See, e.g., Companies Act 1985, c. 6, § 310(3) (Eng.). For Canada, see Canada Business Corporations Act, R.S.C., ch. 44, § 124(6) (1985).} or excuse directors from liability for breach of fiduciary duties.\footnote{Del. Code Ann. tit. 8, § 102(b)(7) (2006). The practice is also prevalent in the United Kingdom since the applicable law now permits companies to pay for directors’ insurance; see also Brian R. Cheffins & Bernard S. Black, Outside Director Liability Across Countries, 84 Tex. L. Rev. 1385 (2006).}

Section 288(1) of CAMA states that, if so provided by the memorandum, the liability may be unlimited for directors, managers, or managing
directors of a limited liability company.\textsuperscript{145} Even where the memorandum does not contain an unlimited liability provision at registration, it is still possible to attain such a result, so long as there is a provision in the by-laws permitting the company to alter its memorandum by special resolution to render unlimited the liability of the company’s directors, managers, or managing director.\textsuperscript{146} In practical terms, this device is likely to be unpopular with most directors or managing directors. But where a company is highly leveraged, the company’s creditors may require that the company’s memorandum be altered to permit for unlimited liability of directors or managing directors.

While a company is at liberty to opt-in or opt-out of the deregulation of limited liability, it is suggested that companies in which inside ownership is significant and concentrated, and in which ownership and management are combined, may be required to make the liability of the owner-manager unlimited. Since the removal of limited liability will internalize the risks of investment and decision making, it is expected that the company’s creditors, who are the likely beneficiaries, will find the deregulation of limited liability under CAMA particularly useful. To the extent that the managers bear the wealth effects of their decisions, other residual claimants will also benefit, as the managers may be forced to significantly reduce shirking or opportunistic behaviors. Deregulation of limited liability under CAMA is particularly suitable to the peculiar challenges of corporate governance in Nigeria. Most public companies have majority owners who are most likely to act as managers. Since the usual fear is that owner-managers may disregard the interests of minority shareholders, the latter may have their fears attenuated because the threat of expanded liability will probably align the interest of owner-managers with those of the minority shareholders.\textsuperscript{147}

\textsuperscript{145} Note that the deregulation of limited liability under CAMA is different from a typical case of unlimited liability under which shareholders are required to satisfy the excess of the corporation’s debts from their personal wealth. The CAMA device is selective in that it is only the liability of top management or the directors that may be unlimited. Whether a company that opts into this regime will be able to attract quality personnel to serve as directors is another question. The same result may, however, be achieved where the chief executive officer and most members of the board of directors are major shareholders.

\textsuperscript{146} See Companies and Allied Matters Act (CAMA), (1990) Cap. C20, § 289(1).

\textsuperscript{147} The usual recourse or remedies open to minority shareholders, i.e., derivative action, oppression, or prejudicial remedy action, require a great deal of cost or information to prosecute. The monitoring that comes with the deregulation of limited liability is a significant improvement over the traditional remedies under sections 303 and 311 of CAMA.
C. Conflicts of Interest, Self-Dealing, and Positional Conflicts

Melvin Eisenberg is emphatic on the normative proposition in this regard for the modern company law: “agents whose interests may materially diverge from the interests of their principals should not have the power to unilaterally determine or materially vary the rules that govern those divergencies of interest.”148 A related proposition is that such rules should also normally not be subject to determination or material variation even with shareholder approval. In essence, the governing rules should be mandatory at the core. How does CAMA respond?

1. Negotiation of Directors’ Liability

In the first decade of the twentieth century, the English courts enforced provisions in the articles of association of registered companies, which freed directors of liability for breaches of fiduciary duties.149 However, the practice was soon outlawed under the English company law.150 Contemporary corporate law in the United States, courtesy of the amendment to the Delaware corporation code, permits corporations to free directors from liability for a breach of fiduciary duty of care.151 In addition, state corporate laws in the United States, Canada, and in the United Kingdom empower companies to buy officers and directors’ liability insurance policies.152

Provisions in corporate codes which exempt directors and officers from liability for breaches of fiduciary duties and those which empower the purchase of liability insurance policies arguably gain prominence as a compromise or trade-off for securing the services of qualified personnel to serve on corporate boards, many of whom had expressed concerns about the magnitude of liability relative to the gains they expected to make from serving as directors.153 However, it can hardly be doubted that a no-liability regime may induce directors to shirk or engage in behaviors inimical to the interest of the company. If directors knew they would not bear the wealth effects of their conducts, fiduciary duties might not motivate them to invest the time necessary to run the com-

148. See Eisenberg, Corporation Law, supra note 63, at 1474.
149. See In re Brazilian Rubber Plantation & Estates Ltd. [1911] 1 Ch. 425, 440 (U.K.).
150. See Companies Act, 19 & 20 Geo. 5, c. 23, § 146 (1929) (Eng.).
152. See supra note 143.
153. See Cheffins & Black, supra note 144.
pany’s business efficiently. The position of a director is a crucial one, in view of the predictable impact that an inattention to serious business matters could have on the investors.

It is in this regard that the position under CAMA deserves some commendation. The Act provides that:

no provision, whether contained in the articles or resolutions of a company, or in any contract shall relieve any director from the duty to act in accordance with this section or relieve him from any liability incurred as a result of any breach of the duties conferred upon him under this section.\(^{155}\)

This position can be compared with those in other jurisdictions, e.g., the United States, where a company may excuse directors from liability in its bylaws or obligate the company to take out insurance policies regarding directors’ liability.\(^{156}\)

2. Treatment of “Golden Parachute” by CAMA

Golden parachute is a term associated with American corporate law, and refers to a payment (or agreement to make a substantial payment) made to directors or the managing director of a target company as consideration for loss of office following a takeover of the company, whether by friendly or hostile bid or even in a merger transaction.\(^{157}\) At its worst, a golden parachute is one defensive mechanism used to insulate management from the threat of a takeover, by making it costlier for a bidder to launch a bid or ensuring that a putative acquirer goes through the board rather than bypass it in a hostile bid transaction.

While not so recognized in Nigeria, this concept is not completely new to our laws. Given that most public companies have significant majority owners, most of whom are in management positions, most directors and

\(^{154}\) It is worthy of note that the contractarians have argued that it is efficient for the parties to be able to negotiate or contract around fiduciary duties, since such negotiation would reflect the parties’ wishes. See Frank H. Easterbrook & Daniel E. Fischel, Corporate Control Transactions, 91 YALE L.J. 698, 734 (1982). For a criticism of this conception, see Victor Brudney, Contract and Fiduciary Duty in Corporate Law, 38 B.C. L. Rev. 595 (1997) (arguing the reasons why the corporate law fiduciary duties, contrasted with commercial agent-principal fiduciary duties, should not be variable at the instance of the parties); Deborah A. DeMott, Beyond Metaphor: An Analysis of Fiduciary Obligation, 37 DUKE L.J. 879 (1988) (criticizing the Jordan court for analogizing fiduciary duty with “off-the-rack” rules).

\(^{155}\) See Companies and Allied Matters Act (CAMA), (1990) Cap. C20, § 278(8).

\(^{156}\) See Hammond, supra note 142, at 544–45.

corporate managers in Nigeria may become prone to using such an exit device. In view of the concentrated ownership structure of Nigerian companies, a takeover of a company in Nigeria may not take place without the approval of the company’s board. Thus, the increased likelihood of merger transactions as the means to effect changes in the control of companies in Nigeria may pave the way for the target management to secure its financial future with a robust exit price.

CAMA outlaws any payment by way of compensation for loss of office, or as consideration for or in connection with retirement from office, by a company to any director of a company, except when particulars of the proposed payment and other allied terms have been disclosed to and approved by members of the company.158 Additionally, an exit payment cannot be made without meeting those same requirements.159 A payment made to the director in contravention of the Act is regarded as illegal, and the director shall be deemed as trustee for the sum so received.160

Moreover, CAMA imposes a duty on a director who expects to receive any such payment following a transfer of the company's shares to a person or body corporate to ensure that the particulars of the proposed payment are included in the notice of the offer sent to the shareholders.161 What is unclear, however, is whether the interested director is entitled to vote in respect of the proposal or payment when it is tabled at the meeting.

The 1999 Investment and Securities Act (ISA) is, however, likely to introduce a twist, in that section 111(6) of the ISA only requires that particulars of any proposed payment to directors in the context of a takeover bid be contained in the directors’ circular sent to the shareholders. Such disclosure is not required for merger transactions, and ISA does require that the provisions of “all existing enactments,” which include those of CAMA, “shall be read with such modification as to bring them into con-

158. See Companies and Allied Matters Act (CAMA), (1990) Cap. C20, §271. Note that in the United States and Canada, issues relating to directors’ remuneration or compensation are not for the shareholders to determine in general meeting of members. Directors’ compensation falls squarely in the category of the company’s business which only the board of directors is empowered to handle. For Canada, see Canada Business Corporation Act, R.S.C., ch. C-44, § 102 (1985), and for Delaware, see Del. Code Ann. tit. 8 § 141(a) (2007). In the respective jurisdictions in which they apply, these provisions confer on corporate boards the power to manage or supervise the management of the businesses of incorporated companies.

159. Companies and Allied Matters Act (CAMA), (1990) Cap. C20, §272(1).

160. Id. § 272(2).

161. Id. §§ 273(1), 274.
formity” with the provisions of ISA. In respect of takeover bid transactions, the more permissive ISA provisions may negate the purpose of sections 271 to 273 of CAMA, except the control which CAMA provisions are designed to serve is recognized and adopted in ISA.

D. Removal of Directors from Office

Agency costs are often exacerbated when directors cannot be removed from office in deserving circumstances. Removal of directors is tied to so many issues, all of which are dealt with under CAMA in ways that limit the directors from insulating themselves from removal from office. The issues range from the procedure for removal, the voting rights attached to shares, shareholders coalition, and the right to issue shares. For instance, under CAMA, directors are removable from office with or without cause, irrespective of the contract between the director and the company. The best option open to a director so removed from office is to claim compensation or damages in an action for breach of contract.

This position contrasts sharply with the state of Delaware, where directors cannot be removed without cause if the certificate of incorporation establishes staggered terms for directors. Although a staggered board system, a practice that dominates American corporate law, is theoretically possible in Nigeria, such practice is of little practical effect given the provision of section 262(1), which allows for the removal of directors “notwithstanding anything in its articles or in any agreement between it and him.”

Besides, the mode of appointment of directors also has some impact on how directors are removed and on agency costs. The American practice, whereby management nominates a slate of directors for the shareholders to vote on, does not help to ensure directors’ accountability to sharehold-

---

162. Investments and Securities Decree No. 45 (1999), § 261(1) (Nigeria). Note that the Investment and Securities Act, which was originally promulgated as a decree under the then Military Administration, like other decrees, is to be so called by virtue of section 315(1)(a) of the Constitution of the Federal Republic of Nigeria (1999).

163. See Companies and Allied Matters Act (CAMA), (1990) Cap. C20, §§ 262, 268(2); see also Longe v. First Bank of Nigeria Plc [2006] 3 NWLR (Pt. 967) 228 (Court of Appeal) (affirming the judgment of the lower court which held that the Bank had the right under the CAMA to remove the appellant, the bank’s managing director, from office).


165. Section 259(1) of CAMA dealing with the rotation of directors is a default provision that allows a company to make different arrangement as to retirement of directors in the articles. Thus a company is at liberty to provide for a staggered board of arrangement, under which only a portion of the board members can retire after a three-year term.

166. Companies and Allied Matters Act (CAMA), (1990) Cap. C20, § 262(1).
ers. To challenge directors, shareholders of U.S. corporations are often forced to spearhead proxy contests, which are few and far between because the contests are usually very expensive.  

In Nigeria, however, shareholders are at liberty to propose any person as a director, and there may also be an agreement in the article empowering named persons to appoint directors. Shareholders may submit a proposal for a resolution to remove directors or on any matter, and equally submit written statements in respect of matters to be transacted at general meetings. Although it is required that reasonable payment be made for transmission, the company is obligated to circulate a statement of resolution requisitioned by qualified shareholders. While the information contents of the new proxy rules under the SEC regulations made pursuant to ISA may make shareholder participation in the proxy process more informed, it is yet to be tested what impact the rules will have on the shareholders, given that proxy solicitation is expansively defined without taking care of the problems similar definition has created for the shareholders in other jurisdictions, as a result of which there have been some clarifications in the affected jurisdiction in order to reduce the chilling effect of the definition.

E. Directors’ Remuneration

Remuneration of top-level management and of the directors is a key determinant of agency costs. If it is right, remuneration can align man-
agement interest with those of the shareholders. It is the theory that, when tied to the company’s stocks, directors’ compensation or remuneration engenders in management and directors a sense of identification with, or creates a stake in, the company. Uncontrolled, however, remuneration can constitute a burden on the company’s free cash flow, which can hurt the corporation, its creditors, and shareholders.

Outrageous compensation packages for corporate executives and their exponential increase relative to the value added have had their roles in the collapse of many recent celebrated corporate failures. Considering American management remuneration, Kevin Murphy wrote that from 1992 through 2000, the median total compensation of CEOs in S&P 500 companies nearly tripled from US$2.3 million in 1992 to US$6.5 million in 2000. Option-driven escalation in CEO pay levels, Murphy observed, “is not limited to S&P 500 Industrials.” Evidence shows “that median pay in S&P 500 Financial Services companies increased 300 percent, form $2.6 million to almost $11 million from 1992 to 2000, while pay in smaller firms (defined as companies in the S&P MidCap 400 and SmallCap 600) more than doubled, from $823,000 to $1.8 million.” Some commentators attribute exorbitant executive compensation packages to the domineering influence of CEOs over compensation committees. Lucian Bechuck et al. argued that, “much of what is observed in the world of executive compensation” can be explained by managerial opportunism and influence over captive boards of directors. Although

175. See Michael C. Jensen, CEO Incentives—Its Not How Much You Pay, But How, in CORPORATE GOVERNANCE AT THE CROSSROADS 192, 195 (Donald Chew, Jr. & Stuart Gillian, eds., 2005) (“The most powerful link between shareholder wealth and executive wealth is direct ownership of shares by the CEO.”).
176. See ENRON REPORT, supra note 35.
177. The Enron case still presents a classic example of how ambitious and outrageous compensation packages could hurt the company. Id.
178. See Kevin M. Murphy, Explaining Executive Compensation: Managerial Power Versus the Perceived Cost of Stock Options, 69 U. CHI. L. REV. 847, 847 (2002) (the increase in CEO pay in S&P 500 Industrialists during the 1990s primarily reflects a dramatic growth in stock options, which swelled from 27% to 51% of total compensation, representing a five-fold increase in dollar terms).
179. Id.
180. Id.
CEO compensation packages in the United States are higher than their international counterparts, \(^{183}\) there seems to be a global upward swing in CEO pay that is hardly justifiable in terms of performance. \(^{184}\)

Today, directors’ compensation constitutes a major issue in any corporate governance reform, principally because the costs they add are sometimes too burdensome for the company and too high to benefit the corporation, its shareholders, and other stakeholders. \(^{185}\) As it is consistent with modern corporate practice to categorize executive compensation as belonging to the realm of the company’s business, only the board of directors (through its compensation committee) gets to decide issues relating to compensation. However, experiences have shown that advertised independent boards or committees are not as independent, given the complicated business ties most directors have with the corporations on whose boards they serve. The fact that the chief executive has a hand in the nomination of directors, independent or interested, is also a factor in gauging the level of resistance to expect from directors on the issue of compensation. \(^{186}\)

Under CAMA, remuneration of directors is not fixed by the board of directors, but by the company in a general meeting, and is payable only out of the company fund. \(^{187}\) Besides, it is unlawful for the company to pay directors’ remuneration free of income tax, or calculated by reference to or varying with the amount of his income tax, except by a contract to that effect, which must exist at the time of the Act and be contained in the articles. \(^{188}\) It is also unlawful for a company to make a loan to any of the directors of its holding company, or enter into any guaran-

---

183. See Thomas, supra note 181, at 1173.
184. See ENRON REPORT, supra note 35.
186. Bebchuck, Fried, & Walker, supra note 182; Brudney, Independent Director, supra note 56, at 612–14. Note that Murphy criticizes the view that explains high CEO compensation in terms of the CEOs desire to extract rent. See Murphy, supra note 178, at 850 (distinguishing between CEOs ability to extract rent and bargain for higher pay); Jensen & Fuller, Director to Do?, supra note 59, at 4–6.
187. See Companies and Allied Matters Act (CAMA), (1990) Cap. C20, § 267. The opportunity given to the shareholders to consider directors’ remuneration is important, especially at a time when the likelihood that directors may drain the corporation through sumptuous remuneration package is on the rise.
188. See id. § 269.
tee or provide any security in connection with a loan made to directors by any other person.  

While CAMA expressly requires that the company approve directors’ remuneration, it provides that the board of directors shall fix the remuneration of the managing director. Without clear guidance on the factors to be considered in fixing remuneration, it is obvious that the possibility of abuse of this power by a subservient board, as seen in other jurisdictions, cannot be ruled out, notwithstanding that compensation committees are a common feature of most board structures and in corporate governance projects.

It is suggested that shareholder involvement in fixing executive compensation is desirable, particularly when it involves stock options, as well as the publication of executive compensation figures in the company’s proxy statements. In Nigeria, such publication will not only provide shareholders with information on executive compensation, but also precipitate possible review, critique, and intervention by the shareholders of the company where the compensation is considered to be outrageous. It should be remembered that, unlike in the United States or Canada where shareholders are not legally entitled to intervene, shareholders of companies registered in Nigeria can, by law, make recommendations to the board regarding any action to be taken by the board, notwithstanding that the board is empowered to manage the business of the company.

Under CAMA, shareholders participate in determining directors’ remuneration or compensation in one significant respect. In relation to the employment contract of a director intended to be employed for more than five years where termination by the company is limited by agreement, the company is required to table the agreement with its terms before the general meeting for a special resolution approving the agreement.

F. Dealing with Conflicts of Interest

CAMA has some mandatory provisions to deal with directors’ conflicts of interest, outlawing certain conduct or mandating disclosure, while leaving others for the shareholders to decide. CAMA also deals with di-

189. See id. § 270.
190. See supra note 35 and accompanying text.
192. See Companies and Allied Matters Act (CAMA), (1990) Cap. C20, § 63(5).
193. See id. § 291.
rect self-dealing by directors.\textsuperscript{194} For instance, promoters are regarded as fiduciaries, and are accountable for any secret profit made or for misuse of corporate information.\textsuperscript{195} The company may rescind any contract, unless after the promoter’s full disclosure of all material facts, the contract is ratified by the company’s independent directors and all the members of the company, or by the shareholders in a general meeting at which the promoter and other interested shareholders are disqualified from voting.\textsuperscript{196}

While the law does not forbid directors from dealing with their companies, it requires that interested directors, whether directly or indirectly, in a contract or proposed contract with the company should declare the nature of their interests at a meeting of the directors where the contract is first discussed.\textsuperscript{197} Furthermore, the Act specifically disqualifies, subject to certain exceptions,\textsuperscript{198} a company from entering into any arrangement under which a director of the company or its holding company or a person connected\textsuperscript{199} with such a director acquires or is to acquire one or

\begin{enumerate}
\item\textsuperscript{194} For the importance of controlling self-dealing, see Black, Legal and Institutional Preconditions, supra note 110, at 804–15.
\item\textsuperscript{195} See Companies and Allied Matters Act (CAMA), (1990) Cap. C20, § 62(2).
\item\textsuperscript{196} See id. § 62(3). The fact that CAMA relaxes the limitation period in the enforcement of the company’s right against promoters is equally important in the reduction of agency costs.
\item\textsuperscript{197} See id. § 277. It is important to note the effect of a failure to comply with this provision. First, section 277(3) provides that a general notice that the director is a member of a specified company interested in the contract or that the director is interested without more is sufficient. Id. § 277(3). Moreover, failure to comply only attracts a small amount in penalty, except that subsection (5) preserves any rule of law restricting directors of a company from having any interest in contracts with company. Id. § 277(5). It is here suggested that specific disclosure of the particular interest of a director is necessary, considering that such disclosure is made to his or her colleagues, some of whom may change their minds or decide differently if they know the real nature of the transaction in which the disclosing director is involved.

On a practical note, Bernard Black has argued that only independent outside members of the board should consider the merit of a transaction in which a director is interested. See Bernard Black, The Core Fiduciary Duties of Outside Directors, 2001 Asia Bus. L. Rev. 3 [hereinafter Black, Core Fiduciary Duties] (“The procedural strategy for approval by disinterested directors can work only if a company has a reasonable number of independent directors. It can work \textit{well} only if these directors are \textit{in fact} independent of the executives. Otherwise, the procedures can become camouflage for a transaction that in fact benefits the insiders at the company’s expense.”).

\item\textsuperscript{198} For other exceptions see Companies and Allied Matters Act (CAMA), (1990) Cap. C20, § 286. Other subsections of this provision also defines the liability of the directors for a contravention of the provisions of section 284 of the Act.
\item\textsuperscript{199} Section 286(8) defines several circumstances under which a person “is connected with” a director. Id. § 286(8). The expansive definition is essential to avoid sham transac-
more non-cash assets of a specified threshold, or the company is to acquire similar assets from the persons aforementioned. Before a company can enter into an arrangement in this context, it is a precondition that the arrangement is first approved by a resolution of the company in general meeting, and if the director or connected person is a director of its holding company, by a resolution in the general meeting of the holding company. It is significant to note that CAMA codifies the common law rule on the competence of directors to take advantage of corporate opportunities, and tapers their impact by a requirement of ex ante disclosure to the company in general meeting.

It is also crucial to the discussion of agency costs to note that directors are treated as trustees of the company’s moneys and properties, and are to exercise their powers honestly in the interest of the company and all shareholders, rather than in their own or sectional interests. In this regard, the Act forbids a director from making secret profits by accepting from any person a bribe, gift, or commission or a share in the profit made by that person from a transaction involving his or her company as a quid pro quo for facilitating the transaction between the company and the person. However, if the gift is unsolicited or given as a sign of post-transaction gratitude, the director may keep the gift, provided that the fact of the gift is reported to the board of directors and a note is made in the minutes book of the directors. The problem with this allowance is that the prevailing corruption culture in Nigeria may induce ex ante negotiations of rewards for facilitating a transaction. Consequently, prohibition of such practice under CAMA may prove very ineffective to arrest the practice. By permitting post-transaction rewards or gifts, parties may circumvent the law if the beneficiary of the transaction simply presents the reward afterwards in order to give it a cloak of legality.

\[
\]

\[
\text{\textit{Id.} (1).}
\]

\[
\text{\textit{Id.} (2). See id. § 284(1).}
\]

\[
\text{\textit{Id.} (3). See id. § 287(1).}
\]

\[
\text{\textit{Id.} (4). See id. § 287(3).}
\]
G. Shirking, Agency Costs, and the Directors’ Duties of Care and Skill

Agency costs are not generated only when agents divert the property or assets of their principals. Refusal to pay attention to duty or to invest the necessary time required by the business, commonly referred to as shirking, can generate agency costs. As investigations into the role of directors in the collapse of Enron showed, shirking played a vital role in the collapse process. The Enron directors were found to have given far less attention to their responsibilities and were highly dependent on management, a situation which prevented them from being acquainted with the details of operations and from asking critical questions of management. Happily, CAMA specifically gives some attention to shirking.

The Act recognizes that lax duties of care and skill will undoubtedly exacerbate agency costs, as they may induce directors to shirk. The Nigerian law reformers found the common law principles on the directors’ duties of care unsatisfactory and recommended a more functional set of duties that underscore the seriousness of the role of directors in corporate governance. Regardless of how directors are appointed to serve on a company’s board, it is this writer’s view that section 282(4) of CAMA appears to enjoin directors to recognize the enormity of their role and be ready to give to it the time and commitment the duty requires. The section also expects directors to candidly decline the offer to serve if, given the particular situation, they cannot give enough attention to the company’s business.

The CAMA provision, which calls on directors to exercise “that care and skill which a reasonably prudent director would exercise in comparable circumstances,” is an invitation to the court to apply an objective standard to the conduct of directors in light of the peculiar circumstances of each case. In this respect, it is expected that courts will be assisted

208. See ENRON REPORT, supra note 35 (“The Board witnessed numerous indications of questionable practices by Enron management over several years, but chose to ignore them to the detriment of Enron shareholders.”).

209. Id. at 59.

210. The provisions of the CAMA on the duty of care and skill of directors effectively undermine the rationes decidendi in cases such as In re Equitable Fire Insurance Co. Ltd., (1925) Ch. 407, and In re Denham’s & Co., (1884) 25 Ch.D. 752 (Eng.).

211. See supra note 198.

212. Id. In fact, Nigeria seems to be ahead of the United Kingdom in this respect, except that, through judicial activism, the U.K. courts can achieve the same objective as that in section 284 of CAMA. See Norman v. Theodore Goddard, [1992] B.C.C. 14 (Ch.D.) (Eng.); see also Cheffins & Black, supra note 144.

213. Nigerian courts may be guided by the recent decision of the Supreme Court of Canada in People Department Stores Inc. (Trustees of) v. Wise, [2004] 3 S.C.R. 461 which emphasizes the objective component of the assessment of whether a director
in determining what a reasonably prudent director will do by the admonitions contained in various corporate governance projects, just as the various opportunities offered by companies to train directors in the modern art of governance may become part of the court’s assessment of what to expect from a director. Specifically, it is no longer a defense to a claim of breach of the duty of care when a director was absent from meetings, because “the absence from the board’s deliberations, unless justified, shall not relieve a director from his or her responsibility for the actions of the board.”

More importantly, CAMA imposes the same standard of care on both executive and non-executive directors. The emphasis on equality in responsibility and dedication expected from both the executive and non-executive directors also addresses the propriety of outside directors conceding to management on most crucial or critical issues of governance. At the minimum, it signifies that outside directors should see themselves more as business managers than mere advisers. By implication, section 282(4) of CAMA requires non-executive (outside) directors to be more proactive in the discharge of their duties. As the Court of Appeal of New South Wales (Australia) rightly observed in Daniels v. Anderson, the “concept of a sleeping or passive director has not survived and is inconsistent with the requirements of current company legislation such as, at the relevant time, [sections] 229 and 269 of the Companies (New South Wales) Code.”

Arguably, a court of law may be justified in re-breached the fiduciary duty of care under Canada Business Corporations Act, R.S. 1985, c. C-44, s. 122(1)(b)).

214. See Bishopsgate Investment Management Ltd v. Maxwell, [1993] B.C.C. 120, 139 (Eng.) (“[T]he law may be evolving in response to changes in public attitudes to corporate governance.”). The global nature of the concerns in corporate governance, which is gradually leading to some form of convergence of practices, should also be crucial to how the courts should perceive the position of directors in the modern company. See Black, Core Fiduciary Duties, supra note 197.

215. See Companies and Allied Matters Act (CAMA), (1990) Cap. C20, § 283(3).

216. See id. § 282(4).

217. See Myles L. Mace, Directors: Myth and Reality, in Theories of Corporate Governance: The Philosophical Foundations of Corporate Governance 96, 99 (Thomas Clarke ed., 2004) (“the lack of active discussion of major issues at typical board meetings and the absence of discerning questions by board members result in most board meetings resembling the performance of traditional and well-established, almost religious rituals.”). For an account of the traditional relationship between directors and management, see also Robert Monks, The Director’s New Clothes (Or, the Myth of Corporate Accountability), in Corporate Governance at the Crossroads 151 (Donald Chew, Jr. & Stuart Gillian eds., 2005).


219. Id. at 310.
quiring that non-executive directors establish the basis of their judgment, by showing the efforts made outside the assistance given by the inside members of the board to gather more information, particularly if the facts of the case show that more information is required before any good decision can be taken. While non-executive directors may simply rely on the presentations of inside members of the board, the court, arguably, may hold that, given the nature of what is being deliberated upon, a proper discharge of the directors’ duties will require non-executive directors to seek independent information from other key employees of the company. This sort of provision is quite commendable, as they attune with the modern dictates of good governance. Besides, they could, if properly given a purposive interpretation by the court, be effective in addressing some of the agency cost concerns.

It is important to note that a breach of the directors’ duties may be asserted and enforced in the course of the winding up of a company. Section 506(1) of CAMA allows a liquidator, receiver, and creditor, among others, to prove against any person that the latter participated in carrying on the business of the company recklessly or with intent to defraud its creditors. The court is empowered to hold such a person personally liable “without any limitation of liability for all or any of the debts or other liabilities of the company.” In particular, this provision is a warning to outside directors, and it reinforces the need for them to be more involved.

220. Nigerian courts can learn a lot from Australian decisions on the scope of directors’ duty of care and skill. Interestingly, the recent judgments in Australian courts which espouse the modern principles of directors’ duties of care, skill, and diligence were inspired by the judgment of Judge Pollock of the Supreme Court of New Jersey in Francis v United Jersey Bank, 432 A.2d 814 (N.J. 1981). In Francis, Judge Pollock stated that a director: (a) should become familiar with the fundamentals of the business in which the corporation is engaged; (b) is under a continuing obligation to keep informed about the activities of the corporation; (c) is required to monitor corporate affairs and policies; (d) is required to maintain familiarity with the financial status of the corporation, by regular reviews of financial statements; and (e) may need to inquire further into matters revealed by a review of financial statements. Id. at 821–23. For an example of an Australian decision, see Daniels v. Anderson, (1995) 118 F.L.R. 248, 309 (Clarke JA & Sheller JA); see also Australian Sec. & Investments Comm’n v. Loiterton, (2004) NSWSC 172.

221. One can only appreciate the value of the Nigerian provisions for liability for breaches of duty of care, and particularly section 279(7) of CAMA in light of the fact that section 102(b)(7) of the Delaware Code was enacted as a direct response to the decision of the Delaware Supreme Court in Smith v. Van Gorkom, 488 A.2d 858 (Del. 1985), which held directors liability for breach of duty of care. See also Mark A. Sargent, Two Cheers for the Maryland Director and Officer Liability Statute, 18 U. BALTIMORE L. REV. 278 (1989) (discussing the genesis of the amendment made to the Delaware Code in the area of director liability).

222. See Companies and Allied Matters Act (CAMA), (1990) Cap. C20, § 506(1).
in governance, rather than turn a blind eye to the running of the affairs of the company as business managers. Moreover, the provision permits those empowered under it to seek remedy against directors where doing so would have been very difficult while the company is a going concern.

H. Minority Empowerment

1. Derivative Action

Derivative action is one minority empowerment device adopted under Nigerian law, apparently to reduce agency costs. To be sure, the remedy against prejudicial or oppressive actions, which the legislature empowers the shareholders to seek, is another form of minority empowerment, albeit, one which protects the personal interests of the complaining shareholders rather than that of the company. The focus in this section is on the derivative action remedy.

A derivative action instituted by minority shareholders does not benefit only the shareholders that bear the cost of the litigation. It is an action intended to benefit the company as a whole by remedying a wrong committed against the company either by management or when management is reluctant to act on behalf of the company. Given this fact, the greater the difficulty in instituting a derivative action, the less likely minority shareholders will have recourse to remedy a wrong against the company. While minority shareholders will indirectly benefit from a successfully prosecuted derivative action, there is no doubt that conceptually the action may help reduce agency costs.

Several elements of the CAMA provisions dealing with derivative action deserve some mention. First, a range of applicants may submit to the court for leave to commence the action. Second, the fact that an alleged breach of a right or duty owed to the company has been or may be approved by the shareholders of the company is not sufficient for a stay or dismissal by the court of an action already commenced. While the court may take such facts into consideration, the Act empowers the court

---

223. Id. §§ 300–30.
224. Under section 309 of CAMA, “applicant” means:

(a) a registered shareholder or a beneficial owner and a former registered owner or beneficial owner of a security of a company; (b) a director or an officer or a former director or officer of a company; (c) the Commission; or (d) any other person who in the discretion of the court, is a proper person to make an application under section 303 of the Act.

Companies and Allied Matters Act (CAMA), (1990) Cap. C20, § 309.
225. See also id. § 305.
to exercise its discretion and look into the justice or equity of the matter and decide what is right in the circumstances.\textsuperscript{226} Because the focus of a derivative action is the protection of the company’s right, it is important that majority shareholders are not allowed to determine what sort of right the company may enforce, and by so doing, compromise the company’s interest, particularly when to do so may benefit them indirectly. Third, the Act gives courts the discretion to order that an applicant be reimbursed for the cost of the action.\textsuperscript{227} Because no guidance is provided, it is for the court to look into the merit of the applicant’s effort in deciding whether or not to order reimbursement.

The provisions dealing with derivative action under the Act, however, suffer from some fundamental defects, which may undermine sincere efforts to reduce agency costs.\textsuperscript{228} In particular, the conditions for bringing a derivative action implicitly attenuate the value associated with the action as a device for reducing agency costs. No action can be brought unless the applicant can show to the satisfaction of the court that the wrongdoers are the directors who are in control and that the directors will not take necessary action.\textsuperscript{229}

Limiting the wrongdoers to the directors who are in control of the company reduces the scope of the action. In most cases, even when directors are not the wrongdoers, they may collaborate with the wrongdoers and refuse to bring an action. Thus, derivative action under CAMA may not cover cases of wrongs done to the company by persons other than the directors in control of the company. Even in the situation covered by the Act, minority shareholders are bound to confront information asymmetry. Derivative action designed to correct wrongs done to the company by directors in control is information intensive. Except where minority shareholders have representation on the board of directors, they

\textsuperscript{226} See id. § 304(2).
\textsuperscript{227} See id. § 304(2)(d). Moreover, the Act provides that an applicant seeking to enforce a company’s right through a derivative action shall not be required to give security for costs in any application made or action brought or intervened in under section 303 of the Act. See id. § 307. The court may equally order the company to pay an applicant interim costs before the final disposition of the case. This is apparently to reduce the financial impact on an applicant of a prolong litigation. See id. § 308.
\textsuperscript{228} See Brian R. Cheffins, Reforming the Derivative Action: The Canadian Experience and British Prospects, 1 CO. FIN. & INSOLV. L. 227 (1997).
\textsuperscript{229} The CAMA position can be contrasted with that under section 239(2) of the Canada Business Corporations Act, where no such condition as contained in section 303(a) of CAMA exists. The Nigerian approach is, it is submitted, overly restrictive. Note that except in relation to section 303(2)(a) of CAMA, the provisions of CBCA and CAMA on derivative action are \textit{in pari material}. 
may not have access to enough information with which to convince the court that a wrong has been committed. Yet, the Act does not make allowance for discovery, which would have facilitated the disclosure of information that may be crucial to the action. The good news, however, is that present or past directors or the Commission, who have better access to information about boards’ operations and activities are equally empowered to seek the court’s permission to commence derivative actions. Moreover, it is hoped that putative petitioners may take advantage of the rules of court on discovery to overcome information asymmetry.

2. Administrative Investigation

Investigation of company affairs by an administrative agency, in this case, the Corporate Affairs Commission (CAC),\(^{230}\) is one potent weapon that may significantly lessen managerial profusion and address part of the agency problems. The power given to the CAC to appoint competent persons to investigate a company may work well to stymie the acuteness of information asymmetry.\(^ {231}\) The knowledge by corporate managers that investigation may be conducted on the direction of the CAC into how they have run the affairs and managed the business of the company may also constitute a check on their conduct. In light of the fact of ownership structure of most Nigerian companies, the investigation device will go a long way in helping to address not just the concerns of minority shareholders, but also help change how corporate managers see their responsibilities.

Under CAMA, the CAC may, on the application of shareholders holding one-quarter of the class of shares issued, appoint inspectors to investigate the affairs of a company and report on them as the CAC may direct.\(^ {232}\) In this case, the applicants will be required to provide such evidence to the CAC to prove that they have good reasons for requiring investigation.\(^ {233}\) Obviously, the requirement of evidence to support a request to investigate is designed to give some discretion to the CAC in the matter and to discourage frivolous applications. Moreover, a court may direct that the affairs of a company be investigated, in which case the CAC has no discretion and must conduct an investigation.\(^ {234}\)

\(^{230}\) The Commission is the body responsible for the administration of the CAMA and its functions are more particularly set out in section 7 of the Act. See C.O. Okonkwo, The Corporate Affairs Committee, in ESSAYS ON COMPANY LAW 14 (E.O. Akanki, ed., 1992).

\(^{231}\) See Companies and Allied Matters Act (CAMA), (1990) Cap. C20, § 314.

\(^{232}\) See id. §§ 314(1), (2)(a).

\(^{233}\) See id. § 314(3).

\(^{234}\) See id. § 315(1). Note that sections 304 and 312 of the CAMA empowers a court to make such orders as it may think fit in the proceedings instituted under section 303 or
may also appoint inspectors to investigate the affairs of a company under the circumstances listed in section 315(2) of CAMA. The power of inspectors appointed under sections 314 and 315 of the Act to examine books, records, and accounts maintained by directors of the company under investigation, or to examine officers (past or present) or agents of the company is a considerable weapon, which, if properly deployed, may prove very useful.

I. Information Disclosure and Accountability

Under the Investment and Securities Act, public companies seeking to raise capital from the securities market are obligated to make very extensive disclosure of the companies’ activities and prospects. Indeed, the long list of items on which information is required in the prospectus is bound to significantly reduce information asymmetry and help investors evaluate the company. Because management or the board of directors faces both criminal and civil liabilities should they falsify or misrepresent material information or fail to disclose it in the prospectus, it is arguable that the monitoring effect of disclosure will produce its desired effect on management. Moreover, the new proxy regime, which requires management to disclose information on specific transactions, is equally supportive of the overall effort to make corporate managers more accountable. However, the fact that ISA is yet to properly grapple with the importance of continuous disclosure of information of material changes as they occur may undermine the effectiveness of external control mechanisms, particularly the market for corporate control.

J. Strengthening the Board Structure Through Mandatory Audit Committee

Ordinarily, CAMA empowers a corporate board to perform its duties through committees. The audit committee is one such committee commonly established by public companies worldwide. Indeed, the role and composition of the audit committee has come under intense scrutiny following the collapse of Enron and other big corporations in the United States. Interestingly, CAMA mandates every public company to estab-

---

311 of the Act. The court may well conclude that investigation is necessary in the particular circumstance of the case before it.

235. See id. § 317.
236. See Investment and Securities Decree No. 45 (1999), § 50; Rule 225 of the Nigerian SEC Rules and Regulations (pursuant to ISA, 1999).
237. See Companies and Allied Matters Act (CAMA), (1990) Cap. C20, § 64(a).
lish an audit committee, and stipulates as to the committee’s membership, an approach this writer believes may prove beneficial in the overall drive to reduce agency costs. Specifically, section 359(4) of CAMA requires that the audit committee “shall consist of an equal number of directors and representatives of the shareholders of the company (subject to a maximum of six) . . . .” Indeed, one may be permitted to say that an audit committee is a committee of the company, rather than that of the board. Unlike section 64(a) of CAMA, which permits a board of directors to exercise its powers through committees consisting of such members of the board as the latter consider necessary, section 359(4) mandates a public company, rather than the board, to establish an audit committee.

That shareholders have direct representation on audit committees suggests that directors do not dominate audit committees in Nigeria. Not being members of the board, the shareholder representatives on an audit committee are expected, theoretically, to provide the necessary balance in the way the committee is expected to perform its functions. The Nigerian approach contrasts sharply with that in the United States and most common law countries where the concern is mainly with the independence of director-members of audit committees; it is somewhat anomalous in those countries for non-directors to serve on audit committees, because an audit committee is seen as a committee of the board, rather than that of the company.

It is, however, noteworthy that the Nigerian approach with regards to membership of the audit committee may pose some problems in practical terms. While contemporary corporate governance reforms of audit committees focus on the quality and independence of members, CAMA does not address who the shareholders may appoint as members of an audit committee. But it is becoming increasingly crucial for the members of an audit committee to be knowledgeable in financial matters, so they can make meaningful contributions in the course of the committee’s deliberations. Moreover, one can only hope that the shareholder representatives on audit committees will not compromise their independence, especially as they may see the opportunity given to them to serve as a rare privilege to join the directors in sharing the spoil of office, rather than an opportunity to add value to the corporation on whose committee they serve.

board’s relationship with external auditors). For Nigeria, see Code of Corporate Governance for Banks in Nigeria Post Consolidation (2006), §§ 3.12, 4.15, 4.16, 5.3.12, 7.1.4; NYSE REPORT, supra note 3.

239. See Companies and Allied Matters Act (CAMA), (1990) Cap. C20, § 359(3)–(4).
K. The Code of Corporate Governance Practices for Banks Post Consolidation

In Nigeria, the separation of ownership and control may be realized much faster in the financial (especially banking) sector than in other sectors. The Code of Corporate Governance Practices issued by the CBN deserves mention here, especially because, unlike most corporate governance codes, the CBN code is mandatory. Some of its provisions are definitely designed to reduce agency costs. The fact that banks must comply with it and include the Code’s compliance status report in the audited financial statements underscores the banks’ responsibility to actively take steps that policymakers believe could prove useful in dealing with agency costs.

It is evident from the preamble to the Code that one of the CBN’s main objectives is to provide an additional framework for effective governance that could minimize agency costs. The CBN noted, among other things, that “poor corporate governance was identified as one of the major factors in virtually all known instances of a financial institution’s distress in the country.” Other specific corporate governance concerns which the Code identifies as requiring attention include fraudulent and self-serving practices among members of the board, management, and staff; over-bearing influence of the chairman or managing director/chief executive officer, especially in family-controlled banks; weak internal controls; abuses in lending, including lending in excess of single obligor limit; and incompetent directorship.

With the foregoing backdrop, one is able to relate better to some of the recommendations in the Code, which the CBN made mandatory for banks. The Code now requires a separation of the office of the chief executive officer and that of chairman of the board of directors. It further provides that no two members of the same extended family should occupy the position of chairman and that of the chief executive officer or executive director of a bank. It requires that a committee of non-executive directors should determine the remuneration of executive directors, and that the non-executive directors’ remuneration should be limited to sitting allowances, directors’ fees, and reimbursable travel and

241. Id. § 6.1.15.
242. See id. § 1.3.
243. Id. §§ 2.3, 2.4, 2.5, 2.10, 2.11.
244. Id. §§ 5.2.1, 5.2.3.
hotel expenses. More importantly, the Code requires banks to establish “whistle-blowing” procedure that encourages (including assurance of confidentiality) all stakeholders to report any unethical activity or breach of the corporate governance code using, among others, a special email or hotline to both the bank and the CBN.

V. CONCLUSION

It must be noted that the use of agency costs as a tool of analysis within the corporate law and governance systems is still a novel subject in Nigeria; in part because efforts by scholars to explore the various interconnections between law, economic, and other aspects of social sciences to provide theoretical frameworks assessing the effectiveness of the law as an institution as well as the understanding of motivations of those entrusted to implement the law, is yet to predominate our scholarship. This Article is a modest effort to articulate: (a) the inevitability of agency costs and their persistence within the Nigerian corporate law and governance systems; and (b) various theoretical approaches that should inform the initiatives to deal with them.

Having noted that Nigerian companies combine both concentrated and dispersed ownership structures, whether by choice or policy design, the Article addresses the peculiarity of the agency problems associable with those structures. On balance, CAMA’s response to agency problems and costs has been impressive. As this Article has discussed, however, there is significant room for improvement in Nigerian corporate governance.

---

246. Id. § 6.1.12.
JAPAN'S GOVERNMENT PROCUREMENT REGIMES FOR PUBLIC WORKS: A COMPARATIVE INTRODUCTION

Shigeki Kusunoki*

ABSTRACT

In Japan, the Act for Promoting Quality Assurance in Public Works was enacted in spring 2005 and came into force soon thereafter. This is an epoch-making act in the history of Japanese public procurement regimes and practices, mainly in that, with respect to bidding procedures: (1) it declares that a comprehensive evaluation method shall generally be used for public works; (2) it permits procuring entities to dialogue with candidates to discuss improvement of their submitted technical proposals; and (3) it permits a cap on the estimated price to be set just before the scheduled date of bidding. The author will introduce Japan’s basic public procurement regimes and the Act’s impact on them, and describe Japan vis-à-vis the United States, European Union, and the Government Procurement Agreement of the World Trade Organization (WTO-GPA or GPA) on the issue of contractor award processes. The hurdles for implementing the WTO-GPA that Japan has to overcome will be addressed in the concluding remarks.

I. INTRODUCTION

As the result of the negotiations of the Uruguay Round for the General Agreement on Tariffs and Trade (GATT), as well as simultaneous multilateral trade negotiations, the Government Procurement Agreement of the World Trade Organization (WTO)1 was signed in Marrakech, Morocco in 1994 and came into effect in January 1996.2 In December 1995, just before it came into force, Japan ratified and then promulgated the Agreement.3

---

* Associate Professor, Faculty of Law, Sophia University. The author thanks the Japan Society for the Promotion of Science for its financial support.
2. GPA, supra note 1, art. 24(1); ARROWSMITH, supra note 1, at 40.
Ten years have passed since the GPA entered into force. Until today, the member countries, as well as observers like China, have reformed their domestic government procurement regimes in an attempt to implement the GPA. Among them, Japan has grappled with many reforms and measures, including ones to tighten up the conditions for awards in cases of non-competitive tendering (single-source tendering) and to establish bid-protest institutions.

Remarkably, as in other member and observer countries, not only the need to implement the GPA to fulfill international treaty obligations, but also particular domestic considerations, provided the impetus for Japan’s recent legal and practical reforms with respect to public procurement. Specifically, the reforms in Japan needed to address not only the country’s GPA obligations but also the country’s specific institutional, historical, cultural, and social conditions, such as the group mentality deeply rooted in Japanese culture which has resulted in bid-rigging problems as well as the involvement of government officials in the majority of bid-rigging cases.

As a starting point to address these obligations and specific conditions in Japan, the Act for Promoting Quality Assurance in Public Works (APQA) was enacted in spring 2005. Although the immediate effect of this important Act is limited as it covers only public works, it has the strong potential to be a breakthrough for the necessary reform of Japanese public procurement laws, regulations, and practices in general. Among its most notable provisions, the Act provides that the

---


6. Id. at 643–56.

7. See, e.g., Guy de Jonquieres, Take a Bite Out of Japan’s Collusion Culture, FIN. TIMES, May 31, 2005, at 17.

8. Regarding recent cases government officials involved in bid-rigging cases, see, e.g., Bid-rigging Attitudes See Shift, YOMIURI SHIMBUN [DAILY YOMIURI] (Tokyo), Dec. 10, 2006, at 3.

“comprehensive evaluation” method, and not the “price-only competition” method, is to be generally employed;\textsuperscript{10} procuring entities and candidates may discuss and make adjustments to the technical proposals in competitive bidding procedures;\textsuperscript{11} and the “maximum estimated price” usually set at the time of a bidding announcement may be set by procuring entities just before the date of bidding.\textsuperscript{12} Until the end of the last century, the main aims of reforming public works regimes were to enhance competitiveness and to deter misconduct, responding to criticism against the frequent cases of bid rigging and bribery in the 1990s. After the enactment of the APQA, however, the movement for Japan’s procurement reforms could be boldly steered toward “flexibility” and “diversification.”\textsuperscript{13}

This article aims at providing detailed information concerning the developments in Japan described above. It is hoped that this article will be useful to readers in deepening their understanding of Japan, and thereby contribute to smooth negotiations among WTO members. Moreover, it is hoped that the article will also be of use to academics and legal practitioners as a comparative analysis of laws and policies.

Hereafter, Part II surveys and highlights the basic legal schemes for government procurement in Japan. Part III introduces the APQA and its legislative history. Next, Part IV compares the relevant laws in Japan, the United States, and the European Union, and discusses the anticipated hurdles that Japan must overcome to implement the GPA fully. Finally, Part V contains the author’s conclusion.

II. CONTRACTOR SELECTION PROCESSES IN JAPAN’S PUBLIC PROCUREMENT

Generally speaking, there are two types of administrative agencies that handle government procurement, namely national agencies and local agencies. In Japan, different laws and regulations apply to the two types of administrative agencies. National agencies are regulated by the Accounting Act (AA)\textsuperscript{14} and its implementing Order Concerning

\begin{itemize}
  \item \textsuperscript{10} \textit{Id.} art. 3; \textit{see infra} Part III.B.3.
  \item \textsuperscript{11} APQA, \textit{supra} note 9, art. 13; \textit{see infra} Part III.B.6.
  \item \textsuperscript{12} APQA, \textit{supra} note 9, art. 14; \textit{see infra} Part III.B.7.
  \item \textsuperscript{13} It is difficult to understand the meaning and impact of this legislation without knowledge of the Japanese laws and regulations concerning public accounts, including the basic provisions for public procurement procedures. These laws and regulations will be explained in Part II of this article.
  \item \textsuperscript{14} \textit{Kaikeih\=o} [Accounting Act], Law No. 35 of 1947 [hereinafter AA].
\end{itemize}
Budget, Auditing and Accounting (OBAA), and local agencies are regulated by the Local Autonomy Act and its implementing Order Concerning Enforcement of the Local Autonomy Act. However, these two sets of acts and orders are similar with respect to their basic framework for contractor selection processes. Therefore, this article will hereafter highlight only the main points of the legal provisions regulating national agencies and will mention the local agencies’ regulations only when there is no equivalent for national agencies.

A. Contractor Selection Procedures

First, there is a set of rules concerning whether an award is to be made to a contractor through a competitive selection procedure, and if so, how competitive the process will be. This is the famous classification distinguishing “open competitive tendering,” “designated competitive tendering,” and “non-competitive (single-source) tendering.”

Open competitive tendering is a procedure in which any qualified person may submit a tender. This is the general procedure provided in the relevant acts and orders, regardless whether national or local. However, this procedure does not mean that anyone may unconditionally take part in the bidding process. Laws and regulations oblige procuring entities to exclude from such process persons falling into certain categories and allow procuring entities to do so in certain cases. Moreover, chief executives of procuring entities may from the outset impose conditions of qualification for participation in particular bidding procedures.

Designated competitive tendering and non-competitive tendering may be used only if the conditions prescribed by the relevant laws and

15. Yosan kessan oyobi kaikeirei [Order Concerning Budget, Auditing and Accounting], Imperial Edict No. 165 of 1947 [hereinafter OBAA].
18. AA, supra note 14, art. 29(3).
19. See, e.g., id. art. 29(3)(2); OBAA, supra note 15, art. 70 (providing, inter alia, for the exclusion of a person who is incompetent to contract and a bankrupt without any possibility of reinstatement).
20. See, e.g., AA, supra note 14, art. 29(3)(2); OBAA, supra note 15, art. 71 (providing, inter alia, that procuring entities may bar persons who are found to have violated the Anti-Monopoly Act from competitive bidding for a maximum period of two years).
21. OBAA, supra note 15, art. 72(1). These rules apply to designated competitive tendering. Id. art. 95. Open competitive tendering with these restrictions is called “conditional open competitive tendering.”
Designated competitive tendering is a procedure in which only persons invited (i.e., designated) by the procuring entities may submit a tender. Designated competitive tendering may be used only in the following cases: (1) when, due to the nature or purpose of the contract, only a small number of operators are expected to participate in competitive bidding and the use of open competitive tendering is not necessary; (2) when, due to the nature or purpose of the contract, the use of open competitive tendering is deemed to be disadvantageous; (3) when the estimated contract price is lower than the threshold level; or (4) in other cases where the order so provides.

Non-competitive tendering is a procedure other than the two competitive tendering procedures previously described. Typically, procuring entities using this procedure contact targeted operators individually, negotiate with them to calculate the estimated expenses to be submitted to the agency, and enter into a contract.

Further, the relevant laws and regulations require procuring entities to select contractors as competitively as possible even in cases where the entities use designated competitive tendering or non-competitive tendering. That is, in designated competitive tendering cases, procuring entities are required to meet certain regulations.
entities should designate at least ten operators if possible, and in non-competitive tendering cases, procuring entities should collect written estimates from at least two operators if possible.

B. The “Price-Only Competition” Method and the “Comprehensive Evaluation” Method

The issue of which type of contractor selection procedure a procuring entity should use, discussed in the previous section, can be said to be an issue concerning whether the selection process is “competitive” or not and how competitive it is, whereas the distinction between the price-only competition method and the comprehensive evaluation method relates to what candidates compete over, that is, the “factors of competition.” First, the price-only competition method can generally be described as a method in which the bidder offering the lowest price will receive the award. Under the law, this method should be used in principle if a procuring entity uses a competitive tendering procedure, regardless whether it is “open” or “designated.”

In contrast, the comprehensive evaluation method can be described roughly as a method where not only price but also other conditions such as quality, techniques, or skills are to be considered in the aggregate and the most advantageous candidate will be awarded the contract when a procuring entity selects a contractor. Under the law, this method should be used exceptionally, that is, only when certain conditions are met.

As will be mentioned later, the latter method is not considered exceptional in the U.S. or the E.U., whereas it is so considered in Japan. This exceptional treatment is emphatically reflected by the provision stating that “[m]inisters or chiefs of departments or agencies should consult with the Minister of the Department of Finance” when they wish to use this method.

---

30. Id. art. 97(1).
31. Id. art. 99(6).
32. See AA, supra note 14, art. 29(6)(1).
33. Id.
34. Id. art. 29(6)(2).
35. OBAA, supra note 15, art. 91(2). This “consultation” requirement was originally interpreted as requiring individual consultation for each procurement item, so the comprehensive evaluation method was seldom used. Later, the competent ministries and the Minister of the Department of Finance agreed that this requirement should be interpreted as not demanding individual consultation as long as there was a comprehensive agreement through comprehensive consultation and they applied this revised interpretation to public works. Thereafter, the “Guidelines for the Comprehensive Evaluation Method in Public Works” were released in 2000. Now, as long as the case falls within the purview of these guidelines, there is no need for consultation. This fact
C. Maximum Estimated Price

A specific characteristic of Japan’s scheme for selection of public works contractors in comparison with other countries is the existence of an upper-limit “estimated price” over which a bid is deemed invalid and a government procuring entity may not contract. This is very strict and there is no exception.\(^{36}\) This limitation is commonly applied regardless of the type of contractor selection procedure.\(^{37}\)

Under the law, a procuring entity shall estimate the bidding price or the contract price of the item at issue in accordance with its design and other specifications and other relevant characteristics,\(^{38}\) and the estimated price shall be properly calculated through consideration of such relevant factors as actual market prices, the situation of supply and demand, the ease or difficulty of execution, the quantity, and the length of time necessary for execution.\(^{39}\) This estimated price is generally characterized as “just an estimated cost calculated in advance.” This characterization in itself is not the basis for the maximum nature of the estimated price. It could be based largely on the government’s own ideas regarding the governmental budget and disbursement.

While the maximum estimated price eliminates prices that are too high, there are also schemes that forestall prices that are too low. The law allows a procuring entity to exclude a candidate whose bidding price is deemed to be so low that the contract might not be executed properly or that the contract would be considered improper because it is likely to disturb the order of fair trade.\(^{40}\) In such a case, a procuring entity shall estimate

\[^{36}\text{In the case of an auction, a minimum price limit will be set. AA, supra note 14, art. 29-6(1). As this issue is outside the scope of this article, a precise explanation is omitted.}\]

\[^{37}\text{Id. art. 26(6) (prescribing the estimated price for open competitive tendering and designated competitive tendering); OBAA, supra note 15, art. 99-5 (prescribing the estimated price for non-competitive tendering).}\]

\[^{38}\text{OBAA, supra note 15, art. 79.}\]

\[^{39}\text{Id. art. 80(2).}\]

\[^{40}\text{AA, supra note 14, art. 29(6)(1).}\]
entity has an obligation to investigate the relevant facts. Moreover, as far as local governments are concerned, a minimum price limitation may be set.

D. Secondary Policies

The Act on Ensuring the Receipt of Orders from the Government and Other Public Agencies by Small and Medium-Sized Enterprises (Public Agency Order Act or PAOA) was enacted in 1966 to implement the aims of the Basic Act Concerning Small and Medium-Sized Enterprises. As the full name of the Act itself reflects, the Public Agency Order Act aims to enhance development of small and medium-sized enterprises by ensuring procurement by public agencies from them through measures to expand opportunities for such enterprises to receive orders. This Act, composed of only seven articles, has long influenced Japanese public procurement practices.

The Public Agency Order Act provides that the government and other public agencies “shall make efforts to expand opportunities for the receipt of orders by small and medium-sized enterprises, while giving due consideration to the proper execution of the budget” when public contracts are concluded. The Act further provides that, as means to carry out these obligations, each year (1) the Minister of Economy, Trade and Industry shall, through consultation with the ministers and heads of other competent ministries and agencies, draft guidelines to expand opportunities for small and medium-sized enterprises to receive orders, taking into consideration the budget and project plans for the year; (2) the Minister of Economy, Trade and Industry shall request approval by the Cabinet of the guidelines; (3) the heads of each of the relevant ministries and agencies shall report their

41. OBAA, supra note 15, art. 86(1).
42. Order Concerning Enforcement of the Local Autonomy Act, supra note 17, art. 167-10(2).
43. Kankōju ni tsuiteno chūshō kigyōsha no juchū no kakusho ni kansuru hōritsu [Act on Ensuring the Receipt of Orders from the Government and Other Public Agencies by Small and Medium-Sized Enterprises], Law No. 97 of 1966 [hereinafter Public Agency Order Act (PAOA)].
44. Chūshō kigyō kihonhō [Basic Act Concerning Small and Medium-Sized Enterprises], Law No. 154 of 1963, art. 20 (providing that the government shall take necessary measures to ensure proper trade by small and medium-sized enterprises).
45. PAOA, supra note 43, art. 1.
46. Id. art. 3. Article 2 of this Act defines “small and medium enterprises.” Id. art. 2.
47. Id. art. 4(1)–(2).
48. Id. art. 4(2).
achievements of contracts with small and medium-sized enterprises to the Minister of Economy, Trade and Industry after the end of the fiscal or business year; and (4) local self-governing bodies, applying national policies, shall endeavor to take actions to ensure opportunities for the receipt of orders by small and medium-sized enterprises.

In practice, government procuring entities have tried either to enhance small and medium-sized enterprises’ entrance into the public procurement market by dividing works into smaller categories or dividing orders into smaller lots in order to make the contract price low enough, or to expand the receipt of orders by small and medium-sized enterprises through the use of designated competitive tendering or non-competitive tendering. In national procurement, small and medium-sized enterprises have received roughly forty to fifty percent of the total amount of governmental contracts during the past decade.

E. Practices to Adjust Opportunities Through “Designation:” Advantages and Disadvantages

The Japanese Act Concerning Prohibition of Private Monopolization and Maintenance of Fair Trade—the equivalent of antitrust law in the United States or competition law in the European Union—was not actively enforced for a long time between its enactment in 1947 and the end of the 1980s. Many regulations limited newcomers’ entry into many business fields such as transportation, electricity, and the media. This demonstrates that competition was not a basis of Japanese economic management.

49. Id. art. 5.
50. Id. art. 7.
51. See, e.g., Jichitai ha kansei dango no ne wo tachikire [Local Governments Must Root Out Bid-Rigging with the Involvement of Government Officials], NIHON KEIZAI SHIMBUN, Dec. 3, 2006 (Morning Ed.), at 2.
52. See id.
54. Shiteki dokusen no kinshi oyobi kosei torihiki no kakuho ni kansuru hōritsu [Japanese Act Concerning Prohibition of Private Monopolization and Maintenance of Fair Trade], Law No. 54 of 1947.
55. Some commentators point out that Japanese tend to prefer “cooperation” rather than “competition” based on their cultural background. See, e.g., THE JAPANESE MIND: UNDERSTANDING CONTEMPORARY JAPANESE CULTURE 195–97 (Roger J. Davies & Osamu Ikeno eds., 2002); YOSHIRO SUGIMOTO, AN INTRODUCTION TO JAPANESE SOCIETY (CONTEMPORARY JAPANESE SOCIETY) (2d ed. 2003).
Japanese public procurement was not managed competitively either. As already stated in Section A of this Part, open competitive tendering is now adopted as a general rule. Until a decade ago, however, the majority of contractor selection procedures used designated competitive tendering. Generally speaking, three reasons can be cited for this. First, it is advantageous if procuring entities can exclude improper candidates from contractor selection procedures through their designation. Second, designated competitive tendering is an effective way to implement the Public Agency Order Act. Third, there are paternalistic demands on the government to adopt designated competitive tendering to assure long-term stable profits for a limited number of companies which are expected to supply high-quality items or works.

As to the third reason, assurance of companies’ long-term stable profits can be achieved not only through the use of designated competitive tendering, but also through a system of adjustment or distribution of profits, that is, bid rigging. The adoption of designated competitive tendering encourages candidates to engage in bid rigging because the number of candidates is limited through the designation, the list of designated candidates can frequently be fixed, and candidates can collude more easily. Moreover, in many cases, procuring entities tacitly permit and sometimes actively get involved in bid rigging. Many commentators informally contend that competition is restrained by bid rigging in most procurement cases in Japan. Naturally, this anti-competitive characteristic of public procurement has supported a cozy relationship among politicians, bureaucrats, and businesspersons for a long time.

57. See, e.g., Yasushi Ohno, Kokyokōji ni okeru nyusatsu keiyaku hoshiki nokadai [Reconsidering Bidding and Contracting Methods of Public Works], 27 KAIKEI KENSA KENKYU 159, 162.
58. See, e.g., Kusunoki, supra note 35, at 17.
59. See Ohno, supra note 57, at 161; Kusunoki, supra note 35, at 14–15. This idea is very controversial. Some commentators may insist that the more intense the competition, the higher the quality of the items and works the successful candidates will supply.
61. Id. at 39–41.
62. Id. at 27.
63. Id. at 39–41.
Criticisms of this problem have grown louder since a number of infamous bid-rigging and bribery cases, including one involving then Construction Minister Kishiro Nakamura and Kajima Corporation, the largest construction company in Japan, triggered the public’s anger in the 1990s. Designated competitive tendering then became a target of criticism as a hotbed of unjust activities. Almost simultaneously, the sanction regime and the enforcement of the Anti-Monopoly Act were strengthened. Furthermore, most national and some local administrative agencies started to treat open competitive tendering as the general rule for procurement to implement the WTO-GPA after it came into force in 1996.

The price-only competition method was, however, left as the general rule. Public works is one of the fields most disadvantaged by this. Coupled with the long-term recession and the drastic curtailment of the public project budgets throughout the 1990s, the change to open competitive bidding forced civil engineering and construction companies to compete much harder than before and this increased competition led to dumping. As a result, there have been many cases where the contract price was much lower than the estimated price that had been deemed to be reasonable.

64. A. Didrick Castberg, Prosecutorial Independence in Japan, 16 UCLA PAC. BASIN L.J. 38, 83–84 (1997). Nakamura was alleged to have received a bribe of ten million yen from Kiyoyama Shinji, a vice president of Kajima, in exchange “for Nakamura’s convincing the [Japanese] Fair Trade Commission [JFTC] not to pursue collusion charges against a cartel of construction firms in Saitama prefecture.” Id. at 83; see also Woodall, supra note 60, at 126.


67. Until the end of the last century, however, the speed of the reforms was slow.

68. There has not been any change to AA article 29(6)(1) which provides that the price-only competition method is a general rule. AA, supra note 14, art. 29(6)(1).

69. The Junichiro Koizumi administration (2001–06) has been based on the concept of “a small government.” This concept has been the engine to budget curtailment. See, e.g., Haruki Sasamori, Shift to Small Govt Accelerates, YOMIURI SHIMBUN, Dec. 28, 2005, at 4.

III. NEW LEGISLATION ON PUBLIC WORKS: THE ACT FOR PROMOTING QUALITY ASSURANCE IN PUBLIC WORKS (APQA)

A. Background of the Legislation 71

In response to the cries of construction companies in the wake of intense competition, the Liberal Democratic Party (LDP), the ruling party for a long time after WWII, started to move to address their hardships. 72 In June 2003, the LDP launched an internal survey and a study group on the issue.

In the same month, the Japanese Fair Trade Commission (JFTC) launched “The Study Group Concerning Public Procurement and Competition Policy,” which aimed to suggest new schemes and practices to deter bid rigging and maximize “value for money.” 73 The recommendations of the study group’s report published in November 2003 include “the positive active use of the comprehensive evaluation method” and “the introduction of the competitive dialogue method,” both later incorporated into the provisions of the APQA.

The JFTC’s report published simultaneously suggested that amendment of the Anti-Monopoly Act was needed to strengthen the sanctions against various types of misconduct. 75 This added fuel to the LDP members’ sense of an impending crisis because they thought the stricter sanctions would lead to an increase in dumping by construction

71. Information about the background and legislative history of the APQA can generally be found only in specialized newspapers written in Japanese.

72. The construction industry is one of the LDP’s important money pipelines and one of its power bases. See Woodall, supra note 60, ch. 3.


companies. The LDP therefore rushed to enact new legislation on public works.

In autumn 2004, Diet members introduced a bill to enact the APQA at the same time the Cabinet submitted a bill to amend the Anti-Monopoly Act. The former was enacted three weeks after the latter in spring 2005.

B. Introduction to the Contents of the APQA

1. Purpose

Article 1 of the APQA states that its purpose is to “set[] forth a basic philosophy for ensuring the quality of public works in order to clarify the responsibilities of the central government and other stakeholders” and “basic policies to promote quality in public works in order to improve the public welfare and contribute to the sound development of the national economy.”

2. Definition of “Public Works”

Article 2 provides that the “public works” that are covered by the APQA are “as defined in Article 2.2 of the Act for Promoting Proper Tendering and Contracting for Public Works.”

3. Basic Philosophy

Article 3 of the APQA describes the basic philosophy of the Act as follows:

1. In that public works, providing social capital that supports the well-being and economic activities of the public, have important socioeconomic implications, the central and local governments, as well as other entities that place and receive orders for public works,
should ensure the quality of public works for present and future generations of the Japanese people in fulfilling their respective roles.

2. In that construction work has such unique characteristics as that its quality can be confirmed only after structures are provided for use, its quality depends to a great degree upon the technological capabilities of contractors, and its conditions differ significantly between individual projects, various factors in addition to price should be considered to ensure the quality of public works; due consideration should also be given to economic efficiency, resulting in the conclusion of contracts that comprehensively consider pricing and quality.

3. In that work efficiency, safety, environmental impact and other factors are important considerations in ensuring the quality of public works, quality assurance should employ the most appropriate technologies available.

4. To ensure the quality of public works, due attention should be given to ensuring the transparency of tendering and contracting processes and the content of contracts, the fairness of competition for contracts, the removal of construction companies that are not qualified as contractors, the elimination of improper activities such as collusion and bid-rigging, and the use of proper construction practices.

5. To ensure the quality of public works, due consideration should be given to the private companies employed in public works projects, including the proper evaluation of their capabilities; the proper reflection of these capabilities in tendering and contracting; and the use of their technical proposals (herein meaning proposals on technology utilization submitted for public works contracts to be awarded competitively), originality, and ingenuity.

6. To ensure the quality of public works, due attention should be given to the conclusion of fair contracts based on agreements between parties negotiating on an equal footing, and to the good-faith implementation of these contracts.

7. To ensure the quality of public works, the quality of surveys on and designs for public works shall be ensured in accordance with the principles set forth in the preceding paragraphs, in that the quality of such surveys and designs significantly affect the quality of public works.\(^{80}\)

\(^{80}\) APQA, supra note 9, art. 3.
4. Responsibilities

Articles 4, 5, 6, and 7 of the APQA provide for the responsibilities of the national and local governments, procuring entities, and contractors. Article 4 provides that the national government “shall assume responsibility for formulating and implementing comprehensive measures to ensure the quality of public works.” Article 5 commands the local governments to assume the same responsibility “in cooperation with the central government, while giving due consideration to local needs.” Article 6 provides that the procuring entities “shall properly perform . . . the production of written specifications, evaluation of prices, determination of tendering and contracting methods, selection of the contractor, supervision and inspection of work, and confirmation and evaluation of the progress of construction during the work period and at the time of completion.” Finally, Article 7 requires public works contractors to perform their work “pursuant to the basic philosophy, and shall improve their technological capabilities to that end.”

5. Basic Principles

Articles 8, 9, and 10 provide for basic principles and responsibilities pursuant thereto to be established by the national government. Article 8 directs the national government to set forth, and then give public notice of, “basic principles for the comprehensive implementation of measures to ensure the quality of public works,” giving “consideration to the autonomy of quasi-governmental agencies . . . and of local governments.” Article 9 prescribes the obligation of the heads of ministries and agencies, heads and representatives of quasi-governmental and independent administrative agencies, and local government heads to “implement necessary measures to promote the quality of public works in accordance with basic principles.” Finally, Article 10 provides that the national government “shall establish systematic cooperation between concerned administrative organizations” in formulating and implementing the basic principles.

81. Id. art. 4.
82. Id. art. 5.
83. Id. art. 6.
84. Id. art. 7.
85. APQA, supra note 9, art. 8.
86. Id. art. 9.
87. Id. art. 10.
6. Evaluation of the Technical Capabilities of Bidding Participants and Technical Proposals

Articles 11 through 13 are provisions concerning the evaluation of the technical capabilities of and technical proposals from the participants in competitive bidding for public works contracts. First, Article 11 requires procuring entities to review the capabilities of candidates, “including their experience in constructing public works, their past construction experience and the expertise of the[ir] engineers.” Next, Article 12 requires procuring entities generally to request that candidates submit technical proposals, and in cases where technical proposals are requested, the entities “shall give public notice of both the request and the method of evaluating proposals in advance;” “shall properly examine and evaluate” the proposals, “implement[ing] measures to ensure . . . neutrality and fairness,” including complaint resolution measures; and generally “shall subsequently make public the results of their evaluations.” Finally, Article 13 prescribes the dialogue method for improvement of technical proposals. Specifically, it states that a procuring entity may request contractors to improve submitted proposals or give them the opportunity to do so, and, in such a case, the procuring entity shall “provide an overview of the technical proposal improvement process.” This latter provision is meant to keep the process of technical proposal improvement transparent.

7. Ex Post Facto Determination of a Maximum Estimated Price

Article 14 of the APQA provides for the ex post facto determination of a maximum estimated price following evaluation of technical proposals. Specifically, in a case where a procuring entity “request[s] technical proposals that involve advanced technologies,” it may set a cap on the estimated price “based on the results of its evaluations of proposals.” Moreover, “[i]n examining technical proposals,” the procuring entity is obligated to “seek the opinions of knowledgeable persons who can offer fair judgments from a neutral position.”

8. Assistance from Other People and Organizations

Procuring entities need sufficient administrative skills and experience to handle bidding procedures in order to carry out public works
properly. In cases where the contents of works are highly specialized and complicated techniques are needed, however, many procuring entities, especially those of local governments, do not have sufficient skills. Article 15 of the APQA therefore provides that the procuring entities are obligated to entrust performance of their duties to other people or organizations that have sufficient skills and experience. 93

C. Significance and Impact of the New Act

1. Comprehensive Evaluation Not to Be an Exception in Public Works

Under the APQA, the comprehensive evaluation method will be used more often than before, in place of the price-only competition method, which has been used on a regular basis. As the quotation of Article 3(2) above indicates, this provision of the Act provides that various factors in addition to price should be considered to ensure the quality of public works. 94

It is said that the LDP inserted the vague expression “various factors in addition to price” into this article because it wanted to weaken pressure from the Democratic Party of Japan (DPJ), the largest opposition party, in order to speed up passage of the legislation. The dominant opinion of the members of the DPJ—a party which depends heavily on the support of local and small-sized enterprises—was that “quality” should be defined to include various factors such as contributions to local events, contributions to charities, and contributions to disaster measures. 95 These are all activities that local and small-sized enterprises actively pursue in their home areas. As a result, it is uncertain exactly how comprehensive the comprehensive evaluation method will actually turn out to be in practice.

2. Dialogue Method

Article 13 of the APQA, which, as explained above, provides for a method of dialogue between a candidate and a procuring entity to improve a submitted technical proposal, can be said to be the core of the Act. Under the Accounting Act, the Local Autonomy Act, and related

93. Id. art. 15(1)–(3).
94. Id. art. 3(2).
95. There are many criticisms of this vagueness. See, e.g., Tarou Sawaki, Hinkakuho eno aru huan [Uneasiness About the Act for Promoting Quality Assurance in Public Works], 54 (4) CE [CIVIL ENGINEERING] 33 (2005) (explaining the main text’s story between the LDP and the DPJ).
orders, the holding of a dialogue between the time of notice and that of bidding are neither permitted nor prohibited in competitive tendering procedures.\textsuperscript{96} Article 13, however, encourages and can be used as justification for the use of the dialogue method by clearly describing the process for dialogue between a candidate and a procuring entity to improve a submitted technical proposal prior to the bidding date.

One potential problem is arbitrariness on the part of procuring entities in cases where the dialogue method is used. In this regard, Article 13 makes it obligatory for procuring entities to make public an overview of the process for improvement of technical proposals. The APQA, however, does not provide any further details.

3. Estimated Price: Ex Post Facto Determination

The method of determining estimated prices was drastically changed by Article 14, which provides that a procuring entity “may cap estimates based on the results of its evaluations of proposals” in cases where it requests “technical proposals that involve advanced technologies.”\textsuperscript{97} In such cases, the time of determination is expected to be several days prior to the day of bidding.\textsuperscript{98} Generally, estimated prices are determined at the stage of the procurement notice. As to such prices, the Order Concerning Budget, Auditing and Accounting requires the entity to estimate price “on the basis of the specifications and the design documents” and related factors concerning the item, and to “keep the document in which this estimated price is written or recorded at the time when submitted bids are opened.”\textsuperscript{99} If there is a very large informational gap concerning the technical proposals between the procuring entity and candidates and if the need for dialogue to improve the submitted proposals is very strong, this means that the procuring entity lacks sufficient data to determine the estimated price and it cannot make such a determination in advance. In such a situation, an ex post facto determination of the estimated price is inevitable.

\textsuperscript{96} See AA, supra note 14, art. 29(1)–(12) (ch. 4); Local Autonomy Act, supra note 16, ch. 9, § 6; OBAA, supra note 15, arts. 68–102 (ch. 7); Order Concerning Enforcement of the Local Autonomy Act, supra note 17, ch. 5, § 6. There is no provision concerning “negotiation” in the competitive bidding procedures. Indeed, a few procuring entities had used the dialogue method before the APQA came into force.

\textsuperscript{97} APQA, supra note 9, art. 14.

\textsuperscript{98} In one previous case, the estimated price was determined four days before the day of bidding.

\textsuperscript{99} OBAA, supra note 15, art. 79.
4. Monitoring and Bid Protests

To assure the quality of public works, it seems appropriate to adopt a flexible system of procurement methods. The greater the number of various methods available, however, the more important the procedures for monitoring procuring entities become. In competitive tendering, where price is the sole evaluated item, there are no serious problems because the successful bidders are determined automatically based only on the prices that they submit. In competitive tendering where the comprehensive evaluation method is adopted—especially in cases where candidates submit technical proposals, they and the procuring entities hold discussions, and the estimated price is determined ex post facto—the need for monitoring procuring entities increases and the techniques for undertaking such monitoring are complicated. In Japan, third-party institutions like the Committee for the Oversight of Bidding, which consists of part-time members including lawyers, scholars, and journalists, tend to be emphasized.

Moreover, it is expected that the more varied and flexible the procuring methods are, the greater the number of unsuccessful candidates who will make complaints. Unsuccessful candidates are supposed to be the most effective monitoring parties because they are the most interested parties who are close to the procuring entities. Therefore, establishment of sufficient and effective bid-protest procedures is one of the most urgent tasks in government procurement regimes.

Neither a monitoring scheme nor a bid-protest scheme is addressed in the APQA. Only the supplementary resolutions for the bill by the respective committees on Land and Transport in the House of Representatives and the House of Councilors deal with these matters, providing in identical language that one of the objectives of the measures the government should implement to enforce the law is “[t]o properly reflect the opinions of third parties such as knowledgeable persons regarding the tendering and contracting process for public works, and to properly handle complaints from concerned parties, including the enactment of legislation where necessary.”

To implement the WTO-GPA, in 1995 Japan established the Office for the Government Procurement Challenge System (CHANS) to deal

100. As the author mentions later, there is the marked tendency in the Japanese community for people to think that “a dispute is best avoided.” Therefore, the author has reservations concerning the effectiveness of any bid-protest scheme that is established.

with bid protests involving national procuring entities. However, this system had seldom been used as of July 2006. This is probably due to the Japanese anti-competitive and cooperative mindset ingrained in the domestic construction and civil engineering industries. The examining committee in this system (the Government Procurement Review Board), like the Committee for the Oversight of Bidding, is a third-party institution consisting of part-time members. Moreover, as of the same date, few local governments had established bid-protest procedures.

D. Practical Responses

Generally speaking, as of July 2006, most government agencies had not yet sufficiently prepared for the enforcement of the APQA. In August 2005, the Cabinet endorsed the basic principles provided in Article 8 of the Act. The Ministry of Land, Infrastructure and Transport (MLIT) then published guidelines concerning the enforcement of the Act in September 2005. Regional branches of the MLIT, other national agencies, and local governments are currently considering or beginning to undertake measures to implement the Act.


104. Due to time and space limitations, this article does not address the subject of legal review because the author believes that it would be very complicated to organize and examine country and regional regulations and practices for a comparative analysis of the United States, the European Union, and Japan. The author of course recognizes the importance of this matter.

105. See, e.g., Interview by Kensetsu Kogyo Shimbun with Masashi Waki, Member of the Diet (Feb. 15, 2006), available at http://www.waki-m.jp/column/column060217.html.


IV. JAPAN VIS-À-VIS THE UNITED STATES, THE EUROPEAN UNION, AND THE WTO-GPA

This Part analyzes the Japanese laws in comparison with the corresponding U.S. and E.U. laws (in part E.U. member states’ laws) and points out the hurdles that Japan must overcome to implement the WTO-GPA fully.

A. Comparison with the United States and the European Union

1. Contractor Award Process

With regard to the contractor award process, a notable characteristic of the relevant Japanese laws is that the comprehensive evaluation


method has not been adopted as the general rule for either national or local government procurement. 111 Prior to the enactment of the APQA, this was true even for items where sophisticated techniques were needed in public works, such as in airport or dam construction. Procuring entities have generally dealt with problems concerning quality or techniques not by adoption of the comprehensive evaluation method but instead by use of designated competitive tendering. 112 Though the law provides that designated competitive tendering may be adopted exceptionally, in practice the reverse has continued for a long time. 113 As previously pointed out, one basis for the APQA’s enactment was the increasing difficulty in excluding incompetent contractors from public procurements through “designation” because designated competitive tendering was considered to be a vice as a result of the aforementioned Nakamura and other bid-rigging and bribery scandals that broke out in the 1990s. 114 Therefore, procuring entities became obligated to use open competitive tendering as the rule and this has led to cut-throat competition among candidates. 115

As a result of the enactment of the APQA, the basic contractor award procedure for public works in Japan changed from the combination of “designated competitive tendering with a price-only evaluation method” to “open competitive tendering with a comprehensive evaluation method.” As will be shown, this change has brought Japanese law more in line with U.S. law and practice. It must be remembered, however, that this change in Japanese procedure currently affects only public works, due to the limited applicability of the APQA.

Unlike in Japan, neither U.S. nor E.U. law differentiate between the procedure for government procurement of public works and that for other types of items or works. In the United States, the generally used “full and open competition” procedure, equivalent to the primary procedure of “open competitive tendering” in Japan, uses two types of evaluation methods, namely, the “sealed bidding” method, 116 in which the contract is awarded to the bidder with the lowest price, 117 and the

111. In the last few years, only the Ministry of Land, Infrastructure and Transport has tried to use the comprehensive evaluation method on a regular basis. As a whole, however, this method is exceptionally used. 112. About use of designated competitive tendering in order to maintain quality of public works, see supra Part II.E. 113. See id. 114. See id. 115. See supra Part III.A. 116. 48 C.F.R. pt. 14 (2006). 117. Under sealed bidding, contracts are awarded “to that responsible bidder whose bid, conforming to the invitation for bids, will be most advantageous to the Govern-
“competitive proposal” method, in which the candidate presenting the proposal “that represents the best value” is awarded the contract. Sealed bidding may be used only in cases where certain conditions are met. In the European Union, on the other hand, combinations of the “open procedure with a price-only evaluation method,” the “restrictive procedure with a price-only evaluation method,” the “open procedure with a comprehensive evaluation method,” or the “restrictive procedure with a comprehensive evaluation method” may be used on a case-by-case basis. The European Union’s “open procedure” and “restrictive procedure” are equivalent, respectively, to open competitive tendering and designated competitive tendering in Japan. Among the four methods, there is no designation as to which should be used as a general rule and which should be used as an exception.

Another important development in the Japanese public procurement regimes, discussed previously, is that the APQA permits the use of dialogue in both types of competitive tendering procedures. This may be understood as an approach similar to the U.S. regimes. Namely, under U.S. law, both the “full and open competition” and

---

120. 48 C.F.R. § 6.401(a) (2006). This subpart provides:

“Contracting officers shall solicit sealed bids” if the four following conditions are met:

1. Time permits the solicitation, submission, and evaluation of sealed bids,
2. The award will be made on the basis of price and other price-related factors,
3. It is not necessary to conduct discussions with the responding offerors about their bids, and
4. There is a reasonable expectation of receiving more than one sealed bid.

Id.

121. In practice, the member states might treat the open procedure as a general rule and the restrictive procedure as an exception or vice-versa.
122. European Parliament and Council Directive 2004/18/EC, 2004 O.J. (L 134) arts. 28, 53. However, this does not mean the laws and regulations of all member states that implement the Directive distinguish between the rule and the exception(s) on this point.
123. See APQA, supra note 9, art. 13.
“full and open competition after exclusion of sources” procedures allow the use of the competitive proposal method in which negotiation between the procuring entity and candidates is expected. In the European Union, neither the open procedure nor the restricted procedure inherently anticipate negotiation. Instead, the “competitive dialogue” method in which procuring entities negotiate with candidates in a competitive environment was introduced by a new directive issued in 2004.

To the extent that a method by which a procuring entity and a candidate can hold discussions as part of a competitive procedure exists in public procurement law, however, the United States, the European Union, and Japan currently concur.

2. Maximum Estimated Price and Reasonable Price

Among the United States, the European Union, and Japan, the concept of capping the estimate for a contract price in public procurement exists only in Japan. The Japanese public procurement regimes are therefore special in this respect. Although in a given case in the United States the procuring entity calculates a price it deems reasonable in light of the specifications of the targeted items and unit prices of mate-

124. It should be noted that in the United States, procuring agencies designate certain entities that are excluded from the award competition, not those that may participate in it. On the other hand, as mentioned in the text accompanying note 123, Japanese designated competitive bidding is similar to the restrictive procedure of the European Union.

125. “Negotiation” means “a procedure that includes the receipt of proposals from offerors, permits bargaining, and usually affords offerors an opportunity to revise their offers before award of a contract.” KEYES, supra note 108, at 269. The FAR provides that “[b]argaining includes persuasion, alteration of assumptions and positions, give-and-take, and may apply to price, schedule, technical requirements, type of contract, or other terms of a proposed contract.” 48 C.F.R. § 15.306(d) (2006).


a procedure in which any economic operator may request to participate and whereby the contracting authority conducts a dialogue with the candidates admitted to that procedure, with the aim of developing one or more suitable alternatives capable of meeting its requirements, and on the basis of which the candidates chosen are invited to tender.

Id. art. 1(11)(c).

128. See supra Part II.C.
ials, this price is the standard of suitability, not the upper limit.\footnote{129} If the price offered by the most advantageous candidate exceeds this reasonable price, the procuring entity then examines whether this price is still within the reasonable range.\footnote{130}

Since, as aforementioned, procuring entities may dialogue with candidates to enable them to improve their proposed techniques pursuant to the APQA, it does not make sense to set the estimated price at the time of the invitation notice. The Act permits the estimated price to be set after holding discussions,\footnote{131} thereby rendering the estimated price extremely flexible, and procuring entities may consider many complicated factors in addition to budgetary restrictions when determining a suitable estimated price.

3. Secondary Policies

Secondary policies (policies other than those based on economic reasonableness) are commonly considered in public procurement law and practice around the world. In the United States, policies of protection of small and medium-sized enterprises and creation of jobs are reflected in several public procurement laws which provide special budgets for certain projects,\footnote{132} using the “full and open competition after exclusion of . . . sources” procedure.\footnote{133} For example, the Small Business Act\footnote{134} establishes a program that authorizes the Small Business Administration to enter into contracts with other agencies and thereafter subcontract their performance to firms eligible for program

\footnote{130. \textit{See} 48 C.F.R. § 14.404-2(f) (2006) (“Any bid may be rejected if the contracting officer determines in writing that it is unreasonable as to price.”); \textit{see also} 48 C.F.R. § 14.408-2(a) (“The contracting officer shall determine . . . that the prices offered are reasonable before awarding the contract.”).}
\footnote{131. \textit{See} APQA, \textit{supra} note 9, art. 14.}
\footnote{132. 48 C.F.R. §§ 6.203–6.205 (2006).}
\footnote{133. 48 C.F.R. § 6.200 (2006).}
\footnote{134. 15 U.S.C. §§ 631–57f (2006).}
participation\textsuperscript{135} based on different considerations than the usual “most favorable offer” evaluation standard.\textsuperscript{136} The E.U. directive does not contain any provisions concerning secondary policies; instead, the matter is left up to the member states. Generally speaking, member states’ national laws tend to support small and medium-sized enterprises and entities which usually must compete with large enterprises and entities at a disadvantage. In Germany, the Act Against Restraints of Competition (\textit{Gesetz gegen Wettbewerbsbeschränkungen})\textsuperscript{137} provides that “[t]he interests of small and medium-sized undertakings shall primarily be taken into account in an appropriate manner by subdividing contracts into trade-specific and partial lots.”\textsuperscript{138} In France, the Code of Public Contracts (\textit{Le Code des Marchés Publics}) provides that “[i]n the event of identical prices or equivalent bids for a contract, a preferential right is granted to bids submitted by a workers’ production cooperative, an agricultural producers’ group, a craftsman, a craftsmen’s cooperative society or an artists’ cooperative society or an adapted company.”\textsuperscript{139} It appears to the author that concerns for secondary policies are not involved in the “most economically advantageous” evaluation standard.\textsuperscript{140}

In applying the guidelines drafted pursuant to the Public Agency Order Act, government procuring entities in Japan, as in the European Union, tend to carry out secondary policies through specially designed entity-specific programs. In addition, certain projects are designated for award to small and medium-sized enterprises only. This system is abstractly similar to the aforementioned U.S. “full and open competi-


\textsuperscript{136} It is much different at the E.U. member state level. In the famous \textit{Burma/Massachusetts} dispute, a secondary policy concern (political commitment) was included in the award criteria. See ARROWSMITH, \textit{supra} note 1, at 343.


\textsuperscript{138} GWB, \textit{supra} note 137, art. 97(3) (English translation).

\textsuperscript{139} C. MARCHÉS PUB., art. 54(I) (Fr.) (English translation, Legifrance (government) Web site, http://195.83.177.9/code/liste.phtml?lang=uk&c=28&r=1411 (last visited Jan. 28, 2007)).

tion after exclusion of sources.” As pointed out above, the frequency of
the use of the designated competitive tendering procedure has de-
creased significantly in recent times. Carrying out secondary policies
through designation has therefore become difficult. To counter this in
the future, the author believes that the language of the relevant laws
and regulations can be interpreted to include secondary policy con-
cerns as part of the comprehensive evaluation standard. Indeed, the
APQA provides that “various factors in addition to price should be
considered to ensure the quality of public works” \(^\text{141}\) and the interpreta-
tion of the vague term “various factors” has not yet been made clear.
The interpretation is likely to be influenced heavily by future practices.
It cannot be denied that the APQA could bypass secondary policies,
but one should remember that the Public Agency Order Act still exists.
If concerns of secondary policies are included in the factors to be taken
into account under the comprehensive evaluation method, Japan would
be much different from the United States and the European Union with
respect to this point.

4. Monitoring and Bid Protests

The more complicated the evaluation process to award contracts, the
stronger the need to monitor procuring entities. Furthermore, the need
to provide procedural protection of losing candidates to appeal, i.e., to
make a bid protest, is also stronger. Although the APQA opened the
door for flexibility and complexity in public works procurement, the
necessary complementary monitoring and bid-protest mechanisms
have not yet been sufficiently established. As mentioned previously,
the existing monitoring and bid-protest institutions in Japan consist of
part-time outsiders. \(^\text{142}\) Whether such institutions work well is doubtful.

Due to the varying national schemes within the European Union,
only a comparison between Japan and the United States is made
here. \(^\text{143}\) In the United States, competition advocates \(^\text{144}\) and officials of

\(^{141}\) See APQA, supra note 9, art. 3(2).

\(^{142}\) See supra Part III.C.4.

\(^{143}\) The E.U. directive does not make concrete suggestions concerning the moni-
toring institutions or bid-protest procedures in the member states. The member states
must decide these matters on an individual basis. It is therefore difficult to analyze the
E.U. regimes concerning these matters comparatively with Japan and the United States
because of the inconsistency among the E.U. member states. It should be noted,
however, that there are several interesting points in the E.U. member states concerning
monitoring institutions and bid-protest procedures. For example, in Germany, the
specially-established office for bid protests is part of the Federal Cartel Office
(Bundeskartellamt), equivalent to the Federal Trade Commission and the Antitrust
Division of the Department of Justice in the United States, the Competition Director-
the Offices of the Inspector General of the various government agencies supervise government agency procurement activities.\textsuperscript{145} Notably, the Office of the Inspector General is a body highly independent from any other government agency, and controlled directly by the President.\textsuperscript{146} The officials of the Offices of the Inspector General and competition advocates are the inside government officials.\textsuperscript{147} In the United States, therefore, public entities or officials monitor other public entities or officials. This clearly differs from the state of affairs in Japan.\textsuperscript{148}

The U.S. Federal Acquisition Regulation,\textsuperscript{149} which comprehensively regulates federal procurement activities in the U.S., devotes an entire chapter to bid protests.\textsuperscript{150} Compared with Japan, it is remarkable that in the United States not only procuring entities but also the General Accountability Office, which is responsible for the public accounts, work as institutions handling bid protests.\textsuperscript{151} This means that a public entity checks another public entity as to bid protests as well as monitoring. In Japan, the Board of Audit,\textsuperscript{152} equivalent to the U.S. General Audit Office, does not play such a role.

\textsuperscript{144} 48 C.F.R. §§ 6.501, 6.502(b).
\textsuperscript{145} The Inspectors General serve generally to perform audits and investigations of their respective agencies’ operations and to promote, among other things, economy and efficiency. 5 U.S.C. app. 3 §§ 2, 4 (2006).
\textsuperscript{148} To avoid expanding the analysis in this article beyond its intended scope, the author has intentionally excluded references to court cases involving bid protests as a comparative analysis of such cases would be very complex. It should be noted that, because going to court is the last step in making a bid protest, these must not be ignored. The author intends to write another article with a comparative analysis of bid-protest litigation in the future. Here, the author only points out that the courts strongly tend to dismiss the claims of unsuccessful candidates in Japan. It might be interesting to compare the state of affairs respecting this issue in Japan with those of the U.S. and E.U. countries.
\textsuperscript{149} 48 C.F.R. pts. 1–53 (2006).
\textsuperscript{152} The Board of Audit was established by Kaikei kensain ho [Board of Audit Act], Law No. 73 of 1947.
B. The Hurdles for Japanese Laws to Implement the WTO-GPA

As far as public works are concerned, the Japanese public procurement regimes have become closer to the framework of the WTO-GPA since the APQA was enacted. Namely, both open tendering and selective tendering under the WTO-GPA, which respectively correspond to open competitive tendering and designated competitive tendering in Japan, permit a dialogue between procuring entities and candidates similar to the method sanctioned by the APQA in Japan. Additionally, the APQA and the WTO-GPA are in accord on the point that there is no disproportionate emphasis on price in the contractor award procedure.

Despite some similarities as described above, there are also some differences between the two. For example, in the WTO-GPA, only limited tendering is treated as an exception, whereas not only non-competitive tendering, the equivalent of limited tendering, but also designated competitive tendering are considered to be exceptions under the Japanese laws. Furthermore, as aforementioned, the Japanese regimes include an estimated price system which does not exist under the WTO-GPA.

This comparison is only a superficial analysis. The more important matter is how the Japanese regimes work effectively in implementing the WTO-GPA. The harmonization of the frameworks is merely a tool for that purpose. There are still major hurdles to be overcome in the Japanese public procurement regimes, two of which will now be addressed.

The main targets of the WTO-GPA for free and fair international trade can be said to be: (1) non-discrimination, (2) transparency, and (3) removal of unnecessary trade restrictions. The APQA should be complemented by legal rules implementing these targets and their practices. However, entry barriers can still be set up and procuring entities can still arbitrarily treat certain candidates unfairly even as part of the flexible and complex processes established pursuant to the APQA. In fact, the complexity of the processes actually reduces their

---

153. GPA, supra note 1, art. VII (3)(a), (b).
154. Id. art. XIV.
155. Id. art. XIII(4)(b).
156. Id. art. VII(3)(c) ("Limited tendering procedures are those procedures where the entity contacts suppliers individually, only under the conditions specified in Article XV [of the GPA].").
157. See supra Part II.A.
158. ARROWSMITH, supra note 1, at 168–71.
transparency. To overcome these problems is thought to be the real target for implementing the WTO-GPA. This is the first hurdle.

A second hurdle is the treatment of secondary policies. As noted above, it is probable that in the future, secondary policy concerns will be considered as part of the comprehensive evaluation method.\(^{159}\) Regarding the WTO-GPA, there has been a debate about the approved range of secondary policies under Article 23(2).\(^ {160}\) The greater the importance procuring entities attach to secondary policies, the higher the trade barriers will be. This causes a conflict with the aims of the WTO-GPA. Whatever the secondary policies it allows to be considered, the Japanese government must respond sufficiently to the need to assure transparency concerning practices under the APQA.

V. CONCLUDING REMARKS

Japan is now at the halfway point in undertaking reforms of its government procurement regimes. It is uncertain how the APQA will be enforced in practice because of its short history since enactment. The public’s tendency to regard only price as a credible standard even in the public works field is still deeply rooted in Japanese society. The possibility that the price-only competition method will continue to be the primary one used in fields other than public works in the future cannot be denied.

This article presented a comparative analysis of the relevant legal provisions in Japan, the United States, and the European Union and pointed out several matters Japan should keep in mind when implementing the WTO-GPA. Considering its present industrial and social environments, how will Japan advance toward this goal? A great number of hurdles must be overcome.

\(^{159}\) See supra pt. III.C.1.

\(^{160}\) ARROWSMITH, supra note 1, at 345–46.
SLOBODAN MILOSEVIC AND THE GUARANTEE OF SELF-REPRESENTATION

Joanne Williams*

SECTION ONE: INTRODUCTION .............................................................. 554
  1.1: The Right to Self-Representation ........................................ 555
  1.2: Milosevic at the ICTY ......................................................... 559
  1.3: Milosevic: Self-Representation ........................................ 560
    1.3.1: Prosecution Phase ...................................................... 562
    1.3.2: Defense Phase .......................................................... 563
    1.3.3: Why Such Haste? ...................................................... 567

SECTION TWO: BALANCING RIGHTS .................................................. 569
  2.1: Overarching Right to Fair Trial ........................................ 569
  2.2: Fair and Expeditious Trial .............................................. 571
    2.2.1: In Whose Interest Is a Speedy Trial? ......................... 572
    2.2.2: The Fair Trial Rights of the Accused Remain Paramount .......... 574

SECTION THREE: MILOSEVIC AND THE ASSIGNMENT OF COUNSEL ............... 576
  3.1: Intentional Disruption: “Obstructionism” ............................. 576
    3.1.1: The Willful Nature of Obstructionism ......................... 579
    3.1.2: Was Milosevic an Obstructionist? .............................. 580
    3.1.3: Threshold of Obstructionist Conduct .......................... 582
  3.2: Unintentional Disruption .................................................. 587
    3.2.1: Trial Chamber Decision ............................................ 588
    3.2.2: Appeals Chamber Decision ...................................... 590

SECTION FOUR: PROPORTIONALITY ................................................ 594
  4.1: Mode of Representation Assigned: Proportionate to Aim Pursued? ........ 594
    4.1.1: Proportionality ........................................................ 595
    4.1.2: Dignity and Autonomy ............................................. 598
  4.2: Hybrid Representation and Milosevic .................................. 600

* LL.B., Law and European Studies, University of Limerick, Ireland; LL.M., International Human Rights Law, Irish Centre for Human Rights, National University of Ireland-Galway. The author has worked as a Legal Assistant with the Association of Defence Counsel of the International Criminal Tribunal for the former Yugoslavia and as a fellow at OKO, the Criminal Defence Section of the Court of Bosnia and Herzegovina. She is currently completing the final vocational requirements for admission to the Irish Bar in 2007.
This Article examines the nature of the guarantee of self-representation as exercised in the Milosevic case at the International Criminal Tribunal for the former Yugoslavia (ICTY). The paper debates the manner in which the Milosevic Trial Chamber and the Appeals Chamber of the ICTY balanced the competing interests of the prose rights of the Accused and the need for expedition in the trial. It is argued that the Trial Chamber disproportionately restricted the rights of the Accused in a dubious reasoning decision, a mistake that was partially remedied by the Appeals Chamber, but which has left a legacy that potentially endangers the rights of the accused which subsist in an already precarious environment. The Article concludes with an examination of the merits of hybrid representation in high-level cases such as that of Slobodan Milosevic.

SECTION ONE: INTRODUCTION

The death of Slobodan Milosevic in his cell at the United Nations Detention Unit of the International Criminal Tribunal for the former Yugoslavia (ICTY) in The Hague on March 11, 2006, has raised the question of what can now be salvaged from a trial that ran for over four years and generated forests of testimony, exhibits, litigation, decisions, and appeals. One question that must be asked in the wake of the former President’s death is whether it is appropriate for high-profile criminal defendants to represent themselves. “Once left to a handful of political dissidents and lawyer-haters, self-representation no longer is rare.” High-profile defendants often seek to utilize their statutory right to represent themselves in person before international and regional tribunals in a bid to secure control over a specific defense strategy, often politically motivated.

1. Molly Moore & Daniel Williams, Milosevic Found Dead in Prison; Genocide Trial Is Left Without a Final Judgment, WASH. POST, Mar. 12, 2006, at A01.
2. Over forty-nine thousand pages of transcript were recorded by the last day of trial on March 1, 2006. Trial Tr. at 49,190 (Mar. 1, 2006), Prosecutor v. Milosevic, Case No. IT-02-54-T, available at http://www.un.org/icty/transe54/060301IT.htm. By November 22, 2006, there had been 2,256 filings, comprising 63,775 pages; 930 prosecution exhibits; 85,526 pages of prosecution evidence; 117 videos of prosecution evidence; and 1,245,084 pages of prosecution disclosure, 930,553 of which were under ICTY Rule 68.
4. Those who have featured prominently in the public eye include Theodore Kaczynski, John Allen Muhammad, Zacarias Moussaoui, Ted Bundy, Douglas Clark, Charles Manson, Colin Ferguson, Sheik Omar Abdel-Rahman, Vojislav Seselj, Hinga Norman, and Slobodan Milosevic.
An all-encompassing set of principles regarding the scope of the right of criminal defendants to self-representation has not yet been expressed. The relative significance of the right to self-representation “needs to be defined carefully, particularly in situations when it comes into apparent conflict with the due process guarantee of a fair trial.” There are “inherent tensions between these competing interests” which necessitate “close judicial scrutiny.” International criminal courts and tribunals must address this tension in a systematic fashion in order to obtain a satisfactory equilibrium between these competing interests.

This Article contends that the improvised approach adopted by the Trial Chamber of the ICTY in Prosecutor v. Milosevic to assess the importance of the pro se right weighed disproportionately on the side of expedition. This imbalance was somewhat remedied by the subsequent decision of the Appeals Chamber. As it stands, however, the jurisprudence generated by this case has markedly broadened the potential circumstances in which the right to self-representation may be curtailed, leaving open the possibility of further abrogation not only of the pro se right but also of the other minimum guarantees from which the right to a defense is cast.

1.1: The Right to Self-Representation

In Milosevic, the Trial Chamber of the ICTY (Trial Chamber) recognized that “the international and regional conventions (in similar language) plainly articulate a right to defend oneself in person.” The right to self-representation appears in identical terms in Articles 14(3)(d) of the International Covenant on Civil and Political Rights (ICCPR), 21(4)(d) of the ICTY Statute, 20(4)(d) of the Statute of the International Criminal Tribunal for the Former Yugoslavia.

6. The terms “pro se” and “self-representation” will be used interchangeably in this paper.
7. Homiak, supra note 5, at 936.
8. Id.
9. All references herein to Prosecutor v. Milosevic or Milosevic are to the case of Slobodan Milosevic.
tional Criminal Tribunal for Rwanda (ICTR), and 17(4)(d) of the Statute of the Special Court for Sierra Leone (SCSL). Articles 67(1)(d) of the Rome Statute of the International Criminal Court (ICC), 6(3)(c) of the European Convention for the Protection of Human Rights and Fundamental Freedoms (European Convention on Human Rights) (ECHR), 8(2)(d) of the American Convention on Human Rights (ACHR), and 16(d) of the Agreement for the Prosecution and Punishment of the Major War Criminals of the European Axis (Nuremberg Charter) contain analogous provisions.

The right to self-representation is, under the ICTY Statute, one of a basic set of “minimum guarantees” to which the accused is entitled “in full equality.” Among the other guarantees listed in Article 21(4) of the ICTY Statute are defendants’ right to remain silent; to confront the witnesses against them; “to be tried without undue delay;” and to court-appointed counsel “where the interests of justice so require, and without

15. Rome Statute of the International Criminal Court art. 67, opened for signature July 17, 1998, 2187 U.N.T.S. 90 [hereinafter Rome Statute] (granting the right “to conduct the defence in person or through legal assistance of the accused’s choosing, . . . and to have legal assistance assigned by the Court in any case where the interests of justice so require”).
16. Council of Europe, Convention for the Protection of Human Rights and Fundamental Freedoms art. 6(3)(c), Nov. 4, 1950, Europ. T.S. No. 5, 213 U.N.T.S. 221 [hereinafter European Convention on Human Rights (ECHR)] (granting an accused the right “to defend himself in person or through legal assistance of his own choosing or, if he has not sufficient means to pay for legal assistance, to be given it free when the interests of justice so require”).
18. Agreement by the Government of the United Kingdom of Great Britain and Northern Ireland, the Government of the United States of America, the Provisional Government of the French Republic and the Government of the Union of Soviet Socialist Republics for the Prosecution and Punishment of the Major War Criminals of the European Axis, Charter of the International Military Tribunal art. 16(d), Aug. 8, 1945, 82 U.N.T.S. 279 (granting an accused “the right to conduct his own defence before the Tribunal or to have the assistance of Counsel”).
19. ICTY Statute, supra note 12, art. 21(4).
20. Id. art. 21(4)(g).
21. Id. art. 21(4)(e).
22. Id. art. 21(4)(c).
payment by him” in the case of indigence.\textsuperscript{23} The Appeals Chamber of the ICTY has acknowledged that by placing the right to self-representation “on a structural par” with these guarantees, the drafters of the ICTY Statute recognized the right to self-representation as “an indispensable cornerstone of justice.”\textsuperscript{24}

Article 21(4)(d) incorporates a “binary opposition between representation ‘through legal assistance’ and representation ‘in person.’”\textsuperscript{25} The personal character of the individual’s procedural rights is recognized by acknowledging the defendant’s right to make his own choice.\textsuperscript{26} The pro se right “embodies one of the most cherished ideals of civilization: the right of an individual to determine his own destiny.”\textsuperscript{27} Furthermore, the pro se process “affirm[s] the dignity and autonomy of the accused.”\textsuperscript{28} On the other hand, it has been recognized that “[t]he presence of counsel serves both to protect the accused from prosecutorial overreaching and to preserve society’s interest in the integrity of the judicial process.”\textsuperscript{29} The benefits derived from the presence of counsel in certain circumstances have spurred contemporary criminal tribunals to make self-representation a qualified, and not an absolute, guarantee.\textsuperscript{30}

It is widely recognized that the right to self-representation “is not categorically inviolable.”\textsuperscript{31} For example, various common law jurisdictions recognize the capacity of courts to restrict the pro se right in sexual as-

\begin{footnotesize}
\begin{itemize}
\item \textsuperscript{23} Id. art. 21(4)(d).
\item \textsuperscript{24} Milosevic v. Prosecutor, Case No. IT-02-54-AR73.7, Decision on Interlocutory Appeal of the Trial Chamber’s Decision on the Assignment of Defence Counsel, para. 11 (Nov. 1, 2004), available at http://www.un.org/icty/milosevic/appeal.decision-e/041101.htm.
\item \textsuperscript{25} Id.
\item \textsuperscript{26} See Note, The Accused as Co-Counsel: The Case for the Hybrid Defense, 12 VAL. U. L. REV. 329, 350 (1978) [hereinafter Hybrid Defense].
\item \textsuperscript{27} People v. Gordon, 688 N.Y.S.2d 380, 382 (N.Y. Sup. Ct. 1999).
\item \textsuperscript{29} Judith Welcom, Note, Assistance of Counsel: A Right to Hybrid Representation, 57 B.U. L. REV. 570, 571 (1977).
\item \textsuperscript{31} Milosevic v. Prosecutor, IT-02-54-AR73.7, Decision on Interlocutory Appeal of the Trial Chamber’s Decision on the Assignment of Defense Counsel, para. 12 (Nov. 1, 2004).
\end{itemize}
\end{footnotesize}
sault trials “to protect vulnerable witnesses from trauma.” Furthermore, in the United States, the right to self-representation does not extend to appellate proceedings. Civil law jurisdictions further restrict the pro se right by often forcing representation by counsel upon defendants “in serious criminal cases.” This practice has been upheld by the European Court of Human Rights (ECtHR). The importance of the right to self-representation is not diminished by the fact that other interests may supersede it. As eloquently put by Colquitt, “[j]ustice, it seems, begs for the striking of balances and the fashioning of procedural accommodations.” It is, of course, essential to


37. Id. at 127.
“balance the rights and interests of defendants against other important rights and interests in a manner fair to all” parties, without excluding victims, witnesses, or defendants.38

Recognizing that a defendant’s right to represent himself or herself is subject to some limitations does not, however, resolve the issue at hand. It must also be shown that the restriction applied was justified and proportionate to the interest pursued. Section Two explores the balancing of the interests of the accused with the interests of victims, witnesses, and the Tribunal itself. Section Three will analyze the balancing of these rights in the curtailment of the pro se right in Milosevic, and Section Four expands upon the need for proportionality in such balancing exercises.

1.2: Milosevic at the ICTY

Slobodan Milosevic was indicted at the Yugoslavia Tribunal on May 24, 1999, for alleged atrocities committed in Kosovo.39 He remained in power as Yugoslav President until October 200040 and was re-elected leader of Serbia’s Socialist Party in November 2000.41 The former President was finally transferred to The Hague in June 2001,42 and, in November 2001, was charged with twenty-nine criminal counts, including genocide, with regard to his involvement in the Bosnian War from 1992–95.43 Two further indictments were brought against him, with five counts

38. Id. at 65.
42. R. Jeffrey Smith, Serb Leaders Hand Over Milosevic for Trial by War Crimes Tribunal; Extradition Sparks Crisis in Belgrade, WASH. POST, June 29, 2001, at A1.
43. Prosecutor v. Milosevic, Case No. IT-01-51-I Indictment (Nov. 22, 2001), available at http://www.un.org/icty/indictment/english/mil-ii011122e.htm; see William Drozdiak, Milosevic to Face Genocide Trial for Role in the War in Bosnia; Yugoslav Ex-Leader First Head of State to Be So Charged, WASH. POST, Nov. 25, 2001, at A22. The Bosnia Indictment was joined with those concerning Croatia and Kosovo, see infra note 47 and accompanying text, and subsequently amended. Prosecutor v. Milosevic, Case No.
concerning his activities in Kosovo in 1999\textsuperscript{44} and a further thirty-two relating to Croatia in 1991–92.\textsuperscript{45} The three indictments were joined together by the Appeals Chamber,\textsuperscript{46} and the trial commenced on February 12, 2002.\textsuperscript{47} Until his death on March 11, 2006, Milosevic stood accused of sixty-six counts, comprising seventeen substantive crimes.\textsuperscript{48}

1.3: Milosevic: Self-Representation

From the outset, Slobodan Milosevic indicated that he wished to represent himself and, accordingly, did not wish to be represented by coun-

\begin{itemize}
\item Trial Tr. at 1 (Feb. 12, 2002), Prosecutor v. Milosevic, Case No. IT-02-54-T, available at http://www.un.org/icty/transe54/020212IT.htm.
\end{itemize}
This appears to have been motivated by his non-recognition of the Tribunal’s legitimacy. Judge May assured the Accused in July 2001 that “[y]ou do have the right, of course, to defend yourself.” The following month, the Trial Chamber invited the Registrar to appoint amici curiae, not to represent Milosevic but rather to ensure a fair trial and “assist [the Chamber] in the proper determination of the case.” In order to assist the Chamber to secure a fair trial, the amici were to bring “exculpatory or mitigating evidence” to the Trial Chamber’s attention, to inform the Chamber of any defenses the Accused could properly raise, to “mak[e] submissions as to the relevance . . . of the NATO air campaign in Kosovo,” and to identify witnesses whom the Chamber could call. The Chamber further enhanced the rights of the Accused in April 2002 by recognizing his right to communicate with legal advisers and by granting him privileged communication with named Legal Associates.

55. Id.
56. Id.
57. Prosecutor v Milosevic, Case No. IT-02-54-T, Order, para. 10 (Apr. 16, 2002).
58. Prosecutor v. Milosevic, Case No. IT-02-54-T, Order, para. 14 (Apr. 16, 2002) (permitting Milosevic to have privileged communication with lawyers Zdenko Tomano-vic and Dragoslav Ognjanovic). Branko Rakic was later appointed as a third “Associate.” Milosevic, Case No. IT-02-54-T, Reasons for Decision on Assignment of Defence Counsel, para. 5 (Sept. 22, 2004) (citing Milosevic, Order Appointing Branko Rakic as Legal
In response to a suggestion in 2001 from the Prosecution that defense counsel should be assigned to the Accused alongside the Amici, the Chamber reiterated that “the accused has a right to counsel, but he also has a right not to have counsel.” The Trial Chamber consistently upheld this position on the basis that “it would be wrong for the Chamber to impose counsel on the accused, because that would be in breach of the position under customary international law.”

1.3.1: Prosecution Phase

The Trial Chamber first expressed concern about the completion of the trial in November 2002 in light of the state of the Accused’s ill health. The Prosecution again sought to have defense counsel imposed on the Accused on the basis that by proceeding pro se, the Accused had exacerbated his health problems. This, the Prosecution suggested, created “self-imposed” difficulties, dictated the scope of the trial, and obtained for the Accused a trial that was “significantly less complete than it would otherwise be.”

This submission was rejected in December 2002 and written reasons were issued in April 2003. The Trial Chamber reasoned that the “present circumstances” were not such that the tribunal could assign counsel

60. Id. at 18:9–11.
62. Prosecutor v. Milosevic, Case No. IT-02-54-T, Reasons for Decision on the Prosecution Motion Concerning Assignment of Counsel, para. 7 (Apr. 4, 2003) (citing Milosevic, Submission from the Office of the Prosecutor on the Future Conduct of the Case in the Light of the State of the Accused’s Health and the Length and Complexity of the Case (Nov. 8, 2002); Milosevic, Observations by the Amici Curiae on the Imposition of Defence Counsel on Accused (Nov. 18, 2002); Milosevic, Addendum to the Prosecution’s Response to the Confidential Observations by the Amici Curiae on the Health of the Accused and the Future Conduct of the Trial (Nov. 20, 2002)).
63. Id. para. 10 (quoting Milosevic, Submission from the Office of the Prosecutor on the Future Conduct of the Case in the Light of the State of the Accused’s Health and the Length and Complexity of the Case, para. 4 (Nov. 8, 2002)).
64. Trial Tr. at 14574:14–17 (Dec. 18, 2002), Prosecutor v. Milosevic, Case No. IT-02-54-T (“[D]efence counsel will not be imposed upon the accused against his wishes in the present circumstances. It is not normally appropriate in adversarial proceedings such as these. The Trial Chamber will keep the position under review.”).
It noted that the duty imposed by Article 20(1) of the ICTY Statute to ensure a fair and expeditious trial must be implemented “‘with full respect for the rights of the accused,’”67 Crucially, however, the Trial Chamber stressed that “there may be circumstances . . . where it is in the interests of justice to appoint counsel,” and resolved to “keep the position under review.”68

1.3.2: Defense Phase

The ill health of the Accused began to cause more frequent disruption as the Prosecution phase of the trial advanced. During the Prosecution’s case, the trial was adjourned on several occasions for two to three weeks or more to allow the Accused to recuperate.69 Due to Milosevic’s high blood pressure and heart condition, his trial was reduced to a three-day week, and, further, to a four-hour day beginning September 2003.70 In an attempt to speed up proceedings, the Trial Chamber reduced the time

66. Id. para. 18.
67. Id. para. 41.
68. Id. para. 40.
70. Pursuant to the Medical Report of Dr. van Dijkman of August 26, 2002, the Trial Chamber mandated that four consecutive days of rest be given every two weeks of trial. Prosecutor v. Milosevic, Case No. IT-02-54-T, Reasons for Decision on Assignment of Defence Counsel, para. 53 (Sept. 22, 2004) (citing Medical Report of Dr. van Dijkman, Aug. 26, 2002). The trial continued as such until the end of September 2003 when, in accordance with further medical recommendations (Id., citing Medical Report of Dr. van Dijkman, Sept. 26, 2003), “the Trial Chamber decided to sit three days each week, to allow the Accused sufficient time to rest.” Id. (citing Trial Tr. at 27063 (Sept. 30, 2003), Prosecutor v. Milosevic, Case No. IT-02-54-T). From then on, the trial ran from 9a.m. to 1:45p.m. (with two fifteen minute intervals) three days per week.
available to the Prosecution,\textsuperscript{71} causing prosecutors to argue that “their case [wa]s being ‘emasculated.’”\textsuperscript{72}

The Prosecution closed its case on February 25, 2004,\textsuperscript{73} at which stage sixty-six trial days had been lost.\textsuperscript{74} The Prosecution’s case was interrupted on thirteen occasions on account of the Accused’s illness, eight of which related exclusively to the Accused’s high blood pressure.\textsuperscript{75} Between February and June 2004, doctors advised the Accused to rest for fifty-one weekdays.\textsuperscript{76} Due to Milosevic’s poor health, the Tribunal delayed the original starting date for the defense case, originally June 8,

\begin{itemize}
\item \textsuperscript{71} Trial Tr. at 2784 (Apr. 10, 2002), Prosecutor v. Milosevic, Case No. IT-02-54-T (“W[We have decided that the Prosecution should have one year from today to conclude their case. That will give them a total of 14 months in which to finish the case, their case. In the view of the Trial Chamber, no Prosecution case should continue for a period longer than that.”); see Prosecutor v. Milosevic, Case No. IT-02-54-T, Decision on Prosecution’s Request to Have Written Statements Admitted Under Rule 92bis, para. 27 (Mar. 21, 2002) (limiting the number of Prosecution witnesses).
\item \textsuperscript{73} Prosecutor v. Milosevic, Case No. IT-02-54-T, Order Rescheduling and Setting the Time Available to Present the Defence Case, para. 2 (Feb. 25, 2004), available at http://www.un.org/icty/milosevic/trialc/order-e/040225.htm.
\item \textsuperscript{74} Prosecutor v. Milosevic, Case No. IT-02-54-T, Order on Future Conduct of the Trial, para. 6 (July 6, 2004), available at http://www.un.org/icty/milosevic/trialc/order-e/040706.htm.
\item \textsuperscript{75} Prosecutor v. Milosevic, Case No. IT-02-54-T, Reasons for Decision on Assignment of Defence Counsel, para. 56 n.120 (Sept. 22, 2004). Of the sixty-six trial days lost, twenty-eight were in 2002, thirty-one in 2003, and seven in the first two months of 2004. \textit{Id.}
\item \textsuperscript{76} Prosecutor v. Milosevic, Case No. IT-02-54-T, Scheduling Order for a Hearing, para. 11 (Nov. 22, 2005), available at http://www.un.org/icty/milosevic/trialc/order-e/051122.htm (citing “Report by the Registrar Pursuant to the Trial Chamber’s ‘Omnibus Order on Matters Dealt with at the Pre-Defence Conference’, filed on 18 June 2004,” para. 7 (June 25, 2004)). This total amount of days was based on a five-day working week. Under a “three-day-per-week analysis,” the amount of days lost was thirty-one. The Accused was found to have used the equivalent of eleven of these days in preparation of his case. Prosecutor v. Milosevic, Case No. IT-02-54-AR73.7, Appeal Against the Trial Chamber’s Decision on Assignment of Defence Counsel: Corrigendum, para. 5 (Sept. 29, 2004) (referring to Prosecutor v. Slobodan Milosevic, Case No. IT-02-54-T, “Report by the Registrar Pursuant to the Trial Chamber’s ‘Omnibus Order on Matters Dealt with at the Pre-Defence Conference’, filed on 18 June 2004,” para. 7 (June 25, 2004)).
\end{itemize}
2004, on five occasions. Milosevic’s health was progressively becoming a major obstacle to the expeditious completion of the case.

In July 2004, the Trial Chamber noted that on the basis of the time lost due to the Accused’s recurring ill health, it was “necessary to carry out a radical review of the future conduct of the trial.” At this stage of the trial, His Honor Judge May, who had been forced to resign due to ill health (and who subsequently died), had been replaced. Preemptively, the Chamber suggested that “it may be necessary to assign counsel to the Accused, and/or adopt other measures to ensure a fair and expeditious conduct of the trial.” It also noted “the resolve and determination of the Trial Chamber to conclude the presentation of the defense case by October 2005.” This was perhaps the first indication of concern with the expeditious completion of the trial that, by the time of the Accused’s death, had become unyielding.

78. Prosecutor v. Milosevic, Case No. IT-02-54-T, Reasons for Decision on Assignment of Defence Counsel, para. 59 (Sept. 22, 2004).
79. Prosecutor v. Milosevic, Case No. IT-02-54-T, Order on Future Conduct of the Trial, para. 15 (July 6, 2004).
81. Prosecutor v. Milosevic, Case No. IT-02-54-T, Order on Future Conduct of the Trial, para. 21 (July 6, 2004).
82. Prosecutor v. Milosevic, Case No. IT-02-54-T, Further Order on Future Conduct of the Trial, para. 5 (July 19, 2004), available at http://www.un.org/icty/milosevic/trialc/order-e/040719.htm. However, it later became apparent that the defense case would proceed significantly beyond this time. Milosevic, Decision in Relation to Severance, Extension of Time and Rest, para. 25 (Dec. 12, 2005), available at http://www.un.org/icty/milosevic/trialc/decision-e/051212.htm (“The conclusion of the Accused’s allotted time will take the trial well into March 2006. Once rebuttal and rejoinder cases are heard and concluding arguments made, it is likely the trial hearings would still not conclude until the middle of 2006. Judgement drafting will occupy a further substantial period.”).
In an oral ruling on September 2, 2004, the Trial Chamber ordered the assignment of defense counsel to the Accused. Accordingly, Mr. Steven Kay QC and Ms. Gillian Higgins, who previously functioned as Amici Curiae, were appointed to this role. The modalities of the assignment were outlined in an order issued on September 3, according to which it would be the role of the Assigned Counsel “to determine how to present the case for the Accused.” In particular, the Assigned Counsel were to:

- represent the Accused by preparing and examining those witnesses court Assigned Counsel deem it appropriate to call;
- make all submissions on fact and law that they deem it appropriate to make;
- seek from the Trial Chamber such orders as they consider necessary to enable them to present the Accused’s case properly . . . [and]
- discuss with the Accused the conduct of the case, endeavour to obtain his instructions thereon and take account of views expressed by the Accused, while retaining the right to determine what course to follow.

Under this system, it was only with the leave of the Trial Chamber that the Accused could “continue to participate actively in the conduct of his case, including, where appropriate, examining witnesses, following examination by court Assigned Counsel.” This order whipped the helm from the hands of the Accused and installed the newly Assigned Counsel in his place, giving them full control over the course of the defense strategy. As will be seen in Section Four, this disproportionate move stripped the Accused of his dignity and autonomy. The rationale behind the decision was expounded upon in a written ruling issued on September 22, 2004. On this occasion, the Chamber extracted a different conclusion from an examination of much of the same jurisprudence upon which it

---

84. It should be noted that this Article will adopt the term “assignment” as used by the Tribunal. This will avoid the negative connotations of the term “imposition” which is often used in relation to this case.
86. Id. para. 6(1).
87. Id. para. 6(1)(a)–(d). The Tribunal also instructed the Assigned Counsel to “act throughout in the best interests of the Accused.” Id. para. 6(1)(e).
88. Id. para. 6(2).
89. Prosecutor v. Milosevic, Case No. IT-02-54-T, Reasons for Decision on Assignment of Defence Counsel (Sept. 22, 2004).
had based its April 2003 ruling. This conclusion is debated in Section Three.

Following the Accused’s oral instigation, the Assigned Counsel lodged an appeal against the assignment that was rejected by the Appeals Chamber in November 2004.90 The Appeals Chamber upheld the Trial Chamber’s exercise of discretion.91 Crucially, however, the Appeals Chamber instructed the Trial Chamber to radically alter the modalities to be followed by the Assigned Counsel, according a far greater role to the Accused.92 This decision essentially reinstated Milosevic as “captain of the ship.”93 The significance of this judgment is expounded in Section Four. The Trial Chamber, in December 2004, refused the Assigned Counsel’s motion to withdraw,94 and refused leave to appeal this decision.95 The Registrar’s ensuing refusal to allow the Assigned Counsel to withdraw was reaffirmed by the ICTY President in February 2005.96 The Assigned Counsel continued, therefore, to function alongside the accused until the trial was terminated in March 2006.

1.3.3: Why Such Haste?

There is, at present, an unconcealed push for a degree of finality to proceedings in The Hague. The Tribunal “was established as a temporary tribunal with a finite mission.”97 While this was not clearly articulated in the ICTY Statute, it is clear that this was the assumption implied by the Report of the Secretary General upon the establishment of the Tribunal, whereby “the life span of the international tribunal would be linked to the restoration and maintenance of international peace and security in the

91. Id. paras. 15, 19.
92. Id. para. 19.
93. Id.
territory of the former Yugoslavia, and Security Council decisions related thereto.”

Thus, “concerns have been voiced not only by United Nations officials, Member States and others, but also by all the organs of the Tribunals with regard to the slowness of the pace of proceedings, the associated length of detention of accused, [and] the length and cost of Tribunal operations . . . .”

With proceedings remaining “exceptionally lengthy, costly, and complicated,” ambitious strategies had been adopted by the U.N. Security Council with a view to hastening the pace of progress. The ICTY envisaged the completion of “investigations by 2004 . . . first instance trials by 2008,” and all work in 2010. These target deadlines have created a palpable concern with the Tribunal’s swift administration of justice, a concern which clearly manifested itself in the Milosevic case.


103. Id. para. 75.
What emerged in the Milosevic conflict between self-representation and the principle of a speedy trial was a balancing test between the personal rights of the Accused and the interest of the tribunal in achieving a fair and expeditious trial. The Tribunal ultimately found that the efficient administration of justice to prevail in light of the overarching need to secure a fair trial. This decision confirms that self-representation is not an institutional component of the adversary process, but rather a privilege that can be withdrawn in certain circumstances in the interests of fairness and expedition. This section examines the legitimacy of that claim.

2.1: Overarching Right to Fair Trial

In its decision to assign counsel, the Milosevic Trial Chamber found that “[t]he minimum guarantees set out in Article 21(4) of the Statute are elements of the overarching requirement of a fair trial.” Essentially, the Trial Chamber subsumed the right to represent oneself in person into a single “right to a defense,” which in turn forms just one of several elements in the “overarching” right to a fair trial.

In accordance with the Vienna Convention on the Law of Treaties, the Trial Chamber found that “when read in light of the object and purpose of securing [the] . . . right to a . . . fair trial,” the right to represent oneself under Article 21(4)(d) “may be lost if the effect of its exercise is to obstruct the achievement of that object and purpose.” Thus, as articulated by Justice Frankfurter of the U.S. Supreme Court in Adams v. United States ex rel. McCann, “[w]hat were contrived as protections for the accused should not be turned into fetters.”

This reasoning was heavily inspired by the jurisprudence of the ECtHR, which considers all minimum rights included in Article 6(3) of the ECHR in the context of the overall purpose of bringing about a fair trial.

trial.\textsuperscript{108} Under this premise, the list of minimum guarantees set out in Article 6(3) of the ECHR (substantially equivalent to Article 21(4) of the ICTY Statute) reflects aspects of the notion of a fair trial.\textsuperscript{109} Article 21(4)(d) is not, therefore, simply a list of unconnected guarantees, but rather “a compact statement of the rights necessary to a full defense”\textsuperscript{110} which must be considered in the broader context of the right to a fair trial.


\textsuperscript{110} Faretta v. California, 422 U.S. 806, 818 (1975). Faretta makes this statement with regard to the Sixth Amendment to the U.S. Constitution; the rights guaranteed by the Sixth Amendment and Article 21(4) of the ICTY Statute are substantially similar. Compare U.S. CONST. amend. VI with ICTY Statute art. 21(4), May 25, 1993, 32 I.L.M. 1192.
Self-representation is simply a means through which this right can be secured. Recent jurisprudence of the HRC, ECtHR, and ICTY support this premise, as does the U.S. Supreme Court decision in *Faretta*.

### 2.2: Fair and Expeditious Trial

Under Article 21(4)(c) of the ICTY Statute, the accused has the right to be tried “without undue delay.” Moreover, trial chambers have a duty, inter alia, to ensure a “fair and expeditious” trial under Article 20(1).
Article 20(1) does not simply concern the accused, but rather imposes a positive duty on the trial chamber in the public interest. The commonly cited principle of speedy trial refers to a combination of these three guarantees, and entails consideration of a diverse range of interests. The trial chambers must, therefore, in balancing various rights with the principle of a fair and expeditious trial, have due regard for interests other than those of the accused.

2.2.1: In Whose Interest Is a Speedy Trial?

It is primarily the accused who has an interest in a speedy trial. Prompt trials go some way to ensuring that the defendant can mount an effective defense. Speedy trials are primarily designed “(i) to prevent oppressive pretrial incarceration; (ii) to minimize anxiety and concern of the accused; and (iii) to limit the possibility that the defense will be impaired.” Prompt trials ensure that witnesses’ memories do not fade and evidence is not destroyed nor disappears. The limitation of pretrial detention is particularly important at the Yugoslavia Tribunal, where many defendants are held for lengthy periods before trial due to, inter alia, the Tribunal’s overwhelming caseload.

The interests of the Prosecution, victims, and witnesses must also be considered, along with those of unrelated defendants awaiting trial.

---

Evidence of the SCSL, Rule 26 bis (as amended Mar. 7, 2003) (similar language); Rome Statute, supra note 12, art. 64(2) (similar language).

119. ICTY Statute, supra note 12, art. 20(1) (“The Trial Chambers shall ensure that a trial is fair and expeditious and that proceedings are conducted in accordance with the rules of procedure and evidence, with full respect for the rights of the accused and due regard for the protection of victims and witnesses.” (emphasis added)).


122. Akhil Reed Amar, The Constitution and Criminal Procedure: First Principles 90 (1997) (“[I]f government holds the accused in extended pretrial detention, courts must ensure that the accuracy of the trial itself will not thereby be undermined—as might occur if an innocent defendant’s prolonged detention itself causes the loss of key exculpatory evidence.”).


124. Prosecutor v. Tadic, Case No. IT-94-1, Decision on the Prosecutor’s Motion Requesting Protective Measures for Victims and Witnesses, para. 55 (Aug. 10, 1995), available at http://www.un.org/icly/tadic/trialc2/decision-e/100895pm.htm (“A fair trial means not only fair treatment to the defendant but also to the prosecution and to the witnesses.”); see also Dietrich v. The Queen (1992) 177 C.L.R. 292, 335 (Deane, J.) (Austl.) (stating that “the interests of the Crown acting on behalf of the community” must be
and the international community. In an enlightening dissent in *Milosevic*, Judge Shahabuddeeen posited that:

> The fairness of a trial is the result of the fairness of the system of justice employed. The latter depends on the striking of a balance between two competing public interests. First, there is the justly publicized public interest in respecting the rights of the accused. Second, there is the less proclaimed but equal public interest in ensuring that crimes are properly investigated and duly prosecuted.

It was recognized in *Milosevic* that “the Tribunal has its own distinct set of interests at stake in this case, including first and foremost the interest in an outcome that is just, accurate, and reasonably expeditious.”

This premise has consistently been reiterated by the Trial Chamber.

considered in “determining the practical content” of the right to a fair trial (quoting Barton v. The Queen (1980) 147 C.L.R. 75, 101 (Austl.) (Gibbs, A.C.J. & Mason, J.)).


126. Prosecutor v. Aleksovski, Case No. IT-95-14/1 (Appeals Chamber), Decision on Prosecutor’s Appeal on Admissibility of Evidence, para. 25 (Feb. 16, 1999), available at http://www.un.org/icty/aleksovski/appeal/decision-e/90216EV36313.htm (“[T]he Prosecution acts on behalf of and in the interests of the community, including the interests of the victims of the offence charged (in cases before the Tribunal the Prosecutor acts on behalf of the international community).”).


128. Prosecutor v. Milosevic, Case No. IT-02-54-T, Decision Affirming the Registrar’s Denial of Assigned Counsel’s Application to Withdraw, para. 11 (Feb. 7, 2005) (footnote omitted) (emphasis added); see also *Milosevic*, Trial Tr. at 32358:12–19 (Sept. 2, 2004), available at http://www.un.org/icty/transe54/040902IT.htm (“The fundamental duty of the Trial Chamber is to ensure that the trial is fair and expeditious.”).

129. Prosecutor v. Delalic, Case No. IT-96-21-T, Decision on Request by Accused Mucic for Assignment of New Counsel, para. 3 (June 24, 1996), available at http://www.un.org/icty/celebici/trialc2/decision-e/60624DS2.htm (emphasizing the “overriding interest of the administration of justice” and that the Tribunal had to “be satisfied that the reasons [for the defendant’s dissatisfaction with assigned counsel] are genuine and that the request [for assignment of new counsel] is not being made for frivolous reasons or in a desire to pervert the course of justice, e.g., by causing additional delay”); Prosecutor v. Seselj, IT-03-67-PT, Decision on Prosecution’s Motion for Order Appointing Counsel to Assist Vojislav Seselj with His Defence, para. 21 (May 9, 2003). The Seselj Tribunal also stated that “the right to a fair trial . . . is not only a fundamental right of the Accused, but also a fundamental interest of the Tribunal related to its own legitimacy.” *Id.*
and the Appeals Chamber, by other tribunals, and in national jurisdictions. In sum, the Tribunal has a legitimate interest in ensuring that justice is being done and seen to be done.

2.2.2: The Fair Trial Rights of the Accused Remain Paramount

While the principle of a speedy trial necessarily encompasses diverse interests, it is particularly instructive that under Article 20(1) of the ICTY Statute a fair and expeditious trial must be achieved “with full respect for the rights of the accused and due regard for the protection of victims and witnesses.” This formulation suggests that the interests of the accused must be given precedence. A further indication of the primacy of the interests of the accused is found in the assertion made in the ICTR case Kanyabashi that the “object and purpose” of the ICTR Statute is to secure “a fair and expeditious trial.” According to Falvey, “[t]he protection of victims and witnesses, although a laudable goal, must yield to the right to a fair trial when the two conflict.”

130. Milosevic v. Prosecutor, Case No. IT-02-54-AR73.7, Decision on Interlocutory Appeal of the Trial Chamber’s Decision on the Assignment of Defence Counsel, para. 19 (Nov. 1, 2004) (noting “the Tribunal’s basic interest in a reasonably expeditious resolution of the cases before it”).

131. See Prosecutor v. Norman, Case No. SCSL-04-14-T, Decision on the Application of Samuel Hinga Norman for Self Representation under Article 17(4)(d) of the Statute of the Special Court, para. 26 (June 8, 2004) (considering “the public interest, national and international, in the expeditious completion of the trial” in determining whether to allow the accused to represent himself).

132. It has, for instance, consistently been reaffirmed by the High Court of Australia that when “determining the practical content” of the right to a fair trial, “regard must be had to the interests of the Crown acting on behalf of the community as well as to the interests of the accused.” Dietrich v. The Queen (1992) 177 C.L.R. 292, 335 (Deane, J.) (quoting Barton v. The Queen (1980) 147 C.L.R. 75, 101 (Austl.) (Gibbs, A.C.J. & Mason, J.)). Similarly, the U.S. Supreme Court felt that the right of a defendant to be present at trial must be held to be subordinate to the overriding need to maintain orderly and dignified proceedings, which is “essential to the proper administration of criminal justice.” Illinois v. Allen, 397 U.S. 337, 343 (1970).

133. ICTY Statute, supra note 12, art. 20(1) (emphasis added).


The ECtHR has held that Article 6(1) of the ECHR, equivalent to Article 20(1) of the ICTY Statute, “is intended above all to secure the interests of the defence and those of the proper administration of justice.”\textsuperscript{136} Furthermore, the ECtHR has held that “[t]he right to a fair administration of justice . . . cannot be sacrificed to expediency.”\textsuperscript{137} Hence, the safeguards accorded to the accused should not be excessively curtailed in the interest of achieving a speedy conclusion at trial.

The ICTY Rules provide little guidance on which rights take precedence in the event of a conflict between them. This problem will become more pronounced, given the elevated status of victims and witnesses in the Rome Statute\textsuperscript{138} and recent pronouncements of the ECtHR that appear to accord new rights to victims.\textsuperscript{139} It is instructive that Article 68(1) of the Rome Statute dictates that measures taken to protect victims and witnesses “shall not be prejudicial to or inconsistent with the rights of the accused or a fair and impartial trial.”\textsuperscript{140} This formulation again appears to give precedence to the rights of the accused. These authorities must be given due weight in the balancing of rights.

The Trial Chamber’s original approach of allowing Milosevic to proceed pro se indisputably exacerbated the already lengthy nature of that trial.\textsuperscript{141} In response, the Chamber had to balance the pro se right of the accused with its own statutory duty to secure a fair and expeditious trial. The following section will address the manner in which the Chamber balanced these rights. It is not denied that the Chamber indeed had a genuine and legitimate interest in curtailing the pro se right of the Accused. Rather, it is argued that the Chamber erred in the principles employed to achieve this goal.

\textsuperscript{138} See, e.g., Rome Statute, supra note 15, art. 68 (containing provisions for the protection of victims and witnesses).
\textsuperscript{139} See, e.g., M.C. v. Bulgaria, 2003-XII Eur. Ct. H.R. 1, 37–38, paras. 177–78. Here, the ECtHR appears to accord a right of a (rape) victim to compel investigators and prosecutors to confront witnesses in order to assess the credibility of conflicting evidence.
\textsuperscript{140} Rome Statute, supra note 15, art. 68(1); see also id. art. 68(3), 68(5) (containing same provision with respect to specific protective measures).
\textsuperscript{141} It is important to note that a trial’s length depends on many factors, “including the scope of the indictment, the breadth of the dispute between the parties and the complexity of the facts.” Richard May & Marieke Wierda, INTERNATIONAL CRIMINAL EVIDENCE 279 (2002).
SECTION THREE: MILOSEVIC AND THE ASSIGNMENT OF COUNSEL

As demonstrated by the foregoing, it has been clearly established that adjudicative bodies have the power to restrict the right to self-representation in the interest of the fair and expeditious administration of justice. Upon this foundation, it is widely accepted that a defendant’s pro se right may be restricted in the case of *deliberate* trial disruption, to which the author will refer as “obstructionism.” This principle is based on the rationale that self-representation may not be used as a tactic to delay the trial. By necessity, the employment of tactics implies the existence of resolve, volition, and intention. Thus, the concept of obstructionism may be said to relate solely to the willful conduct of the accused deliberately aimed at the disruption of trial proceedings. It follows that this concept does not encompass disruption caused by unintended or extraneous factors.

Controversially, the Trial Chamber ruled in Milosevic that “[t]here is no difference in principle between deliberate misconduct which disrupts the proceedings and any other circumstance which so disrupts the proceedings as to threaten the integrity of the trial.”142 This reasoning effectively led to the conclusion that delays accruing due to the Accused’s ill health had the same potential as a defendant’s deliberately obstructionist actions to damage the integrity of proceedings. In other words, the damage that disruption causes to the integrity of a trial will be the same regardless of the cause or purpose of that disruption. This assertion will be challenged in this section.

It is contended that while the recurring ill health of Milosevic undeniably inhibited the expeditious completion of the trial, it cannot be said to have undermined the integrity of proceedings. Ill health must be differentiated from deliberate obstructionist behavior. The following section outlines why the decision to assign counsel could not defensibly have been based on the principle of obstructionism, thereby leading the Trial Chamber to avoid this rubric, and to create an entirely new and unfounded premise in law in order to fulfill the aims of the completion strategy.

3.1: Intentional Disruption: “Obstructionism”

The concept of obstructionism is derived from the fact that the “pro se right is circumscribed by the requirement that the defendant not disregard

---

the dignity, order and decorum of judicial proceedings.” In the seminal case of *Faretta v. California*, the U.S. Supreme Court acknowledged that the right to proceed pro se may be denied where the defendant *deliberately* undertakes to hinder the trial’s orderly conduct by engaging in obstreperous behavior. This was based on the premise that the right to self-representation was not intended as a license either “to abuse the dignity of the courtroom” or to ignore either the rules of procedure or substantive law. This case clearly sought to regulate voluntary misconduct specifically designed to interrupt proceedings. Nowhere does *Faretta* imply that a defendant may lose his pro se right if he unintentionally consumes too much time in exercising it. It is imperative that international legal proceedings are seen to be conducted efficiently, and with dignity and decorum. In other words, “justice must satisfy the appearance of justice.” Disruptions, particularly where intentional, inevitably damage the public perception of the Tribunal. Accordingly, “the interest of a court in stopping disruption of its


144. *Faretta v. California*, 422 U.S. 806, 834 n.46 (1975) (“[T]he trial judge may terminate self-representation by a defendant who deliberately engages in serious and obstructionist misconduct.” (citation omitted)).

145. *Id.*

146. *Id.*; see also Ramirez Ferrel v. Superior Court of Los Angeles County, 576 P.2d 93, 95 (Cal. 1978) (noting that, as supported by *Faretta*, “an accused should only be deprived of the right to self-representation when he engages in disruptive in-court conduct which is inconsistent with its proper exercise”).

147. *See Faretta*, 422 U.S. at 834 n.46 (“[T]he right to self-representation does not exist . . . to be used as a tactic for . . . disruption . . . .”); see also United States v. Egwaoje, 335 F.3d 579, 586 (7th Cir. 2003) (concluding that the fact that the accused “engaged in a pattern of obfuscation and obstructionism” supported the “knowing and intelligent” nature of his waiver of the right to counsel). The connotations of the word “obfuscation” in this context clearly indicate that the conduct to which the exception applies involves control and resolve. *See Illinois v. Allen*, 397 U.S. 337, 343 (1970); Commonwealth v. Africa, 353 A.2d 855, 864 (Pa. 1976).

148. *See Memorandum of Law Regarding Defendant’s Motion to Proceed Pro Se and Status of Counsel at 11 n.19, United States v. Moussaoui, 2002 U.S. Dist. LEXIS 11135 (E.D. Va. 2002) (No. 01-455-A) (“[T]he defendant’s pro se right is circumscribed by the requirement that the defendant does not disregard the dignity, order and decorum of judicial proceedings.” (citations omitted)).

proceedings, and the consequent threat to the integrity of the trial, is a component of the overarching right to a fair trial.”

In Seselj, the Tribunal advanced the “legitimate interest in ensuring that the trial proceeds in a timely manner without interruptions, adjournments or disruptions” to justify the assignment of standby counsel. This interest was reaffirmed in Delalic. In Prlic, the Tribunal stated that “it is the duty of the Trial Chamber to make sure that the proceedings would not be halted by foreseeable, and therefore avoidable, risks.” Likewise, in the SCSL case of Norman, the great potential for further disruption to the court’s timetable and calendar was among the factors the court considered relevant to the curtailment of the right to self-representation.

Furthermore, in Croissant v. Germany, the ECtHR found that “avoiding interruptions or adjournments corresponds to an interest of justice which is relevant in the present context and may well justify an appointment against the accused’s wishes.” This principle is also recognized by U.S. courts which have denied applications “to proceed pro se be-

150. Milosevic v. Prosecutor, Case No. IT-02-54-AR73.7, Prosecution Response to “Assigned Counsel Appeal Against Trial Chamber’s Decision on the Assignment of Defence Counsel” and “Defence Reply to “Prosecution Motion to Strike Ground of Appeal (3) from Assigned Counsel “Appeal Against Trial Chamber’s Decision on the Assignment of Defence Counsel,”” n.98 (Oct. 11, 2004).
151. Prosecutor v. Seselj, IT-03-67-PT, Decision on Prosecution’s Motion for Order Appointing Counsel to Assist Vojislav Seselj with His Defence, para. 21 (May 9, 2003).
152. Prosecutor v. Delalic, Case No. IT-96-21-T, Decision on the Alternative Request for Renewed Consideration of Delalic’s Motion for an Adjournment Until 22 June or Request for Issue of Subpoenas to Individuals and Requests for Assistance to the Government of Bosnia and Herzegovina, paras. 44–45 (June 22, 1998) (noting that the exercise of the rights contained in Article 21(4) “is subject to the control of the Trial Chamber to ensure a fair and expeditious trial in the interests of justice” and declaring that “the Chamber cannot wait until foreseeable harm is done to the proceedings. It is for the Chamber to prevent such foreseeable harm.”).
153. Prosecutor v. Prlic, Case No. IT-04-74-PT, Decision on Requests for Appointment of Counsel, para. 31 (July 30, 2004) (citing Prosecutor v. Hadzihasanovic, Case No. IT-01-47-PT, Decision on the Prosecution’s Motion for Review of the Decision of the Registrar to Assign Mr. Rodney Dixon as Co-Counsel to the Accused Kubura, para. 44–45 (Mar. 26, 2002), (noting that “the Chamber cannot wait until foreseeable harm is done to the proceedings. It is for the Chamber to prevent such foreseeable harm.”)).
154. Prosecutor v. Norman, Case No. SCSL-04-14-T, Decision on the Application of Samuel Hinga Norman for Self Representation Under Article 17(4)(d) of the Statute of the Special Court, paras. 15–16 (June 8, 2004). The SCSL further considered that the Trial Chamber has a special interest in “protect[ing] the integrity of [its] proceedings” and “ensur[ing] that the administration of justice is not brought into disrepute.” Id. para. 28.
cause of concern for orderly process.”

Further justifications for denying a pro se request have included “the need to ‘minimize disruptions, to avoid inconvenience and delay, [and] to maintain continuity.'”

3.1.1: The Willful Nature of Obstructionism

An ICTY Trial Chamber in Delalic confirmed the willful nature of obstructionism, noting that self-representation may be curtailed where an accused “unreasonably and unilaterally chooses his own dates in such a manner as to prejudicially affect the course of the proceedings and cause delay in respect of the defence of other accused persons.” This conclusion was primarily inspired by the possibility that “an accused . . . may by devious reasons relying on Article 21(4)(e) prolong the trial unnecessarily.” As in Faretta, the rationale underpinning this decision pertains solely to voluntary misconduct.

In a dissenting opinion in Seselj, Judge Antonetti stated that a mere intention to obstruct proceedings is insufficient to justify the curtailment of the right to self-representation. In order to warrant the appointment of counsel, the Tribunal must demonstrate deliberately obstructionist behavior or indisputably extreme conduct by the defendant which, by its

156. Colquitt, supra note 36, at 64 (citing Sapienza v. Vincent, 534 F.2d 1007, 1010 (2d Cir. 1976); People v. Anderson, 247 N.W.2d 857, 860 (Mich. 1976)).

157. Id. at 65 (quoting United States v. Dunlap, 577 F.2d 867, 868 (4th Cir. 1978)); see also Martinez v. Court of Appeal of Cal., 528 U.S. 152, 161–62 (2000) (“[T]he government’s interest in ensuring the integrity and efficiency of the trial at times outweighs the defendant’s interest in acting as his own lawyer.”).

158. Prosecutor v. Delalic, Case No. IT-96-21-T, Decision on the Alternative Request for Renewed Consideration of Delalic’s Motion for an Adjournment Until 22 June or Request for Issue of Subpoenas to Individuals and Requests for Assistance to the Government of Bosnia and Herzegovina, para. 45 (June 22, 1998).

159. Id. Based on these reasons, the Delalic Tribunal ordered the pro se Accused to close his case. Id. para. 48.

160. Prosecutor v. Seselj, Case No. IT-03-67-PT, Decision on the Accused’s Motion to Re-Examine the Decision to Assign Standby Counsel, Opinion Dissidente du Judge Antonetti [Dissenting Opinion of Judge Antonetti], para. 10 (Mar. 1, 2005) (“La Chambre ne peut pas . . . limiter le droit de l’Accusé à assurer personnellement sa défense en se fondant sur des « intentions » obstructionnistes.” [“The Chamber may not limit . . . the right of the Defendant to personally ensure his defense because it is based on obstructionist intentions.”] (translation by Brooklyn Journal of International Law)).

161. Id. (noting that the Tribunal “a le devoir de démontrer que le comportement de l’accusé est constitutif d’une faute témoignant d’un comportement délibérément grave et obstructionniste” [“has the duty to show that the behavior of the defendant constitutes an offense demonstrating deliberately serious and obstructionist conduct”] (translation by Brooklyn Journal of International Law)).
very nature, encumbers the administration of justice.\(^{162}\) Clearly, therefore, in order to constitute obstructionism, disruption must be derived from voluntary misconduct.

The Appeals Chamber in Milosevic found it “particularly instructive”\(^{163}\) to consider the right of an accused “to be tried in his [own] presence.”\(^{164}\) Rule 80(B) of the ICTY Rules allows a Trial Chamber to “order the removal of an accused from the courtroom and continue the proceedings in the absence of the accused if the accused has persisted in disruptive conduct.”\(^{165}\) Here, again, it is voluntary misconduct which triggers the curtailment of the minimum guarantees set out in Article 21(4)(d).

3.1.2: Was Milosevic an Obstructionist?

The Milosevic Trial Chamber explicitly acknowledged that the Accused’s conduct at trial did not constitute obstructionism.\(^{166}\) The Prosecution also overtly accepted this fact.\(^{167}\) In order to understand Milosevíc’s conduct, it is first necessary to comprehend his defense strategy. Milosevic clearly viewed himself as a “political” defendant, attempting to convert his trial into a trial of NATO and the Clinton Administration, which, he claimed, cooperated with “terrorists” in Kosovo in 1999.\(^{168}\)

\(^{162}\) Id. (noting that the Tribunal must establish “un comportement manifestement excessif de nature à entraver la bonne administration de la justice” (“conduct so manifestly excessive the nature of which is likely to hinder the proper administration of justice”)(translation by Brooklyn Journal of International Law)).

\(^{163}\) Milosevic v. Prosecutor, Case No. IT-02-54-AR73.7, Decision on Interlocutory Appeal of the Trial Chamber’s Decision on the Assignment of Defense Counsel, para. 13 (Nov. 1, 2004).

\(^{164}\) Id.; see ICTY Statute, supra note 12, art. 21(4)(d).


\(^{166}\) Prosecutor v. Milosevic, Reasons for Decision on the Prosecution Motion Concerning Assignment of Counsel, para. 40 (Apr. 4, 2003).

\(^{167}\) Prosecutor v. Seselj, IT-03-67-PT, Prosecution’s Motion for Order Appointing Counsel to Assist Vojislav Seselj with his Defence, at para. 11 (Feb. 28, 2003) (“As a former head of state, the accused Milosevic does not need to be disruptive, obstructionist or scandalous in order to remain in the public’s eye. Therefore, despite his rejection of the Tribunal and its authority as such, to date the accused Milosevic has, to a large extent, taken part in the proceedings in an orderly fashion.”) (on file with the author).

\(^{168}\) Sophia Piliouras, U.N. Report, International Criminal Tribunal for the Former Yugoslavia and Milosevic’s Trial, 18 N.Y.L. SCH. J. HUM. RTS. 515, 520–21 (2002) (alleging, inter alia, that the United States provided “close air support to the Kosovo Liberation Army and other Albanian ‘terrorists’ and their foreign Islamic allies,” that NATO
Among other allegations, Milosevic accused the ICTY of “victor’s justice.” Milosevic claimed that the Tribunal yields to the objectives of the U.S. and other NATO powers, countries without whose financial and military support the tribunal could not function. Essentially, he chose self-representation as the most effective means of defending the actions of the Serbian nation and his own political record.

Commentators have noted that Milosevic was derisive on certain occasions during the trial proceedings. This, however, is inherent in the primarily adversarial nature of the trial proceedings. It also ignores the sensitive manner in which the Accused dealt with alleged victims of sexual assault. Accusations that the Accused refused to focus on material considered relevant to the indictment fail to examine the role of the Trial Chamber in regulating the presentation of the evidence, and the efforts of the Accused to present indictment-relevant evidence, particularly relating to Kosovo, during the Defense stage.

Although the Accused did not personally file written submissions and at times failed to state his position on procedural points, he effectively

---

169. Among other factors challenged by Milosevic were the composition of the bench, the manner of his surrender to the Tribunal, the timing of his indictment, and the unfairness of procedures at the Tribunal. Michael P. Scharf, *The Legacy of the Milosevic Trial*, 37 NEW ENG. L. REV. 915, 923–30 (2003).


171. *See* id.

172. *See*, e.g., Madeleine K. Albright, *We Won’t Let War Criminals Walk: With or Without a Balkan Peace Deal, the U.S. Won’t Relent*, WASH. POST, Nov. 19 1995, at C01 (noting that the U.S. had given “more than $13 million in direct contributions and assessments” to the Tribunal).

173. *See*, e.g., Scharf, *supra* note 169, at 919 (“In addition to regularly making disparaging remarks about the court and repeatedly brow-beating witnesses, Milosevic pontificated[d] at length during cross examination of every prosecution witness.”).

174. *See*, e.g., Trial Tr. at 3264:15–17 (Apr. 17, 2002), Prosecutor v. Milosevic, Case No. IT-02-54-T, available at http://www.un.org/icty/trans54/020417IT.htm (May, J.) (“It is an abuse of the process for you to make speeches, Mr. Milosevic, at this stage. It’s also an abuse to go over the same ground.”); id. at 6208:8–10 (June 4, 2002), available at http://www.un.org/icty/trans54/020604ED.htm (May, J.) (“A great deal of time is taken up with repetition and argument and sometimes irrelevancies.”).

175. *See*, e.g., id. at 32078:3–4 (Mar. 25, 2004), available at http://www.un.org/icty/trans54/040325MH.htm (Milosevic: “Of course I have no intention of declaring my views on your administrative issues.”); id. at 32078:9–13 (Meron, J.) (“I am assuming that I should derive the conclusion from your comment that you do not wish to grant a consent to the continuation of the hearings as I cannot understand the comment you have made as amounting to a clear and unequivocal consent to continue the proceedings with a substitute Judge.”).
waived his right to file by implicitly accepting the work of Assigned Counsel. The Accused essentially prompted them to make such submissions on his behalf through oral submissions at trial.\footnote{176} He went so far as to explicitly acknowledge the dedication and commitment of Assigned Counsel.\footnote{177} However, it is evident from the litigation concerning provisional release that he gave explicit and direct instructions to Assigned Counsel from December 2005 to the time of his death.\footnote{178} Furthermore, Milosevic “fully engaged in the cross-examination of Prosecution witnesses”\footnote{179} and presented consecutive witnesses in his defense,\footnote{180} thereby actively partaking in trial proceedings.

3.1.3: Threshold of Obstructionist Conduct

The high threshold that a defendant’s conduct must breach in order to constitute obstructionism is manifest from the ICTR case of Barayagwiza.\footnote{181} The Accused, Jean-Bosco Barayagwiza, declared the Rwanda Tribunal to be “so dependent on the dictatorial anti-Hutu regime of Kigali” that it could not render “independent and impartial justice.”\footnote{182} He

\footnotetext[176]{E.g., Prosecutor v. Milosevic, Case No. IT-02-54-T, Decision on Assigned Counsel Request for Provisional Release, para. 2 (Feb. 23, 2006); Prosecutor v. Milosevic, Case No. IT-02-54-AR73.7, Appeal Against the Trial Chamber’s Decision on Assignment of Defence Counsel: Corrigendum, para. 32 (Sept. 29, 2004), available at http://www.un.org/icty/milosevic/motion/040929.htm.}

\footnotetext[177]{See Prosecutor v. Milosevic, Case No. IT-02-54-T, Transcript of Record at 46690:10–16 (Judge Robinson: “Mr. Milosevic, that, again, is a matter in relation to which you owe a great debt of gratitude to assigned counsel. Through their action, through their professionalism, we are considering now a motion to subpoena certain witnesses, and without their intervention, without their help, we would not have been considering this.” The Accused: “Yes, I know about that. I know about that, Mr. Robinson.”).}

\footnotetext[178]{E.g., Prosecutor v. Milosevic, Case No. IT-02-54-T, Decision on Assigned Counsel Request for Provisional Release, para. 2 (Feb. 23, 2006) (indicating that the Accused first requested provisional release orally, which was then followed by Assigned Counsel’s formal written Request for Provisional Release Pursuant to Rule 65).}

\footnotetext[179]{Prosecutor v. Milosevic, Case No. IT-02-54-AR73.7, Prosecution Response to “Assigned Counsel Appeal Against the Trial Chamber’s Decision on Assignment of Defence Counsel” and to “Defence Reply to Prosecution Motion to Strike Ground of Appeal (3) from Assigned Counsel “Appeal Against the Trial Chamber’s Decision on Assignment of Defence Counsel.””, para. 90 n.144 (Oct. 11, 2004).}

\footnotetext[180]{E.g., Trial Tr. at 34851:9–10 (Jan. 11, 2005), Prosecutor v. Milosevic, Case No. IT-02-54-T, available at http://www.un.org/icty/transe54/050111IT.htm.}

\footnotetext[181]{Prosecutor v. Barayagwiza, Case No. ICTR-97-19.}

stated that the Tribunal was committing “a mockery of justice” and instructed his counsel not to represent him in any way. Likewise, Milosevic contested the legitimacy of the ICTY, accusing the Tribunal of being “not [a] juridical institution” but rather “a political tool.” However, parallels between the two cases end there; the voluntarily disruptive conduct of Barayagwiza breached a threshold which Milosevic did not aspire to reach. Barayagwiza refused to attend proceedings, thereby apparently “boycotting” the tribunal. As a result of Barayagwiza’s instructing his lawyers not to represent him, his lawyers remained passive and did not mount an active defense. Additionally, Barayagwiza did not assert his right to self-representation.

By contrast, to quote the prosecution, “despite his rejection of the tribunal and its authority as such . . . the accused Milosevic has, to a large extent, taken part in the proceedings in an orderly fashion.” The Milosevic Trial Chamber explicitly distinguished the conduct of the former President from that of Barayagwiza, noting that “[n]o such circumstances have, as yet, arisen in this trial.” Thus, although certain aspects of the behavior of Milosevic may parallel the conduct of the ICTR-accused, the primary source of the obstructionism in the Barayagwiza case, i.e., the refusal of the accused to attend court, was absent in the case of the former President.

The conduct of ICTY-accused Seselj provides a further example of the high threshold an accused must breach in order to constitute obstructionism. Vojislav Seselj, representing himself, has consistently derogated from the issues at hand and refused to follow the procedural rules of

183. Id. para. 12.
184. Id. para. 11 (citing Letter from Jean-Bosco Barayagwiza to the Trial Chamber (Oct. 23, 2000)).
187. Id. para. 17.
188. Id., Concurring and Separate Opinion of Judge Gunawardana, para. 3.
189. Prosecutor v. Seselj, Case No. IT-03-67-PT, Prosecution’s Motion for Order Appointing Counsel to Assist Vojislav Seselj with his Defence, para. 11 (Feb. 28, 2003) (on file with the author).
190. Id. para. 40.
191. E.g., Prosecutor v. Seselj, Case No. IT-03-67-PT, Decision on Prosecution’s Motion for Order Appointing Counsel to Assist Vojislav Seselj with his Defence, para. 7 (May 9, 2003) (“The Accused devoted only thirteen pages of his ninety-three page ‘Reply to the Prosecutor’s Motion to impose defence counsel on me against my will’ . . . to the concrete legal question actually at issue.”). The Tribunal classified the remainder as “frivolous abuse of the Tribunal’s Translation Unit.” Id. para. 7 n.7.
Seselj consistently filed handwritten, “excessively long,” and “largely irrelevant” motions. He publicly expressed his intent to cause harm to, or indeed to “attack,” and “destroy” the Tribunal, and to use the proceedings to promote Serb national interests rather than as a means to defend himself against the charges alleged against him. Furthermore, he refused to use a laptop or typewriter as “he was ‘afraid of receiving an electric shock,’” and repeatedly insisted that he only understands the Serbian language despite the fact that Croatian is simply a variant of that language and evidence showed that he understands Eng-

192. Id. para. 23 (noting that, inter alia, the Accused “submitted a hand written petition directly to the Appeals Chamber” knowing that this violated the Rules, that a self-described “legal adviser” to the accused submitted a letter to the Registrar “[un]accompanied by the necessary power of attorney” and who apparently did not meet the requirement that counsel speak one of the working languages of the Tribunal, and that the accused, in a petition, made “frivolous demands framed in language inappropriate for a legal document”).

193. Id.

194. In a 1994 interview for the French film “Crimes et Criminels,” when he was asked what would be his defense in the event of a case being taken against him at the ICTY, the Accused responded: “I am not planning to defend myself, I can only attack.” Prosecutor v. Seselj, Case No. IT-03-67-PT, Prosecution’s Motion for Order Appointing Counsel to Assist Vojislav Seselj with his Defence (Feb. 28, 2003) (on file with the author). He also said: “Personally, I do not recognise this Hague Tribunal. I think it has no legal foundation, but if I am ever invited to The Hague I’ll gladly go there immediately. I would never miss such a show.” Prosecutor v. Seselj, Case No. IT-03-67-PT, Decision on Prosecution’s Motion for Order Appointing Counsel to Assist Vojislav Seselj with his Defence, para. 4 n. 5 (May 9, 2003). Furthermore, at a public rally on the evening before he left Belgrade for The Hague, the accused Seselj said “that he would put ‘the Americans, the Hague tribunal and NATO’ on trial.” Sofia Hilden, Serb Hardliner Seselj Arrives in Hague, Reuters, Feb. 24, 2003. The Accused also made the following comments: “With their stupid charges against me they have come up against the greatest living legal Serb mind . . . . I shall blast them to pieces.” Seselj Dismisses Hague Indictment, ONASA News Agency, Feb. 17, 2003.

195. On February 3, 2003, the news agency Deutsche Presse Agentur reported that the Accused was “reported to have said that ‘he would gladly travel to The Hague to ‘destroy’ the . . . tribunal in case it open[ed] a trial against him.’” Prosecutor v. Seselj, Case No. IT-03-67-PT, Decision on Prosecution’s Motion for Order Appointing Counsel to Assist Vojislav Seselj with his Defence, para. 4 n. 5 (May 9, 2003) (quoting DEUTSCHE PRESSE-AGENTUR, Feb. 3, 2003).

196. See Prosecutor v. Seselj, Case No. IT-03-67-PT, Prosecution’s Motion for Order Appointing Counsel to Assist Vojislav Seselj with his Defence, para. 1 (Feb. 28, 2003) (on file with the author).

197. Prosecutor v. Seselj, Case No. IT-03-67-PT, Trial Tr. at 66, available at http://www.un.org/icty/transes67/030325SSC.htm; Seselj, Decision on Prosecution’s Motion for Order Appointing Counsel to Assist Vojislav Seselj with his Defence, para. 23 (May 9, 2003).
lish. The trial chamber found the foregoing “attitude and actions . . . indicative of obstructionism on [Seselj’s] part.” As recognized by the Prosecution, the similarities between Seselj and Milosevic are remote.

Although Milosevic used the Serbian form of address rather than that customarily used at the Tribunal, Milosevic remained moderately respectful to Tribunal judges. In the highly publicized case of United States v. Zacarias Moussaoui, the right to proceed pro se was revoked where the accused filed frivolous motions which repeatedly insulted the judge, among others. This misconduct indicates a lower threshold than

---

198. Prosecutor v. Seselj, Case No. IT-03-67-PT, Decision on Prosecution’s Motion for Order Appointing Counsel to Assist Vojislav Seselj with his Defence, paras. 24–25 (May 9, 2003).

199. Id. para. 26.

200. Prosecutor v. Seselj, Case No. IT-03-67-PT, Prosecution’s Motion for Order Appointing Counsel to Assist Vojislav Seselj with his Defence, para. 11 (Feb. 28, 2003) (on file with the author). The Prosecution stated that:

As a former head of state, the accused Milosevic does not need to be disruptive, obstructionist or scandalous in order to remain in the public’s eye. Therefore, despite his rejection of the Tribunal and its authority as such, to date the accused Milosevic has, to a large extent, taken part in the proceedings in an orderly fashion. The opposite is to be expected from the accused Seselj. The accused Seselj thrives on the creation of the scandal, conspiracies and publicity.

Id.; Seselj, Decision on Prosecution’s Motion for Order Appointing Counsel to Assist Vojislav Seselj with his Defence, para. 4 (May 9, 2003). The ICTY stated that:

The Prosecution argues that the circumstances of the Accused’s request to represent himself are distinguishable from those of Slobodan Milosevic who is currently the only other accused before this Tribunal conducting his own defence. The Prosecution also requested the imposition of defence counsel in the Milosevic case, but in that case the main reason was concern about the toll that self-representation was taking on Milosevic’s health. In contrast to Milosevic, the Prosecution believes the Accused Seselj has intimated to and may behave in a ‘disruptive, obstructionist or scandalous’ manner.

Id.

201. Molly Moore, Trial of Milosevic Holds Lessons for Iraqi Prosecutors, WASH. POST, Oct. 18, 2005, at A19. Theodor Meron, the president of the ICTY, remarked, “[Milosevic] complies with the rules of the game for the most part. If he insists on calling the judges ‘Mister’ instead of ‘Your Honor,’ I regret that. But it doesn’t mean he’s not otherwise respectful to the judges.” Id.

the Barayagwiza or Seselj standards, but remains inapplicable to Milosevic.

There appears to be a lack of consensus with regard to the minimum threshold to be met to constitute obstructionism in domestic jurisdictions. In the U.S. case of United States v. Davis, counsel was assigned as the accused delved into obscure and irrelevant discussion whenever he was afforded the opportunity to speak. The trial in Duke v. United States was infused with accusations that the prosecution was the result of a conspiracy to frame the accused. The court indicated that the accused could not be allowed to turn the proceeding into a “Roman holiday” or use his pro se status “as a launching ground for missiles, even if . . . he believed his best defense was . . . trying the prosecutors and witnesses for the prosecution.” Although this displays certain comparisons to the Milosevic case, it is arguable that the level of deliberate misconduct in these cases exceeds that of the former President.

Dissentient Justice Blackmun in Faretta v. California suggested that the presentation of a politically inspired defense may defeat the right to self-representation, proposing that the right does not accommodate “the whimsical—albeit voluntary—caprice of every accused who wishes to use his trial as a vehicle for personal or political self-gratification.” Similarly, in United States v. Frazier-El, the defendant’s insistence on representing himself so that he could present “frivolous” arguments in his defense was sufficient to deny him his right to proceed pro se.

This premise stretches the theory of obstructionism to the extent that it excessively restricts the accused’s defense rights. According to Judge Antonetti’s dissenting opinion in Seselj, Faretta established that a judge may discontinue an accused’s self-representation if he or she “deliberately adopts behavior that is serious and obstructionist.” According to

204. Id. at 1021. The obstructionism in Davis was due to the defendant’s mental incapacities. The court found that Davis lacked sufficient mental capacity to represent himself. Id.
205. 255 F.2d 721 (9th Cir. 1958).
206. Id. at 726–27. The accused alleged that a federal judge and members of the U.S. Attorney’s staff were among those involved in framing him. Id. at 727.
207. Id.
210. Prosecutor v. Seselj, Case No. IT-03-67-PT, Decision on the Accused’s Motion to Re-examine the Decision to Assign Standby Counsel, Opinion Dissidente du Juge Antonetti, para. 8 (Mar. 1, 2005) (“[S]i l’accusé adopte délibérément un comportement grave et obstructioniste.” “[I]f the Accused deliberately adopts behavior that is serious and obstructionist.”) (citing Faretta, 422 U.S. at 834 n.46 (author’s translation)).
Judge Antonetti, the accused must seriously interfere with proceedings, repeatedly breach the orders and decisions of the Trial Chamber, interrupt proceedings through bad behavior, or use profanities. Milosevic did not come close to breaching this threshold of deliberate obstructionism.

3.2: Unintentional Disruption

While it is clear that while a deliberately disruptive defendant reaching a certain level of misconduct may be denied the pro se right, “a more complex problem arises when the pro se defendant unintentionally disrupts the proceedings but not to an extent that would justify his permanent removal from the courtroom.” At the outset, it is important to note that although a pro se defendant will inevitably take longer to present his defense than experienced counsel will, “a simple lack of legal knowledge may not be the . . . sole reason for denying a pro se request.” This principle is self-evident, because if a court required a de-

---

211. Id. para. 8 (“[L]e droit de l’accusé à se défendre lui-même peut être restreint au motif que l’accusé perturbe gravement le procès.” “[T]he right of the accused to defend himself may be limited on grounds that the accused seriously disrupts the proceeding.”) (emphasis added) (translation by Brooklyn Journal of International Law).

212. Id. para. 10. Judge Antonetti stated that

La notion d’obstruction à la justice et au bon déroulement du procès doit en effet s’entendre de la violation répétée des ordres et décisions de la Chambre, de la perturbation du déroulement du procès par des écarts de conduite, de l’utilisation d’un langage outrageant ou de toute autre faute témoignant d’un comportement délibérément grave et obstructionniste. La Chambre ne peut pas à ce stade limiter le droit de l’Accusé à assurer personnellement sa défense en se fondant sur des « intentions » obstructionnistes.”

[The notion of the obstruction of justice and the proper conduct of proceedings must involve repeated violations of order and decisions of the Chamber, disturbances of the proceedings through bad conduct, the use of insulting language or other conduct exhibiting deliberate, serious and obstructionist behaviour. The Chamber cannot, at this stage, restrict the right of the accused to present his own defence on the basis of obstructionist “intentions.”]

Id. (author’s translation).

213. Homiak, supra note 5, at 920.

214. See United States v. Dougherty, 473 F.2d 1113, 1124 (D.C. Cir. 1972) (“Given the general likelihood that pro se defendants have only rudimentary acquaintanceship with the rules of evidence and courtroom protocol, a measure of unorthodoxy, confusion and delay is likely, perhaps inevitable, in pro se cases.”) (footnote omitted)).

215. Homiak, supra note 5, at 910; see also Faretta v. California, 422 U.S. 806, 835 (1975) (“[A] defendant need not himself have the skill and experience of a lawyer in order competently and intelligently to choose self-representation . . . .”)
fendant to possess legal knowledge as a pre-requisite to granting the right to represent himself, the right to self-representation “would become essentially meaningless.” 216 In other words, an accused may elect to represent himself even if his lack of experience causes significant inefficiency in the proceedings. 217

What is contentious in the context of the Milosevic case is the decision of the Trial Chamber to base its decision to assign counsel on the consequences of the recurring ill health of the Accused. It is the contention of this paper that this innovative expansion of the circumstances in which the pro se right may be restricted, without the circumscription necessary to confine the decision to the specific circumstances of the Milosevic case, leaves open the possibility that the right to self-representation and other minimum guarantees which are composite elements of the right to a fair trial will in the future be abrogated.

3.2.1: Trial Chamber Decision

As seen in Section 1.3.1, the recurrent ill health of Milosevic inhibited the expeditious completion of the case. 218 The Trial Chamber indicated that medical reports filed in July and August 2004 made it “plain . . . that the Accused is not fit enough to defend himself and that should he continue to represent himself, there will be further delays in the progress of the trial.” 219 Doctors advised that Milosevic could suffer “a hypertensive emergency, a potentially life-threatening condition.” 220 On this basis, the Tribunal felt obliged to assign counsel in the interest of the orderly administration of justice. 221

The Trial Chamber conceded that no precedent existed for this reasoning. 222 Nevertheless, it promulgated the theory that “[d]isruption of a trial, whatever the circumstances, may give rise to the risk of a miscarriage of justice because the whole proceedings have not been conducted

216. Homiak, supra note 5, at 910.
218. Trial Tr. at 32357:9–10 (Sept. 2, 2004), Prosecutor v. Milosevic, Case No. IT-02-54-T, available at http://www.un.org/icty/tras54/040920IT.htm (“The health of the accused has been a major problem in the progress of the trial.”).
219. Id. at 32358:4–6.
220. Id. at 32357:24–25.
221. Prosecutor v. Milosevic, Case No. IT-02-54-T, Reasons for Decision on Assignment of Defence Counsel, para. 65 (Sept. 22, 2004) (citing the concern that the trial could “last for an unreasonably long time” and noting the Tribunal’s “duty to ensure a fair and expeditious trial and its responsibility to preserve the integrity of its proceedings”).
222. Id. para. 37 (“[E]xtensive research has not led to the identification of any case in any jurisdiction where counsel has been assigned to an accused person because he was unfit to conduct his case as the result of impaired physical health . . . .”).
and concluded fairly.”

This decision was based on an examination of the jurisprudence of the ECtHR and U.N. Human Rights Committee (HRC) as well the jurisprudence of the ICTY, ICTR, and SCSL. In contrast to its 2003 ruling, the Trial Chamber sought to demonstrate that the various adjudicative bodies recognized a potentially broad range of possible exceptions to the right to self-representation.

Furthermore, in its 2004 ruling, the Trial Chamber relied heavily on the ECtHR decision in *Croissant* despite having found in 2003 that case to be “distinguishable from the instant case” to the extent that it involved objection to appointment of additional counsel. It also distinguished from the case at hand the views of the HRC in *Brian & Michael Hill v. Spain*, despite having found this decision to be “highly relevant to the correct interpretation of Article 21(4) (d)” in its 2003 ruling. However, in its 2003 ruling, ignoring jurisprudence to the contrary, the Trial Chamber appeared to interpret the HRC’s views as prohibiting exceptions to the pro se right.

---

223. Id. para. 33.
224. Id. paras. 38–44.
225. Id.
226. Id. para. 43.
228. Hill v. Spain, Commc’n No. 526/1993, U.N. Doc. CCPR/C/59/D/526/1993, para. 14.2 (June 23, 1997). This case concerned a Spanish law which specified that although an accused had the right to defend himself, “such defence should take place by competent counsel, paid by the State when necessary.” Id. The HRC found, without further discussion, that the right to defend oneself under Article 14(3)(d) of the ICCPR had not been respected. Id.
231. Prosecutor v. Milosevic, Case No. IT-02-54-T, Reasons for Decision on the Prosecution Motion Concerning Assignment of Counsel, para. 36 (Apr. 4, 2003) (“[T]he only case on the issue decided under these conventions which the Trial Chamber has been
Side-stepping the inaccuracy of its earlier assertions, the Trial Chamber in its 2004 ruling disregarded the HRC’s views on the basis that the Committee had given no reason for its determination and that it was not faced with circumstances which could be compared to the Milosevic case.\textsuperscript{232} Further, although it had distinguished Barayagwiza (ICTR) in its 2003 ruling, the Trial Chamber regarded this case\textsuperscript{233} as well as Norman (SCSL)\textsuperscript{234} and the interim case of Seselj (ICTY)\textsuperscript{235} to be authority for the fact that “such factors as the ability of the accused to conduct his own defence, as well as his attitude and actions” should be taken into account when determining what course might be appropriate in the circumstances.\textsuperscript{236} This assertion markedly broadens the principles elucidated in these judgments.

3.2.2: Appeals Chamber Decision

Despite these inconsistencies in the Trial Chamber’s reasoning, the Appeals Chamber upheld its exercise of discretion in assigning counsel, reaffirming that “it cannot be that the only kind of disruption legitimately cognizable by a Trial Chamber is the intentional variety.”\textsuperscript{237} The reasoning upon which the Appeals Chamber based this premise was no more consistent than that of the Trial Chamber. In the form of a rather puzzling rhetorical question, the Appeals Chamber, in its decision, asked:

How should the Tribunal treat a defendant whose health, while good enough to engage in the ordinary and non-strenuous activities of everyday life, is not sufficiently robust to withstand all the rigors of trial

\textsuperscript{232} Prosecutor v. Milosevic, Case No. IT-02-54-T, Reasons for Decision on Assignment of Defence Counsel, para. 44 (Sept. 22, 2004).

\textsuperscript{233} Prosecutor v. Barayagwiza, Case No. ICTR-97-19-T, Decision on Defence Counsel Motion to Withdraw, para. 2; Concurring and Separate Opinion of Judge Gunawardana (Nov. 2, 2000).

\textsuperscript{234} Prosecutor v. Norman, Case No. SCSL-04-14-T, Decision on the Application of Samuel Hinga Norman for Self Representation Under Article 17(4)(d) of the Statute of the Special Court, paras. 8, 27 (June 8, 2004).

\textsuperscript{235} Prosecutor v. Seselj, IT-03-67-PT, Decision on Prosecution’s Motion for Order Appointing Counsel to Assist Vojislav Seselj with His Defence, paras. 20, 27, 30; Disposition at 13 (May 9, 2003).

\textsuperscript{236} Prosecutor v. Milosevic, Case No. IT-02-54-T, Reasons for Decision on Assignment of Defence Counsel, para. 38 (Sept. 22, 2004); \textit{Ibid}. paras. 39–41 (discussing the aforementioned cases).

\textsuperscript{237} Prosecutor v. Milosevic, Case No. IT-02-54-AR73.7, Decision on Interlocutory Appeal of the Trial Chamber’s Decision on the Assignment of Defense Counsel, para. 14 (Nov. 1, 2004).
work—the late nights, the stressful cross-examinations, the courtroom confrontations—unless the hearing schedule is reduced to one day a week, or even one day a month?\textsuperscript{238}

Without giving a substantive answer to this monologue, the Appeals Chamber went on to question whether the Tribunal ought to “be forced to choose between setting [such a] . . . defendant free and allowing the case to grind to an effective halt?”\textsuperscript{239} Giving the impression that the solution was self-evident, the Appeals Chamber responded ambiguously that “to ask that question is to answer it.”\textsuperscript{240} No further scrutiny was accorded to the question of non-intentional disruption.

The manner in which the concept of non-intentional disruption was promulgated is wholly unsatisfactory. Neither the Appeals Chamber nor the Trial Chamber attempted to provide any guidance as to the degree of ill health necessary in order for the tribunal to acquire the discretion to limit the right to self-representation. The Tribunal thereby failed to specify to what extent and for how long the ill health of the accused must delay proceedings in order for the tribunal to be able to exercise this discretion. The necessary cause, nature, and potential duration of such ill health remain undetermined.

It also remains unclear how “healthy” participants must be in order to qualify for pro se status before the Tribunal, or how “unhealthy” they must become in order for this minimum guarantee to be denied. Under the precedent established by the Appeals Chamber, it is to be expected that counsel might be assigned to every defendant, like Milosevic, with high blood pressure or potential heart problems. How much further this precedent stretches, however, remains a question of speculation. It is not implausible to surmise that any defendant with any potential health problems, no matter how trivial, could be denied the right to represent themselves in person. This possibility creates obvious anomalies, particularly given that the defense rights of the accused set out in Article 21(4) are guaranteed “in fully equality.”\textsuperscript{241}

Arguably, the Tribunal deliberately phrased its reasoning in ambiguous terms to avoid drawing attention to the novelty of its contention, the inconsistency in the reasoning of the chambers, and the lack of precedent to support it. As it stands, given the lack of definitive guidelines provided by the Appeals Chamber, case-by-case analysis will be required to determine which circumstances will merit curtailing the right to self-

\textsuperscript{238} Id.
\textsuperscript{239} Id.
\textsuperscript{240} Id. (emphasis added).
\textsuperscript{241} ICTY Statute, supra note 12, art. 21(4).
representation where non-intentional delay occurs. This leaves the future of self-representation, as well as the other minimum guarantees in Article 21(4) of the Statute, in unclear territory. Such lack of clarity critically undermines the significance of the right to self-representation and compromises the possibility of achieving a fair trial.

In order to estimate the potential parameters of non-intentional disruption, it may be useful to examine the cases cited by the Appeals Chamber.\textsuperscript{242} The U.S. case of \textit{Johnson v. State}\textsuperscript{243} provides authority for the withdrawal of the right to self-representation where, as in \textit{Johnson}, expert testimony indicates that defendant’s fragile mental state “might well succumb to the added stress of self-representation and deteriorate to the point where it would become necessary to continuously disrupt the proceedings to monitor his ‘present’ mental abilities and competence.”\textsuperscript{244} This suggests that the potential ill health to which the concept of non-intentional disruption applies appears to encompass both physical and mental health. The \textit{Johnson} concept cannot, however, be equated to \textit{Milosevic}, as the capacity of the accused to stand trial was one of the matters at issue in \textit{Johnson}.\textsuperscript{245} This position implies a level of trial disruption that far exceeds the \textit{Milosevic} situation.\textsuperscript{246} It may, perhaps, be surmised that where mental health is at issue, a rather insidious level of disruption is necessary.

The \textit{Milosevic} Appeals Chamber also cited the U.S. case of \textit{Savage v. Estelle},\textsuperscript{247} where the trial court did not allow the pro se defendant to con-

\textsuperscript{242} Milosevic v. Prosecutor, Case No. IT-02-54-AR73.7, Decision on Interlocutory Appeal of the Trial Chamber’s Decision on the Assignment of Defense Counsel, para. 14 n.43 (Nov. 1, 2004).


\textsuperscript{244} \textit{Id.}, 2001 Nev. LEXIS at **28.

\textsuperscript{245} \textit{Id.} (generally).

\textsuperscript{246} See \textit{Prosecutor v. Milosevic}, Case No. IT-02-54-AR73.7, Prosecution Response to “Assigned Counsel Appeal Against the Trial Chamber’s Decision on Assignment of Defense Counsel” and to “Defence Reply to “Prosecution Motion to Strike Ground of Appeal (3) from Assigned Counsel “Appeal Against the Trial Chamber’s Decision on Assignment of Defence Counsel,”” para. 54 (Oct. 11, 2004) (citing Prosecutor v. Strugar, Case No. IT-01-42-T, Decision Re the Defence Motion to Terminate Proceedings (May 26, 2004) (“[T]here is no question that the Accused is unfit to stand trial: he clearly meets all the tests for fitness listed in the test recently formulated in the Strugar case.”); see also Milosevic v. \textit{Prosecutor}, Case No. IT-02-54-AR73.7, Decision on Interlocutory Appeal of the Trial Chamber’s Decision on the Assignment of Defense Counsel, para. 14 n.42 (Nov. 1, 2004) (“We reject Assigned Counsel’s contentions that Milosevic’s inability to represent himself necessarily rendered him unfit to stand trial at all. Trial litigation is an extraordinarily demanding profession. It cannot be that only those defendants capable of meeting its demands are formally fit to stand trial.”).

\textsuperscript{247} 924 F.2d 1459 (9th Cir. 1990).
duct voir dire, cross-examine witnesses, make objections, or argue before the jury without leave of court because of his incapacitating stutter.\(^{248}\)

This case immeasurably broadens the range of situations to which the Milosevic precedent may apply. A stutter is simply a speech impediment which relates in no way to the health of the accused, or his capacity to attend proceedings or stand trial. It certainly cannot be classified as “behavior.” Accordingly, given that the parameters of the exception recognized in Milosevic remain undefined, the potential breadth of conditions to which the decision could apply is unsettling.

Curiously, neither the Appeals Chamber nor the Trial Chamber referred to the proposition in Strugar that “[i]n some cases legal assistance to an accused may be a sufficient measure to compensate for any limitations of capacity of the accused to stand trial.”\(^{249}\) While this proposition initially appears to substantiate Milosevic, Strugar in fact concerned a situation in which the defense claimed that the accused was unable to stand trial, a stage to which it was not foreseen that the Milosevic proceedings would degenerate.\(^{250}\) Hence, the Strugar proposition, rather than substantiating the Milosevic decision, highlights that the assignment of counsel is appropriate only where an extremely prohibitive level of disruption has occurred, i.e., to the extent that proceedings are in danger of completely shutting down.

Furthermore, in Strugar, the provision of legal assistance appears to have been considered suitable where the individual’s incapacity to stand trial was “of such a nature and effect that measures can be taken to sufficiently alleviate the impairment, or its effect, so that the trial can con-

\(^{248}\) Id. at 1461.

\(^{249}\) Prosecutor v. Strugar, Case No. IT-01-42-T, Decision Re The Defence Motion to Terminate Proceedings, para. 39 (May 26, 2004), available at http://www.un.org/icty/strugar/trialc1/decision-e/040526.pdf. The Prosecution proposed that this decision “at a minimum leaves open the possibility, and at a maximum supports the contention, that ill-health can justify the assignment of legal assistance in order to enable a trial to continue.”

The Tribunal suggested that if a defendant’s unfitness is temporary, then an appropriate remedy may be to adjourn until “the accused has sufficiently recovered.” It further stated, however, that “[o]ther cases may require that a trial be abandoned,” meaning presumably that permanent unfitness may lead to termination of the proceedings.

Thus, the measures suggested in Strugar were structured in a hierarchy whereby the measure adopted must be suitable to meet the needs of the circumstances of the case, or, in other words, proportionate to the interest pursued. Section Four will examine the disproportionality of the Appeals Chamber decision in Milosevic and its effect on the dignity and autonomy of the Accused.

SECTION FOUR: PROPORTIONALITY

4.1: Mode of Representation Assigned: Proportionate to Aim Pursued?

As seen in Section 1.3.2, instead of merely appointing counsel to assist Milosevic with the presentation of his defense, counsel was initially assigned to fully represent him. This decision purported to allow the Accused to continue to “actively participate along with counsel in the preparation and presentation of his case.”

His participation was, however, “secondary to that of Assigned Counsel and strictly contingent on the discretionary permission of the Trial Chamber in any given instance.” Thus, the burden of the defense was placed on the Assigned Counsel. This situation may have been based on the supposition expressed in Parren v. State that “[t]here can be but one captain of the ship.”

Subsequent to the assignment of counsel, Milosevic categorically refused to question witnesses or cooperate with Assigned Counsel. This

---

253. Id.
254. Id.
255. Prosecutor v. Milosevic, Case No. IT-02-54-T, Reasons for Decision on Assignment of Defence Counsel, para. 68 (Sept. 22, 2004) (noting that the decision “does not deprive him of his right to speak either by giving evidence, examining and re-examining witnesses as permitted by the Chamber, selecting and submitting documentary evidence, and making final submissions on the evidence”).
256. Milosevic v. Prosecutor, Case No. IT-02-54-AR73.7, Decision on Interlocutory Appeal of the Trial Chamber’s Decision on the Assignment of Defense Counsel, para. 7 (Nov. 1, 2004).
258. See Trial Tr. at 32348:2–3 (Sept. 1, 2004), Prosecutor v. Milosevic, Case No. IT-02-54-T, (Milosevic: “I do not accept any decrease of that right or any renouncing of that
mutinous situation appears to substantiate the theory propounded by the U.S. Supreme Court in *Faretta v. California* that as “unwanted counsel ‘represents’ the defendant only through a tenuous and unacceptable legal fiction,” in this situation “counsel is not an assistant, but a master.”

Ironically, this theory was concurrently recognized by the Trial Chamber as “the classic statement of the right to self-representation.”

4.1.1: Proportionality

While the Milosevic Appeals Chamber upheld the Trial Chamber’s decision to assign counsel, it acknowledged that the Trial Chamber had, by relegating Milosevic to a “visibly second-tier role in the trial,” sharply restricted the Accused’s “ability to participate in the conduct of his case in any way.” The Appeals Chamber reprimanded the Trial Chamber for not recognizing that restrictions on the pro se right “must be limited to the minimum extent necessary to protect the Tribunal’s interest in assuring a reasonably expeditious trial.” This failure, it felt, was “a fundamental error of law.”

Various jurisdictions utilize a “basic proportionality principle” in considering limitations on fundamental rights, which is that “any restriction of a fundamental right must be in service of ‘a sufficiently important objective,’ and must ‘impair the right . . . no more than is necessary to accomplish the objective.’” The ICTY has noted that “[a] measure in public international law is proportional only when it is (1) suitable, (2) necessary and when (3) its degree and scope remain in a reasonable rela-

261. Milosevic v. Prosecutor, Case No. IT-02-54-AR73.7, Decision on Interlocutory Appeal of the Trial Chamber’s Decision on the Assignment of Defense Counsel, para. 16 (Nov. 1, 2004).
262. Milosevic v. Prosecutor, Case No. IT-02-54-AR73.7, Decision on Interlocutory Appeal of the Trial Chamber’s Decision on the Assignment of Defense Counsel, para. 17 (Nov. 1, 2004).
263. *Id*.
tionship to the envisaged target.”

The U.N. HRC has similarly observed that restrictions on the right to freedom of movement, for example, “must be the least intrusive instrument amongst those which might achieve the desired result; and they must be proportionate to the interest to be protected.”

The ECtHR has also held that “there must be a reasonable relationship of proportionality between the means employed and the aim pursued.”

According to this principle, Assigned Counsel may “do as little as needed to ensure judicial economy and an orderly courtroom” should the defendant object to counsel’s assistance. This was acknowledged by Judge Antonetti in a dissenting opinion in Seselj, which was handed down between the decisions of the Trial Chamber and the Appeals Chamber in Milosevic. In Seselj, it was felt that standby counsel should only take over the defense from the Accused at trial “in exceptional circumstances . . . should the Trial Chamber find, following a warning, that the Accused is engaging in disruptive conduct or conduct requiring his removal from the courtroom under Rule 80(B),”

The Appeals Chamber found the significant curtailment of Milosevic’s role in his own defense to be disproportionate in light of the Accused’s


266. U.N. Int’l Human Rights Instruments, Compilation of General Comments and General Recommendations Adopted by Human Rights Treaty Bodies, at 176, para. 14, U.N. Doc. HRI/GEN/1/Rev.6 (May 12, 2003). The HRC also stated that “[t]he application of restrictions in any individual case must . . . meet the test of necessity and the requirements of proportionality.” Id. para. 16.

267. Chassagnou v. France, 1999-III Eur. Ct. H.R. 21, 53. The Chassagnou court also held that only “indisputable imperatives” can justify restrictions on a right protected by the ECHR, and even then only if the restrictions are a “necessary” and “proportionate” means of advancing the state objective. Id. paras. 112–13; see also Croissant v. F.R.G., 237 Eur. Ct. H.R. (ser. A) 20, 43 (1992) (Messrs. Frowein, Weitzel, Schermers & Mrs. Liddy, dissenting) (noting that while the appointment of a third defence counsel was appropriate in the light of “the public interest to ensure the proper running of the trial,” his active intervention in the trial was “neither necessary nor required”).


269. Prosecutor v. Seselj, Case No. IT-03-67-PT, Decision on the Accused’s Motion to Re-Examine the Decision to Assign Standby Counsel, Opinion Dissidente du Judge Antonetti, para. 9 (Mar. 1, 2005) (noting “[I]l convient de garder en mémoire que la mesure d’ingérence doit être nécessaire et proportionnée au but légitime”) (“[It is advisable to remember that the measurement of interference must be necessary and proportionate to the legitimate goal” (translation by Brooklyn Journal of International Law)).

270. Prosecutor v. Seselj, IT-03-67-PT, Decision on Prosecution’s Motion for Order Appointing Counsel to Assist Vojislav Seselj with His Defence, para. 30 (May 9, 2003).
vigorous two-day opening statement without interruption or apparent difficulty and a lack of evidence that Milosevic’s condition was permanent or that he had suffered from any health problems since late July. 271

On these bases, the Appeals Chamber ordered that the practical impact of the assignment of counsel should be minimized “except to the extent required by the interests of justice.” 272 The regime put in place thereafter was to be rooted, therefore, “in the default presumption that, when he [wa]s physically capable of doing so, Milosevic [would] take the lead in presenting his case.” 273 The captaincy of the ship had reverted to the Accused.

Nevertheless, the Trial Chamber was entrusted with the discretion to “steer a careful course” between allowing the Accused to take the lead and safeguarding “the Tribunal’s basic interest in a reasonably expeditious resolution of the cases before it.” 274 Crucially, the trial could continue in the event Milosevic was temporarily incapable of participating. 275 If Milosevic elected not to continue to act as counsel, became belligerent or uncooperative, or proved incapable of conducting even part of the defense, the Trial Chamber could direct Assigned Counsel to proceed with the representation. 276

Indeed, this very situation occurred when, during the presentation of the evidence of defense witness Kosta Bulatovic in April 2005, the Accused was unable to attend court through illness. 277 The Trial Chamber decided nevertheless to hear the remainder of the evidence, noting, “[i]f the decision of the Appeals Chamber is authoritative for anything, it seems to us that it authorises the completion of a witness’s testimony in the temporary absence of the accused.” 278 Thus, while Milosevic stood at


272. Id. para. 19.

273. Id.

274. Id.

275. Id. para. 20.


277. The witness initially gave evidence on April 14, 2005. Trial Tr. at 38503, Prosecutor v. Milosevic, Case No. IT-02-54-T, available at http://www.un.org/icty/transe54/050414IT.htm. He was examined in chief and partially cross-examined before the case was adjourned over the weekend. When the trial resumed on April 19, the Accused was absent through illness. Id. at 38577:5–6, available at http://www.un.org/icty/transe54/050419IT.htm.

278. Trial Tr. at 38591:1–3 (Apr. 19, 2005), Prosecutor v. Milosevic, Case No. IT-02-54-T; see also Milosevic, Case No. IT-02-54-R77.4, Contempt Proceedings Against Kosta Bulatovic, Decision on Contempt of the Tribunal (May 13, 2005), available at http://www.un.org/icty/milosevic/trialc/judgement/bulatovic.htm; Milosevic, Case No.
the helm, the compass remained in the hands of the Trial Chamber. Significantly, the Trial Chamber was careful in this instance to ensure that a video recording and transcript of the proceedings should be delivered to the Accused to enable him to review the remainder of the evidence, and declared that, should it be necessary, Bulatovic could be recalled. This ensured that the Accused remained in control of the presentation of the evidence and of his defense strategy. The Trial Chamber thereafter consistently concerned itself with these interests of the Accused, adjourning the proceedings when Milosevic could not attend due to ill health. Thus, the dignity and autonomy of the Accused were essentially preserved.

4.1.2: Dignity and Autonomy

Dignity and autonomy were first recognized in Faretta to be essential factors of the pro se right: “The right to defend is personal. The defendant, and not his lawyer or the State, will bear the personal consequences of a conviction. It is the defendant, therefore, who must be free personally to decide whether in his particular case counsel is to his advantage.” Supplying a defense to the accused does not sufficiently protect the individual’s autonomy; rather, the defense must be the individual’s own. In order to preserve the Faretta right to proceed pro se, it is crucial that the accused retains final authority over significant decisions. In McKaskle v. Wiggins, the U.S. Supreme Court held that counsel must not unduly interfere with the perception that the accused is acting as lead counsel. Thus, counsel “must generally respect” the defendant’s preference to proceed “solo,” and a defendant may not be “forced to sub-

---

279. Trial Tr. at 38591:5–8 (Apr. 19, 2005), Prosecutor v. Milosevic, Case No. IT-02-54-T.
280. See, e.g., id. at 46635 (Nov. 16, 2005), available at http://www.un.org/icty/transe54/051116IT.htm. Milosevic indicated, “I cannot call the next witness. I don’t feel well, and I can’t go on sitting here, and I want to say that I am opposed to any hearing in my absence.” Id. at lines 9–11. The evidence of Lieutenant Colonel Zlatko Odak had just been completed. Id. The trial was adjourned until the Accused was once again fit to attend trial. Arguably this caution was strengthened by the contempt proceedings that resulted from Kosta Bulatovic’s refusal to give evidence in the absence of the Accused.
283. Id.
284. Id. at 188; see also Parren v. State, 523 A.2d 597, 599 (Md. 1987).
mit to counsel’s unwanted control.” This “may mitigate his distrust both of the criminal justice system and of his counsel” while safeguarding his dignity and enlarging his freedom of choice.

It must not be forgotten, of course, that the assistance of counsel is considered vital to due process, and particularly so in adversarial systems. The recognition of these two interests of the accused—“full representation” and self-representation—as autonomous entities leads to a conflict “between preservation of the reliability of the judicial process and protection of the dignity of the defendant,” as it appears that “each right can only be exercised at the expense of the other.”

The possibility of assigning advisory, standby, or hybrid counsel mitigates this apparent mutual exclusivity and ensures that both the

285. Welcom, supra note 29, at 582.
286. Id. at 585.
288. See Daniel D. Ntanda Nsereko, Ethical Obligations of Counsel in Criminal Proceedings: Representing an Unwilling Client, 12 CRIM. L.F. 487, 488 (2001) (“Vital as representation is to the litigant’s assertion for his rights before the courts and other adjudicative bodies, it is particularly indispensable in an adversarial as opposed to an inquisitorial system of justice.”).
289. Welcom, supra note 29, at 581.
291. Some courts use the terms “advisory counsel,” “standby counsel,” and “hybrid representation” interchangeably. United States v. Singleton, 107 F.3d 1091, 1097 n.2 (4th Cir. 1997). But see, e.g., Locks v. Sumner, 703 F.2d 403, 407 & n.3 (9th Cir. 1983) (discussing use of and distinctions between the terms “standby counsel,” “advisory counsel,” and “co-counsel”). “Advisory counsel” do not participate in trial proceedings while “standby counsel” may be expected to assume the defense if the defendant is not able to continue. Id.
292. Some courts have indicated that standby counsel serves the dual purposes of providing a “safety net” in ensuring a fair trial and of eliminating delays. Anne Bowen Poulin, The Role of Standby Counsel in Criminal Cases: In the Twilight Zone of the Criminal Justice System, 75 N.Y.U. L. REV. 676, 707 (2000) (citing United States v. Bertoli, 994 F.2d 1002, 1018–19 (3d Cir. 1993); State v. Ortisi, 706 A.2d 300, 308–09 (N.J. Super. Ct. App. Div. 1998)). Standby Counsel can participate in the trial only to the extent that they do not usurp actual control or interfere with the perception of control. Id.
293. Standby counsel are not normally permitted to actively represent the defendant. See McKaskle v. Wiggins, 465 U.S. 168, 178 (1984). The McKaskle Court stated that:

[T]he pro se defendant is entitled to preserve actual control over the case he chooses to present to the jury. . . . If standby counsel’s participation over the defendant’s objection effectively allows counsel to make or substantially interfere with any significant tactical decisions, or to control the questioning of witnesses, or to speak instead of the defendant on any matter of importance, the Faretta right is eroded.
interests of justice and those of the defendant are duly protected. As
counsel in Milosevic essentially functioned as “hybrid” counsel, as
they shared the role of defense counsel with the Accused. This consti-
tuted a “co-counsel” model which involved actual assistance in the
trial proceedings. Thus, the Accused simultaneously exercised his rights
to counsel and to represent himself, consistent with the nature of hybrid
representation.

4.2: Hybrid Representation and Milosevic

The hybrid model was the most suitable for the Milosevic case. Famili-
arity with the vast and countless facts, figures, locations, and witnesses
made the Accused more suited to conducting certain aspects of his de-
fense. This was particularly important as political and moral beliefs
were central to the Accused’s defense. Hybrid representation allowed
Milosevic to publicly conduct cross-examination and examination-in-
chief while relinquishing to counsel those procedural and substantive
tasks which required the legal skills at which they excel.

Hybrid representation “is more likely to encourage efficiency and order
than to promote chaos.” It also serves to preserve judicial neutrality.
This is because inexperienced pro se defendants often rely on judges for
assistance in legal and procedural matters, and hybrid representation

Id. (emphasis added). When standby counsel assume the role of lead counsel, the
relationship has evolved into an inadvertent form of hybrid representation. This consists of
“concurrent representation by both defendant and counsel.” State v. Cornell, 878 P.2d
1352, 1363 (Ariz. 1994) (en banc). Hybrid counsel participate in or assist in the pre-
paration of such activities as opening statements, examination of witnesses and closing argu-

294. On the various benefits and inconveniences of these models, see Richard H.
Chused, Faretta and the Personal Defense: The Role of a Represented Defendant in Trial
Tactics, 65 CAL. L. REV. 636 (1977); Colquitt, supra note 36; Poulin, supra note 292;
Marie Higgins Williams, The Pro Se Criminal Defendant, Standby Counsel, and the
Judge: A Proposal for Better-Defined Roles, 71 U. COLO. L. REV. 789 (2000); Michael P.
Erhard, Note, The Pro Se Defendant’s Right to Counsel, 41 U. CIN. L. REV. 927, 940
(1972); Welcom, supra note 29.

295. Locks v. Sumner, 703 F.2d 403, 407 (9th Cir. 1983).
296. Welcom, supra note 29, at 570.
297. See U.S ex rel. Davis v. McMann, 386 F.2d 611, 620 (2d Cir. 1967) (“[A] defen-
dant who knows the facts of his case may indeed be better off defending himself without
an attorney than with an inadequately prepared one.”).
298. See Scharf, supra note 169; see also Erhard, supra note 294, at 932.
299. See Welcom, supra note 29, at 573.
300. Colquitt, supra note 36, at 107.
301. Id. at 108.
302. Chused, supra note 294, at 653 n.73 (citation omitted).
can “protect[] the trial judge from the moral, if not legal, obligation to help the defendant,” thus allowing “the judge to remain a neutral arbitrator rather than . . . becom[ing] an active participant in the trial.”

Furthermore, this also allows the victim, the prosecutor, and the witnesses to see the judge as impartial.

Problems arise, however, from the ambiguity of the relationship between the hybrid counsel and the accused. Without a defined role, hybrid counsel are cast into “an uncomfortable twilight zone.” Normally holding a position of complete autonomy and control, counsel must allow the accused to feel in control. Thus, hybrid representation “poses a risk of clashing wills.” Moreover, while hybrid counsel are only supposed to assist when the accused so requests, well-intentioned co-counsel often cannot remain idle while the defendant makes mistakes. This ultimately could lead to the accused feeling “short-changed” or to believe that counsel is forcibly interfering with his case. This “undermines the appearance of fairness and places counsel in the untenable position of supporting a hostile pro se defendant.”

The Assigned Counsel in Milosevic overcame these obstacles, curbing their public role in trial proceedings to the disputation of procedural matters beyond the expertise of the pro se Accused. The Assigned Counsel comprehensively fulfilled their dual obligation to the Chamber and to the Accused through the filing of innumerable public and confidential written submissions on behalf of the Accused, and responding on his behalf and in his interests to those of the Office of the Prosecution, the Chambers and the Registry. It is a credit to the professionalism, patience, and sound judgment of Mr. Kay and Ms. Higgins that they fulfilled their

303. Welcom, supra note 29, at 584.
304. See Prosecutor v. Norman, Case No. SCSL-04-14-T, Decision on the Application of Samuel Hinga Norman for Self Representation Under Article 17(4)(d) of the Statute of the Special Court, para. 26 (June 8, 2004).
305. See Howard, supra note 268, at 859.
306. Poulin, supra note 292, at 676.
308. Howard, supra note 268, at 861.
309. Williams, supra note 294, at 807.
310. Poulin, supra note 292, at 687.
311. The assignment of counsel “not only entail[s] obligations towards the client, but also implies that [counsel] represents the interest of the Tribunal to ensure that the Accused receives a fair trial.” Prosecutor v. Barayagwiza, Case No. ICTR-97-19-T, Decision on Defence Counsel Motion to Withdraw, para. 21 (Nov. 2, 2000); see also United States v. Swinney, 970 F.2d 494, 498–500 (8th Cir. 1992) (noting that the “contentious relationship” of the accused with counsel was not grounds for the withdrawal of the latter, given that the relationship of the accused with substitute counsel could be expected to be equally fractious).
duties without compromising the wishes of the Accused. Ultimately, Milosevic maintained full control over the witnesses called in his defense, and over his defense strategy.

In conclusion, regardless of the substantive deficiencies of the Trial Chamber decision, the Appeals Chamber must be commended for its application of the principle of proportionality to the manner in which the Assigned Counsel would function in trial proceedings. It is hoped that the application of the principle of non-intentional disruption to the right of self-representation, particularly in the domain of ill health, will be strictly circumscribed, and conducted in such a manner as to truly accord “full equality” to the minimum defense guarantees of the accused, as required by Article 21(4) of the ICTY Statute.
WORLD WIDE WEB: USING INTERNET GOVERNANCE STRUCTURES TO ADDRESS INTELLECTUAL PROPERTY AND INTERNATIONAL DEVELOPMENT

I. INTRODUCTION

Each year, shortages of flu vaccines cause panic and controversy over the U.S. patent system. Global demand for such drugs as Tamiflu far exceeds the production levels needed to satisfy that demand.¹ Populations of developing nations are ravaged with AIDS and without the ability to pay for appropriate drugs, even when the drugs are available in remote and rural regions.² However, with the development of the Internet, organizations are using mobile vans to bring life-saving health care information, including Internet sources when possible, to those remote areas or underserved populations.³ The glaring lack of access to information, however, plays out around the world in many sectors—education, arts, and science to name a few. Frustration is mounting and voices are begging to be heard. To paraphrase a common argument in the debate over intellectual property regulation, information wants to be shared.⁴

United Nations agencies are currently working on issues surrounding the Internet and global development—bringing the global community online to serve larger development goals, such as providing medicine to AIDS-ravished nations, making pre-natal and other healthcare information available via mobile Internet vans, and bringing literacy to children in impoverished villages. Two such agencies are the Working Group for

---

¹ See David Brown, Run on Drug for Avian Flu has Physicians Worried, WASH. POST, Oct. 22, 2005, at A1 [hereinafter Avian Flu article].
⁴ “Information wants to be free” is a mantra used by scholars arguing for a regime focused on the preservation of the public domain and fair use provisions in the world of intellectual property. R. Polk Wagner, Information Wants to be Free: Intellectual Property and the Mythologies of Control, 103 COLUM. L. REV. 995 (2003) [hereinafter Wagner I].
Internet Governance and the World Intellectual Property Organization. Separately, these organizations are bringing issues of technology and global infrastructure to the fore of the debate on international development objectives. Together, this Note will argue, these groups could better address these objectives and make real progress in the developing world.

In July 2005, the Working Group for Internet Governance (WGIG), under the auspices of the United Nations and the World Summit on the Information Society (WSIS), issued a report on the future of Internet governance. The WGIG has identified important public policy issues and overarching framework elements vital to the future of a global perspective for the Internet. The WGIG report focused on the need for a global forum for discussion and negotiation between all nations with an emphasis on the inclusion of the developing world. Meanwhile, the World Intellectual Property Organization (WIPO) and the “Access to Knowledge” movement are advocating a treaty which would address the technological advances required to promote access to information and educational resources traditionally protected as intellectual property of more developed nations.

Generally, WGIG and WIPO occupy mutually exclusive spheres of influence, though there is significant overlap between intellectual property issues and Internet governance. Lawrence Lessig, a preeminent Internet law scholar, expressed this conundrum in his work *Free Culture*. Lessig served as a keynote speaker during a preparatory conference for the

---

6. Id. at 5–8.
7. Id. at 10–12.
8. Participants in this movement include Doctors without Borders, Electronic Frontier Foundation, and other international organizations and individuals. See discussion infra text accompanying notes 168–73.
10. Lawrence Lessig is a Professor of Law at Stanford Law School and founder of the law school’s Center for Internet and Society. Lessig has authored several books on Internet and the law including *Free Culture, The Future of Ideas*, and *Code And Other Laws of Cyberspace*. He is also the founder of the Creative Commons project (http://www.creativecommons.org) and is involved with several other Internet and law-related organizations. More information on Lessig and his work is available at his website, http://www.lessig.org.
WSIS in 2003. In an interview before the address, Lessig focused on the balance between intellectual property protections and access needed in the burgeoning Information Society. This statement was met with great consternation—the president of WSIS responded that intellectual property “questions were the exclusive domain of [the World Intellectual Property Organization]” and not the WSIS. However, this response highlighted for Lessig the need to consider the importance of intellectual property in an “information society”—for, as he argues, conversations about the information society cannot occur without addressing the balance between the holders of intellectual property rights and those who would benefit from the protected information. Without the widespread sharing of knowledge, society instead becomes divided into the informed and those struggling to grasp hold of information.

In this ignored zone of overlap, developing nations and the developed world (particularly the United States) are caught in a power struggle on these very issues—the struggle between control and information sharing. Since the United States is the current overseer of Internet governance through the Internet Corporation for Assigned Names and Numbers (ICANN), the United States has expressed significant concern over the possibility of internationalizing control of the Internet. The officials remarked that the Internet structure is complex and should remain within

12. Id. at 263.
13. Id.
14. Id.
15. For the purposes of this Note, “developed nations” will refer primarily to the United States as a major holder of intellectual property and a major player in the United Nations and ICANN. “Developing nations” will generally refer to those nations who are struggling for greater access to information on the global playing field. Brazil and Argentina have led a coalition of nations to achieve balance in the global intellectual property realm and will be counted among “developing nations” within this context.
17. ICANN is an international non-profit organization which oversees the technical aspects of the Internet, with oversight by the U.S. Department of Commerce. ICANN will be discussed more thoroughly later in the Note.
18. IP Watch Articles, supra note 16.
the control of ICANN and the United States so that the Internet “remains stable and secure.” They further asserted that the United States would make no changes to disrupt the “effective and efficient” administration of the DNS system. Of course, these reactions to any possibility of change in control do not address the criticisms of ICANN as ineffective and inefficient. These criticisms will be addressed later in this Note.

This conflict between the United States and those who would internationalize control over the Internet mirrors the struggle over WIPO’s development agenda which has been negotiated between the United States and a coalition of developing nations. Brazil, Argentina, and several other nations presented a proposal in line with the “Access to Knowledge” movement, encouraging the World Intellectual Property Organization to take active steps to consider the needs of developing nations in its policies and regulations. Because these developing nations are concerned with global sharing of knowledge and much of the protected material is generated and owned in the United States, any search for the common ground seems to be lost in the struggle. Argentina and Brazil’s proposal seeks moderation in intellectual property in order to further development goals, however the counter proposal by the United States blindly affirms the status quo without true consideration of the Argentina-Brazil proposal. Any chance at achieving goals collectively, addressing the needs of both sides of the debate, disappears without a more collective approach. The United States and developing nations have seen similar conflicts throughout recent history in intellectual property treaty negotiations.

This Note will explore the proper balance between protection of intellectual property rights and global information sharing, specifically protecting copyright while promoting education and Internet-based archive

20. Id.
21. See IP Watch Articles, supra note 16.
22. The Access to Knowledge movement was created by numerous groups and individuals outside of the traditional WIPO process to encourage greater consideration of development goals in its intellectual property regulation practices. See discussion infra text accompanying notes 169–84.
23. IP Watch Articles, supra note 16.
24. Id.
25. Previous draft treaties presented before WIPO created conflict between the United States’ desire for “fortified copyright in cyberspace” and the developing nations who want greater access to information. Many nations rejected the proposals in past votes. See JESSICA LITMAN, DIGITAL COPYRIGHT 128–29 (2001).
and library systems. It will argue, using key tenets of the WGIG and WIPO proposals, that there is a common ground on the issues which will satisfy the shared goal of education and access to knowledge in developing countries while addressing the intellectual property protection concerns of the United States.

This Note proceeds in four parts. Part II will examine the current status of Internet governance by ICANN and the U.S. Department of Commerce. It will then address the WGIG proposals and the problems of Internet governance with respect to intellectual property issues. Part III will introduce the current state of global intellectual property regulation, with the World Intellectual Property Organization at the forefront, as well as global efforts to include development goals in IP discussions. Part IV will evaluate theoretical and practical perspectives including existing technologies designed to balance access and control of protected works. In conclusion, this Note will propose using these technologies to create a middle ground—a balance of intellectual property rights and access, using the Internet as a tool for the calculated spread of information instead of viewing it as a lawless frontier.

II. CURRENT STATUS OF INTERNET GOVERNANCE: ICANN AND WGIG

A. International Corporation for Assigning Names and Numbers (ICANN)

The technical aspects of the Internet, including domain name regulation, are currently governed by ICANN. ICANN is a non-profit organization, created by the U.S. Department of Commerce (DOC), which ultimately organizes and maintains order on the Internet. ICANN handles “Internet Protocol (IP) address space allocation, protocol identifier assignment, generic (gTLD) and country code (ccTLD) Top-Level Domain name system management, and root server system management functions.” Welcome to ICANN FAQ, http://www.icann.org/new.html (last visited Mar. 23, 2007). These technical aspects of the Internet are crucial stepping stones to accessing the Internet. Through
randum of Understanding which established the organization included
the DOC’s intent to transfer control of domain name management from
the United States to the “global community.”  

The board and staff are
comprised of individuals from many nations; meetings are open to public
participation by “all who have an interest in global Internet policy.”

ICANN also consists of several advisory committees such as the Gov-
ernmental Advisory Council and the At-Large Advisory Committee.

Supporting organizations, such as the independent domain name registry
offices, also advise ICANN on an international scale.

ICANN was established in response to a 1998 White Paper issued by
the Clinton Administration to move control of the domain name system
into the private sector from ICANN’s predecessor organization Internet
Assigned Numbers Authority (IANA).  

The White Paper called for the
creation of a policy-making domain name group in the private sector.

The controlling principles of this group would be “stability, competition,
private bottom-up coordination, and representation.”

Since its establishment, there has been some disagreement over the ef-
effectiveness of ICANN and how successfully it has achieved the goals set
out at its inception. In 2002, ICANN’s President issued a proposal to
“fix” what was “broken” about ICANN in its present state.

Primarily, the proposal was concerned with reforming ICANN’s structure, address-
ing transparency and accountability issues, and streamlining the deci-
these protocols and management systems, ICANN is able to control how content is or-
organized and accessed across the globe.

29. See ICANN Fact Sheet, supra note 27.
30. Id.
31. Id.
32. Id.
33. IANA was established to govern Internet development under direct contract from
the U.S. Government. In the 1998 White Paper, the Clinton Administration noted that in
acting as a government contractor, IANA “describes a function more than an entity” and
recommended a formal incorporated structure. IANA has been subsumed by ICANN and
remains descriptive of a central ICANN function: the assignment of names and numbers
on the Internet. U.S. DEP’T OF COMMERCE, WHITE PAPER ON MANAGEMENT OF INTERNET
NAMES AND ADDRESSES, No. 980212036-8146-02 (June 5, 1998), available at
34. Id.
35. See Crawford, supra note 26, at 412.
36. ICANN President Stuart Lynn issued this proposal in order to address a number
of criticisms which had been made both within and without the ICANN system. Id. at
415; see also Stuart Lynn, A Proposal for Reform of ICANN, Feb. 2, 2002,
Proposal]; Mark Ward, Wanted: New Plan to Run the Net, BBC NEWS ONLINE, Mar. 30,
sion-making processes. The proposal also questioned the ability of ICANN to address problems and policy outside of the narrow scope of the domain name system—"how to reflect public interest concerns such as fair competition, privacy, intellectual property, and diversity?" In addressing this issue, the proposal acknowledged the rigidity of the structure of the ICANN system as well as the strength of ICANN’s power in controlling the Internet by asking, “If not ICANN . . . then who would perform these policy functions?” Under the original structure of ICANN, little guidance was given for representing the public interest in Internet policy development.

Just how far ICANN is able to go in regulating the Internet or how far the world wants ICANN to go is unclear as its mission simply states that it will “[c]oordinate policy development reasonably and appropriately related to [its] technical functions.” With such a general statement, the boundaries of ICANN’s mission are murky at best. Global intellectual property issues generally are considered to be in the jurisdiction of the World Intellectual Property Organization of the United Nations, although ICANN’s control has extended to cover some intellectual property issues. However, these policy issues surrounding Internet governance fall squarely between the established organization structures of WGIG and WIPO.

The current structure does allow for government input on issues of public policy through the Government Advisory Council (GAC) of

37. ICANN Proposal, supra note 36.
38. Crawford, supra note 26, at 421.
40. There is also some question as to how much power ICANN actually has to effect broad policy decisions due to its structure and, one could argue, because of the much narrower functions which the 1998 White Paper outlined—(1) to set policy for and direct the allocation of IP number blocks; (2) to oversee the operation of the Internet root server system; (3) to oversee policy for determining the circumstances under which new top level domains would be added to the root system; and (4) to coordinate the assignment of other Internet technical parameters as needed to maintain universal connectivity on the Internet. See Crawford, supra note 26, at 420–24, 429.
41. WIPO is concerned with “developing a balanced and accessible international intellectual property . . . system.” Icann.org, What is WIPO?, http://www.wipo.int/about-wipo/en/what_is_wipo.html (last visited Apr. 4, 2007). ICANN, on the other hand, covers more Internet specific intellectual property issues. See, e.g., Crawford, supra note 26, at 411 (discussing ICANN’s role in issues of trademark and Cybersquatting.)
ICANN.\textsuperscript{42} The GAC is an entity comprised of representatives from national governments, multinational governmental organizations, and public authorities.\textsuperscript{43} Membership is voluntary and the GAC itself admits that global representation is not yet fully achieved, as nations are still not as yet represented.\textsuperscript{44} Furthermore, though the council serves as a forum to advise ICANN, it has no legal authority and no decision-making power, leaving the findings and recommendations of the GAC without bite.\textsuperscript{45}

The ICANN system, including the GAC, however, has been further criticized as merely an instrument of American control because of its inadequate representation of developing nations.\textsuperscript{46} For example, ICANN has been criticized for its “perceived bias toward insiders and large corporations,”\textsuperscript{47} as well as for its myriad deficiencies in the current state of ICANN, including lack of meaningful outsider participation, complex structures, and a lack of openness about procedures.\textsuperscript{48} Specifically, the ICANN board, as a corporate governance entity, makes decisions about the Internet, but there is no true mechanism to represent the Internet “citizenry” analogous to a government structure.\textsuperscript{49} There can be no real outside participation because there is no actual representation of ICANN’s constituency.\textsuperscript{50} These problems highlight some of the criticisms of this system of Internet governance—ICANN is both too narrow and too complex to continue to serve the Internet community in its current capacity.

\textbf{B. New Directions: WGIG and U.N. Proposals on Internet Governance}

How the Internet should be governed and by whom are important questions which incite a broad spectrum of responses. On one end, there is a school which argues that the Internet grew up out of a collective spirit
which defies governments, nations, and affiliations and therefore should stay that way. On the other end, ICANN and other organizations have been structured to give the Internet just that—structure and the resultant governance.

Pioneers of the Internet argue that because the Internet developed organically without a governing body that it should be allowed to continue that way. One important reason for this unregulated vision of the Internet is the lack of consent by Internet users to be governed by any government or other regulating body. Academics have addressed the development of norms in this early Internet society—created as needed within smaller anonymous communities. Instead of creating overarching and overreaching rules, these advocates would keep the Internet generally unregulated, allowing groups of users to come up with their own rules or norms as situations arise. Such groups might be members of a chat forum, online gamers, or other online communities. With this model, consent to the rules comes with membership in the group and autonomy is preserved.

On the other hand, scholars have also argued that regulation is necessary in order to serve the greater public interest. One scholar examined post-Soviet Eastern European development to demonstrate the need for some structured regulation. Where establishment of rules and norms are left to happenstance, certain sectors of society will not be developed. In the situation of the Internet, much of the clamor for strict intellectual property regulation comes from the American movie and music industries—an area where protections for intellectual property rights have not developed in the absence of a governance body.


52. See generally Palfrey, supra note 46; see also Crawford, supra note 26.

53. Declaration of Independence, supra note 51.

54. Id.

55. For an example of norm development in a small online community, see LESSIG, CODE, supra note 51, at 142–43.

56. Declaration of Independence, supra note 51.

57. LESSIG, CODE, supra note 51, at 3–6.

58. Id.

In a move to globalize the debate and solidify the definition of “Internet governance,” the World Summit on Information Society (WSIS) was organized by the United Nations. The U.N. convened the summit both to develop a global vision and common understanding of the information society and to outline a strategic plan of action for development. Furthermore, the summit was to define an agenda detailing future objectives and how to achieve those objectives with resource mobilization. Attendees included representatives from U.N. agencies, member states, and individuals from the private sector and civil society. Ultimately, a Declaration of Principles and Plan of Action solidified the organization’s commitment to “build a people-centred, inclusive and development-oriented Information Society, where everyone can create, access, utilize, and share information and knowledge, enabling individuals, communities, and peoples to achieve their full potential in promoting sustainable development and improving their quality of life . . .” in accordance with the U.N. charter.

Some of the objectives identified at the close of the summit included connecting all “schools, villages, Governments and hospitals” to the Internet and each other “by 2015.” Internet governance was also identified as a major long-term target for discussion and future action. The summit also specifically addressed the importance of intellectual property issues in global development of the information society:

60. The WSIS meeting was held on December 10, 2003 in Geneva. WGIS REPORT, supra note 5, at 3.
62. Id.
63. Id.
On the areas of intellectual property rights and the need for enabling environments, universal access policies, and multilingual, diverse and culturally appropriate content to speed ICT adoption and use—particularly in the world’s most underserved economies—government-level commitment to follow a set of common values and principles has been attained.\textsuperscript{57}

In response to the first phase of the WSIS summit, the Secretary-General established the Working Group on Internet Governance (WGIG) to address a number of Internet-related issues.\textsuperscript{68} The committee mandate recognized that the Internet is a “central element of the infrastructure of the emerging information society.”\textsuperscript{69} In order to represent the breadths of this society, forty members were drawn from “[g]overnments, private sector and civil society, who all participated on an equal footing and in their personal capacity.”\textsuperscript{70}

The members of the WGIG convened regularly to discuss the objectives articulated by the WSIS Declaration of Principles and the WSIS Plan of Action.\textsuperscript{71} These two documents chiefly charge the WSIS and, as its action arm, the WGIG, to promote active and global government participation, encourage infrastructure and global information sharing.\textsuperscript{72} Of particular interest are action items emphasizing the importance of getting valuable information to remote societies and institutions, therefore requiring an inquiry into the proper balance of intellectual property regulation. These action items include fighting illiteracy through development of inexpensive computer interfaces and access to information as well as developing policies which would enable libraries, archives, and cultural institutions to function fully in the Information Society.\textsuperscript{73}

Furthermore, the Plan of Action presents a number of provisions to level the playing field such as encouraging traditional media—books, newspapers, television—to take advantage of new technology and better Internet connectivity to spread cultural information to (and from) rural areas and connecting schools, libraries, and hospitals to the Internet and to each other for further information exchange.\textsuperscript{74} In addition to these policy actions, WGIG was also specifically charged with developing a

\textsuperscript{57} Id.
\textsuperscript{68} WGIG REPORT, supra note 5, at 3.
\textsuperscript{69} Id. at 3.
\textsuperscript{70} Id.
\textsuperscript{71} The committee met four times in 2004 and 2005. Id.
\textsuperscript{72} Declaration of Principles, supra note 62, at §§ 26, 29, 30; Plan of Action, supra note 62, at §§ 9(g), 23(b).
\textsuperscript{73} Id.
\textsuperscript{74} Plan of Action, supra note 62, at § 6(e).
working definition of Internet governance, identifying the relevant public policy issues and examining as well as defining the roles of society’s layers with relation to Internet governance.\(^{75}\)

Of primary importance was the establishment of a practical, yet inclusive definition for a “shared view of Internet governance.”\(^{76}\) The definition was required to be “adequate, generalizable, descriptive, concise and process-oriented.”\(^{77}\) The group analyzed governance mechanisms already in place as well as definitions proposed by various parties to the development process.\(^{78}\) The current consensus is that “Internet governance is the development and application by Governments, the private sector and civil society, in their respective roles, of shared principles, norms, rules, decision-making procedures, and programmes that shape the evolution and use of the Internet.”\(^{79}\) WGIG emphasized that this definition was purposefully inclusive of each constituency in the “mechanisms of Internet governance” though the interests of each group may overlap or diverge.\(^{80}\)

Using this definition of Internet governance as a common starting point is an important first step in forging further consensus on regulation of the Internet. The characterization of governance is indeed “adequate” and “generalizable” including government, private, and civil sectors with broad strokes. The definition, however, in being so inclusive, also leaves considerable space for determining what “shared principles” or “evolution and use of the Internet” will come to mean. At the same time, these terms also ensure that Internet governance on the global scale will not be limited to the technical aspects of regulation as with ICANN. By discussing principles, norms, and programs to shape the Internet, the WGIG explicitly carved out an important niche for policy development in tandem with governance structures and rules.

The WGIG identified a number of policy issues as important to the development of the Internet generally and to the overarching goal of leveling the playing ground between developing and developed nations.\(^{81}\) WGIG identified four main areas of “potentially relevant” Internet governance policy—infrastructure and resources, Internet use (including

---

75. WGIG REPORT, supra note 5, at 3. (“Governments, existing international organizations and other forums, as well as the private sector and civil society in both developing and developed countries.”).
76. Id. at 4.
77. Id. (emphasis in original).
78. Id. at 5.
79. Id.
80. Id.
81. WGIG REPORT, supra note 5.
spam, security and cyber crime), issues of wider impact with existing organizations (including intellectual property rights), and developmental issues (including capacity building). Here WGIG immediately recognizes WIPO as the existing organization to deal with intellectual property rights; however global information sharing, essential at least to WGIG’s development objectives, relies heavily on the balance of information protections. Despite the importance of integrating these objectives at the outset, the spheres of influence are kept separate.

Within these individual areas, WGIG expressed strong interest in making global participation more meaningful, including increasing access to the Internet internationally. Some important mechanisms for achieving this goal are distribution of costs for interconnection to ensure that remote nations are able to access the Internet, creating a “global mechanism” for representing the needs of developing nations, and the promotion of “multilingualism” in the technology as well as Internet content. WGIG’s emphasis on these issues, however, does not mean that they are entirely new to the governance discussion. Recently, ICANN has helped encourage multilingual content by creating the .cat domain for Catalan language web content as well as by adopting guidelines for International Domain Names which enables even broader international web content to grow. WGIG has, with its report, laid a foundation for improving upon the original goals of the private international organization structure established with ICANN.

To promote these policies, WGIG presented four proposals for the future of Internet governance to be discussed and further developed in anticipation of the final WSIS meeting in Tunisia in November 2005. Each of the four models was concerned with increasing participation and access to the process through internationalization. The most significant mechanism, present in each model, was a public forum to ensure that policy discussions, as well as oversight of the governance process generally, would be truly global. WGIG details these elements with an em-

82. Id. at 7.
83. See id. at 6–7.
84. Id.
86. ICANN Fact Sheet, supra note 27.
87. See WGIG REPORT, supra note 5, at 13–16.
89. WGIG REPORT, supra note 5, at 11–12.
phasis on consensus-building and using the future Internet governance structure to promote a cooperative collective system of governance.  

The first model would bring ICANN under the U.N.’s auspices to create a “Global Internet Council” where governments would take the lead in a multilateral governance structure. The second model would keep ICANN intact, giving its Governmental Advisory Council increased importance, and removing ICANN from U.S. control. This model includes no particular oversight organization. In its third model, WGIG proposes the creation of an “International Internet Council” where governments would again take a leading role. This council would take over the functions of the GAC and the role of the U.S. government—creating a host country agreement for ICANN so that unilateral control by the incorporation nation is avoided completely. Finally, WGIG proposed a fourth model which combines three layers of governance—a “Global Internet Policy Council” made up of government representatives and “responsible for international Internet-related policy issues,” an intergovernmental version of ICANN to be known as World ICANN or WICANN, and a Global Internet Governance Forum to specifically fulfill the role of the vital world forum. Though each of these proposals addresses the problems of the ICANN system differently, widespread international participation in the future of the Internet is common to each proposal.

Though the creation of the global policy forum was a central element of each proposal, WGIG was quite vague about which policy questions would be addressed, but some policy recommendations were made. Apart from practical funding considerations, the task-oriented recommendations fall within four public policy areas. Specifically, WGIG makes recommendations for security concerns, freedom of expression, meaningful global participation, and consumer rights. The report does acknowledge that some of these issues have a wider impact and scope than simply Internet governance. Notably, the report refers to intellectual property rights as the main example of such a category in an earlier

90. Id. at 12.
91. Id. at 13.
92. Id. at 14.
93. Id.
94. WGIG REPORT, supra note 5, at 14.
95. Id.
96. Id. at 15–16.
97. Id. at 16–19.
98. Id.
99. Id.
100. WGIG REPORT, supra note 5, at 16–19.
This reference makes the absence of any policy recommendations about intellectual property regulation quite glaring. In its earlier reference to intellectual property rights, the WGIG report defers to the intellectual property oversight organization. This reservation of policy neglects WGIG’s direct charge to address the policies of information, including use of the Internet for libraries, archives, and other cultural information (among other realms with significant intellectual property elements and ramifications). These two elements—Internet governance and intellectual property—walk hand in hand and any organization addressing the Information Society must also address these intellectual property issues. The WSIS officials however insist that this is to be left for WIPO to address.

In November 2005, WSIS reconvened in Tunisia to discuss the proposals and decided to leave the current system of Internet governance through ICANN intact and under U.S. control. However, the summit did reaffirm its commitment to increasing international government participation in Internet governance through the creation of the Internet Governance Forum. This forum will handle “cross-cutting international” public policy issues relating to Internet governance. This can be seen as a critical first step towards integrating the isolated spheres of intellectual property and Internet mechanisms. Through this forum, problems will be identified and recommendations will be made to appropriate governments; however this organization has been given no oversight and no real “teeth” in enforcement or implementation.

The summit also issued the “Tunis Commitment” which stated the agency’s commitment to its original goals (reaffirming the Declaration of Principles and Plan of Action created in Geneva). Notably, this statement of commitment directed discussions to include the ways in which Information and Communication Technologies (ICTs) can aid in addressing

101. Id.
102. Plan of Action, supra note 64.
103. See LESSIG, FREE CULTURE, supra note 11, at 263.
105. Id.
106. Id.
development goals. Particularly, the statement included a commitment to “developing countries, countries with economies in transition, Least Developed Nations,” and other groups of developing countries. This statement marks an important commitment to the spread of information to developing nations; however, there is still a significant need to coordinate these efforts with the recognized authority on protection/access to information, the World Intellectual Property Organization.

In order to establish concrete goals for the future, WSIS has created Action Lines with dedicated moderators to address the issues presented in the Plan of Action. In early 2006, the moderators of these committees met to start the process, including the launching of the Golden Book—a database tracking the initiatives of WSIS participants around the globe. Various projects are being undertaken at the national and international level by private, government, and non-governmental organizations in addition to the actions being taken in the Action Item Committees established by WSIS.

One specific committee, the WSIS Action Line C3 “Access to Information and Knowledge” Committee, met in October 2006 to bring committee members together, set Action Line C3 specific goals, and establish working methods to achieve these goals. The committee established seven areas of concern, including increasing public access to information, libraries and archives generally, as well as research and development in the information sector. Once these areas were identified, the

109. Id. para. 21.
110. WSIS Press Release, supra note 107.
112. Examples of ongoing projects include a Chinese initiative to bring telephone access to nearly 70,000 villages currently without connections, an Egyptian project to digitize a million books and place them in a searchable internet library, and efforts across the African continent to create greater connection and communication technologies. WORLD SUMMIT ON THE INFORMATION SOCIETY, PRESENTATION: THE GOLDEN BOOK, WSIS STOCKTAking, ICT SUCCESS STORIES, Feb. 24, 2006, http://www.itu.int/wsis/goldenbook/Publication/GB-Final.pdf.
114. Id.
committee also established objectives for each sub-topic to further guide future discussion and action.\(^\text{115}\)

This committee is an important step towards increased accessibility of information throughout the world. While one speaker at the meeting addressed the complexities of information sharing, noting that technology can be used to both encourage and prevent sharing,\(^\text{116}\) it is unclear how this committee will resolve Internet governance and intellectual property issues.

### III. GLOBAL INTELLECTUAL PROPERTY REGULATION: WIPO AND ACCESS TO KNOWLEDGE

#### A. World Intellectual Property Organization

Regulation of intellectual property, on the global scale, is apparently within the exclusive jurisdiction of the World Intellectual Property Organization (WIPO). WIPO is a special agency of the United Nations created to specifically deal with global regulation of intellectual property in the realms of science and the arts.\(^\text{117}\) The organization administers more than twenty international treaties.\(^\text{118}\)

Currently, WIPO is working on a “Digital Agenda” over the upcoming years to address technological developments and intellectual property.\(^\text{119}\) This “Digital Agenda,” however, seems to be charged with exactly the same tasks and policy considerations as WSIS and its WGIG.\(^\text{120}\) WIPO’s agenda consists of a work program dedicated to the integration of developing countries in making policy and to expansion of access to information.\(^\text{121}\)

The WIPO Digital Agenda, approved in 1999, is set up to deal mainly with the coordination of intellectual property offices in member countries.\(^\text{122}\) Specifically, WIPONet, a system created by the organization, links these offices to provide a secure space for communication, resources for greater Internet connectivity (including remote access to WIPO meetings), and resources for establishing a presence on the

---

\(^{115}\) Id.

\(^{116}\) Id. para. 6; see also Part IV infra.


\(^{118}\) Id.


\(^{120}\) See id.

\(^{121}\) See id.

\(^{122}\) See id.
web. This system functions, in a way, like ICANN: focusing on infrastructure and technical aspects of intellectual property, without providing a forum to address policy issues resulting from changing technology.

In addition to WIPO’s own Digital Agenda, Brazil and Argentina drafted a formal proposal, in August 2004, to create a WIPO Development Agenda. Building on a number of previous technology and development initiatives including the WSIS and WGIG principles, the proposal centers on “knowledge gap” between wealthy and poor nations. The proposal particularly notes the stringent protections on intellectual property in wealthy nations leading to the inability to share knowledge vital to the development process in poorer countries. Argentina and Brazil urged WIPO to make global development an integral concept in its ongoing activities and policies with a member-driven approach.

Primarily, the Argentina/Brazil proposal stresses the importance of creating a balance in intellectual property policy by recognizing the vital importance of the public interest side of information and innovation as well as structuring digital policies to value open source models of information sharing. In addition, enforcement mechanisms are worked into the balance—the coalition proposes that enforcement be made “fair and


125. The proposal builds upon the:

United Nations . . . Millennium Development Goals . . . [t]he Programme of Action for the Least Developed Nations for the Decade 2001–2010, the Monterrey Consensus, the Johannesburg Declaration on Sustainable Development and the Plan of Implementation agreed at the World Summit on Sustainable Development, the Declaration of Principles and the Plan of Action of the first phase of the World Summit on the Information Society, and most recently, the Sao Paolo Consensus adopted at UNCTAD XI . . .

Id. at Annex 1.

126. *See generally id.* Examples of this include medicine for impoverished nations suffering from the AIDS epidemic, books and literacy materials for children where there are no resources for libraries, and archival mechanisms for dying cultures in rural communities.


128. *Id.* at Annex 2–3, 5.
equitable” to take into consideration the developing nation status and its effect on future technologies and innovations. Apart from these broad policy issues, the proposal also includes specific task items to be considered by WIPO. These task items include enforcement and legitimizing mechanisms, such as a “High-Level Declaration” to address development concerns of the world community, amendments to the WIPO Convention to “specifically include” the development concerns into the objectives and functions of the Organization, and the establishment of a Working Group on the Development Agenda. The proposal’s suggested language for the WIPO Convention adds “fully taking into account the development needs of its Member States, particularly developing countries and least-developed countries” under the “Objectives” section of the Convention. This addition would quell much of the debate over whether or not addressing development goals is within the purview of WIPO.

In October 2004, a meeting was convened to discuss the Brazil-Argentina Proposal with the other member states. At this conference, WIPO agreed to consider the Development Agenda and to consider the broader impact of intellectual property on the global community. The Assembly decided to convene Inter-sessional Intergovernmental meetings to examine the proposals and any to be made by other member states. Discussion was also to be undertaken immediately with other intergovernmental organizations and agencies both within the United Nations and without, including the WTO. Notably, the WIPO decision referenced the “internationally agreed development goals” which were the cornerstone of the Argentina-Brazil proposal.

The United States countered the Brazil/Argentina Proposal by affirming WIPO’s current commitment to development (by connecting nations)

129. Id. at Annex 4.
130. Id. at Appendix.
131. The proposal also addresses the inclusion of provisions in present and future treaties relating to the development agenda, specific transfer of technology committees, participation by non-governmental organizations and civil society, and joint efforts with other international organizations such as the WTO and UNCATD. Id. at Appendix.
132. Argentina/Brazil Proposal, supra note 124, at Appendix. Id.
133. Id.
136. Id.
137. Id.
and dismissing intellectual property as one small part of the “necessary infrastructure” required for the “journey from developing to developed” nation status.\footnote{138} In fact, the United States refers to the agreements creating WIPO and argues that since WIPO was created “subject to the competence and responsibilities of the United Nations” and other agencies, WIPO is not the appropriate place to handle development issues.\footnote{139}

Instead of committing to balance, the U.S. proposal touts the importance of intellectual property protection and instead proposes the creation of a database, using existing WIPO technology, to match up needs and resources amongst member nations.\footnote{140} Ultimately, the United States’ stated goal for this approach is to coordinate development efforts,\footnote{141} instead of creating an open environment for development to occur organically. Participation in this database appears to be voluntary. Furthermore, as the United States would have it, the program would be entirely market-driven.\footnote{142} This approach is distinctly different than the member-driven Argentina-Brazil Proposal, and to this author, seems to be only a slight modification of the current situation with developing nations at the mercy of developed nations’ willingness to share information.

This conflict of views between Brazil and the United States is not a new one. The United States has threatened sanctions against Brazil to coerce a promise for greater protection of American intellectual property.\footnote{143} Brazil is also kept on a “priority watch list” for intellectual property violations.\footnote{144} These threats are a very real consequence of the conflict between the United States and other nations. Brazil has also threatened to lessen protection for copyrighted materials in the United States in

\begin{flushleft}

139. Id. at Annex 2–3.

140. Id. at Annex 3–4.

141. Id.

142. Id. at Annex 7.


144. Brazil is on the priority watch list for intellectual property violations maintained by the Office of the United States Trade Representative. The Office recognizes that progress is being made, however the report expresses concerns about the lack of protections for American intellectual property in some areas. OFFICE OF THE UNITED STATES TRADE REPRESENTATIVE, 2006 SPECIAL 301 REPORT, available at http://www.ustr.gov/assets/Document_Library/Reports_Publications/2006/2006_Special_301_Review/asset_upload_file473_9336.pdf.
\end{flushleft}
response to American policies on Brazilian cotton farming. While intellectual property should be properly protected, these power struggles between nations (between development and regulation) should be addressed on a global scale through a supra-governmental structure which specializes in the issues fueling the conflict—Internet governance and intellectual property.

In a move that seems to echo the U.S. position in this conflict, Mexico also submitted a proposal urging against any substantive change in the current WIPO agenda/structure. This proposal was far less elaborate than the American proposal; though it similarly emphasized the use of WIPONet to help nations. By focusing on this structure, however, these two proposals emphasize the importance of protection of IP, not sharing or balance. Increased protection of IP is beneficial to the wealthy nations—those that own intellectual property—and not beneficial to struggling developing nations who would benefit from a freer flow of information. Finally, the U.K. submission was not a proposal per se, but rather was a strategy paper, asserting no need for a change to the status quo.

Of course, one major issue with using the strictly structured system of proposals and meetings to address issues in a rapidly developing field is the actual speed with which these large, broad organizations can move in order to react and function with that change. Fortunately, there exists a group of individuals, non-governmental organizations, and other entities from outside of these organizations urging change and directing the policy conversations. In May 2005, experts from nations around the world gathered together to come up with a Draft Access to Knowledge Treaty. This treaty included the Geneva Declaration on the Future of the World Intellectual Property Organization, which serves as the statement of U.N.-outsider support for a new WIPO treaty addressing the

147. Governments Meet, supra note 146.
148. Representatives from the United States, Serbia, South Africa, United Kingdom, the Netherlands, Spain, Greece, Italy, Germany, Malaysia, France, India, Canada, Korea, Brazil, Chile, and others were present for the workshop which resulted in the draft treaty language. Indicare, Access to Knowledge: Make It Happen, www.indicare.org/tiki-read_article.php?articleId=102; see also Draft Treaty on Access to Knowledge, May 9, 2005, available at http://www.cptech.org/ip/wipo/a2k.pdf.
place of development goals in its future policy determinations. Hundreds of scientists, academics, representatives from non-profit organizations, and others began to sign the Declaration. The preamble of this document pledges WIPO to the goals of greater access to knowledge in both the arts and sciences. Many of the observations and policy items mirror the concerns of the Brazil-Argentina Coalition.

In September 2005, WIPO held further meetings on development issues in Geneva. Though several developed nations including the United States, Japan, and the United Kingdom balked, the “vast majority” of nations continued to push to keep the development agenda at the fore of WIPO’s ongoing agenda. This commitment was further solidified at WIPO’s meetings a month later. In October 2005, the development agenda survived efforts to have it re-directed to an inactive WIPO committee by the United States, Japan, and several E.U. states. Throughout 2006, little progress was made in WIPO, but it is crucial that the debate continue.

Despite the strides made by developing nations in keeping their concerns on the table, a clear struggle still exists between developing and developed nations on intellectual property issues. This struggle also arises where Internet governance issues are at stake. Nevertheless, with a strong commitment to address this struggle by both WGIG and WIPO, there is hope for compromise, to find a common ground.

IV. INTELLECTUAL PROPERTY AND INTERNET GOVERNANCE: PROMOTING ACCESS THROUGH INFRASTRUCTURE

A. Theoretical Perspectives on Intellectual Property Regimes

In the ongoing debate over the future of technology and intellectual property, scholars and experts usually fall into one of two camps. On the one hand, there is the “commons” view which suggests that the Internet
should be an “information commons” because “information wants to be free.”\(^{155}\) Frequently, this argument is described as advocating the preservation of “the public domain, arguing for the public’s right to speak and express views freely by not expanding copyright for digital works.”\(^{156}\) One contentious argument in the commons theory is that stringent copyright protection is a deterrent to the creativity and innovation which might result from more liberal regulations allowing for derivative uses.\(^{157}\)

In the conflict between the United States and the developing world, it is this “commons” view taken to its theoretical extreme that has the United States, as a protector of the IP rights-holder, up in arms. The motion picture and recording industries in the United States have mounted a crusade against illegal downloading of music and movies.\(^{158}\) By making electronic versions of copyright-protected materials available to the world in the name of development, these lucrative industries fear losing control over American-produced movies and music to piracy by foreign nations.\(^{159}\) Furthermore, American drug companies control a vast number of the patented medicines needed to combat epidemics as serious and widespread as AIDS as well as those as tentative and threatening as the South Asian bird flu.\(^{160}\) The United States and the pharmaceutical industry have argued extensively that patent protection on these vital medicines is required to allow companies to recoup the costs of intensive research and development of these life-saving drugs.\(^{161}\)

Again, taking this “commons” regime to its theoretical extreme, the United States will be expending large amounts of money on development of drugs, entertainment, and intellectual discourse. The incentives to drug companies, artists, and other companies will be whittled away to nothing.


\(^{158}\) RIAA and MPAA statements, supra note 59.

\(^{159}\) See id.

\(^{160}\) See AIDS Article, supra note 2; see also Avian Flu Article, supra note 1.

\(^{161}\) The balance between the patent protection and the incentives to drug companies to commit significant funds to research and development is a more nuanced concept than there is room for here in this Note.
as the value of these creations is squandered by piracy abroad. The detractors of the “commons” view argue that lack of protection is the true deterrent to innovation and creativity as the general public takes advantage of the free information and provides no financial return to the original author/producer.

This competing view, sometimes called a “proprietary” model, for intellectual property rests on strong, complete protection of intellectual property with all control over the work given to the rights-holder.162 Here, monopolies over technological innovations have been considered “not only acceptable, but necessary” to ensure the development of newer technological innovations.163 Protection of innovations allows for full financial return on the intensive research and development that goes into these protected products (such as software and medicine).164 This represents a long-standing theory that “without an appropriate incentive, inventors will not create new innovations.”165

This proprietary model has been criticized on precisely this theory. While the information monopoly does protect innovations and inventors, the potential for corporate monopolies can result in injury to the consumer.166 Some of the negative fallout from such a proprietary model is reflected in the medical patent crises over AIDS and Asiatic bird flu.167 Pharmaceutical companies expend considerable amounts of research time and money into discovering and producing medicines to treat these epidemics (or potential epidemics). However, because of the monopoly afforded these companies during the medicine’s infancy, innovative treatments are unavailable to populations in poor nations or those without healthcare coverage in wealthy nations.

Some academics have gone so far as to suggest that control will lead to greater availability of “open” information and so arguing that the goals of the “commons” advocates will actually be better served by a more perfect regime of control and protection for intellectual property.168 In fact, this argument presumes access to information since it relies on the tenet that ideas, on which new ideas will build, still leak out into the

162. Weiser, supra note 155.
163. Id. at 577 (citing Richard R. Nelson & Sidney G. Winter, The Schumpeterian Tradeoff Revisited, 72 Am. Econ. R. 114, 144 (1982)).
164. Id. at 576.
165. Id. at 578.
166. Id. at 581 (citations omitted).
167. See text accompanying notes 1 and 2.
168. Wagner I, supra note 4, at 1000 (arguing that “because even perfectly controlled works nonetheless transfer significant information into the public domain . . . additional control likely to stimulate additional works”).
world from strictly protected works. While this may be a valid argument, the main focus of the WGIG and Access to Knowledge discussions has been working within the intellectual property regime to make sure that information (in its protected form) is distributed to rural and remote regions of the world. Only then can there be a truly level playing field for global innovation and education.

Ultimately, these views conflict over one significant use of information—fair use. Fair use is a general exception to intellectual property protection which allows for the public to create derivative works, building on the ideas of inventors in order to create newer innovations and continue the production cycle. Supporters of the “information commons” are in favor of wide exceptions for fair use, thereby valuing the right of the public to benefit from the work of others in creating new innovations. On the other hand, supporters of the “proprietary” model have even gone so far as to suggest that “fair use” is not reasonable in any incarnation. However, some preservation of fair use is arguably an important component of encouraging further innovations across the board. It is within these two extremes that the balance of intellectual property rights and freedoms is most needed.

Each of these possibilities represents extreme, but possible, regimes of intellectual property protection (or lack thereof). However, the future of the Internet and intellectual property need not be black and white. This Note proposes using Internet governance structures (within the purview of the WGIG) to create mechanisms to achieve a balance between the protection of IP rights-holders and the information interests of the developing world.

169. Id.
170. For more information, see Melville B. Nimmer & David Nimmer, 4 Nimmer on Copyright § 13.05 (2006).
171. See, e.g., Lessig, Free Culture, supra note 11, at 145.
173. Phillip Weiser’s “Competitive Platforms” model suggests another approach dealing specifically with software. While this is outside of the focus of this Note, there are scholars who are addressing the failings of a purely “commons” approach or a purely “proprietary” approach in content-specific ways. See generally Weiser, supra note 155.
174. Scholars have also suggested other “third way” approaches to the problem of the Internet and intellectual property, though often focusing on a particular subset of IP rights. Id.
B. Practical Perspectives: Digital Rights Management

Scholars, such as Lawrence Lessig, have argued that the Internet is actually regulated by several factors—norms, market, government, and code. “The code, or the software that makes cyberspace as it is, constitutes a set of constraints on how one can behave in cyberspace. The substance of these constraints vary, but they are experienced as conditions on one’s access to cyberspace.”179 It is through code, the software or technology behind the transfer of electronic information, that the balance between IP rights-holders and the developing world may be achieved.180

A growing tech industry is developing software which is directly designed to control the user’s access to information and cyberspace.181 Digital rights management (DRM) products, such as e-book software,182 broadcasting V-chips,183 and others184 are being developed in the private

175. LESSIG, CODE, supra note 51, at 86.
176. Id.
177. Id.; see also Weiser, supra note 155, at 543–44.
178. A very real example of the government regulating the Internet is ICANN as established by the U.S. Department of Commerce or the UN as a supra-governmental agency discussing Internet regulations through the WGIG.
180. The implementation of this code may require the support or influence of the other three factors. This Note suggests that the WGIG and WIPO have joint jurisdiction over the problems of intellectual property and the Internet and it is here that government structures (as well as norms and market forces) will play a role in the implementation of Internet governance structures to achieve balance in global intellectual property regulation.
182. One example of current reading software is Adobe’s e-book reader. While this software currently exists, the software does not always enable the same uses as a “real-world” or hard copy book. Sharing or reading aloud are sometimes restricted by the author. For example, the e-book version of the popular children’s book Alice in Wonderland includes a restriction against “reading aloud.” While this would seem a preposterous restriction on a “real-world” book, this restriction protects unauthorized recordings to be made and distributed for readers with special needs. See LESSIG, FREE CULTURE, supra note 11, at 152–53. Under the WIPO development agenda, uses such as making books available to readers with special needs would be allowed in order to further the resources available to the needy in developing nations. See Argentina/Brazil Proposal, supra note 124.
183. These broadcasting tools are examples of digital rights management which allow parents to control what television material their children are able to see (V-chip) or to monitor what people are viewing and recording—limiting the uses of broadcasted mate-
sector and seem to be the future of control (and access) on the Internet. These “technological enforcement measures” may also be useful for the copyright holder because they would “lower enforcement costs over time.” Essentially, DRM can help to protect works from infringing uses, but can also be used to help preserve the “fair use” of these works. Clear preservation of the fair use exception is vital to allow libraries, hospitals, and other institutions to be able to receive and disseminate important information to global citizens in remote countries.

Digital rights management, however, is not, in its current state, a perfect solution to the problems addressed in this Note. Scholars have suggested that there are serious questions about expectations of privacy in this new world of digital access controls. Because these technologies collect information about the uses of materials and report that information to the owners of the work, there is a potential for the collection of personal information about the user which could contribute to the growing fears about identity theft. As a solution, anonymity of the user must be programmed (or encoded) into access controls while collecting information about the uses of information.

Critics of DRM systems have also found that “almost all DRM solutions are themselves vulnerable to countermeasures (information about which is easily disseminated).” An example of this circumvention of DRM techniques is DeCSS, a program which was created by reverse engineering to “descramble” code on DVDs which restricted use of the discs to only those DVD players which could “descramble” the code. By descrambling this code on their own, users were able to copy and distribute the content of the DVD freely.

---

185. The government not only has the capacity to oversee the developments of a DRM system for intellectual property management, but it has been argued that the government is in the best position to regulate the use of information in this manner. See generally Williamson, supra note 181, at 312, 351–77.
186. Wagner I, supra note 4, at 1012.
188. Id.
189. Id.
190. Wagner I, supra note 4, at 1016.
192. Id.
At least in the United States, these concerns have been addressed by the legislature. The Digital Millennium Copyright Act was passed in 1999 in order to, inter alia, prevent the circumvention of anti-copying devices protecting intellectual property. This law does provide some exceptions for “fair use” of copyrighted material by libraries, law enforcement, and schools. However, there are still concerns that this law goes too far in protecting copyrights. In fact, some have argued that “fair use” should never be an excuse for circumventing a digital protection, leaving those who would fairly access the information without recourse. Even on this front, technology must also be used to promote a balance of rights between copyright holders and the consuming public.

Because information is valuable on the world market, other scholars have argued that the market should be regulated to make sure that these digital protections are effective. In order to address problems of anti-circumvention, scholars have suggested implementing legal bans on all products which would enable illegal uses or distribution of protected works. Because intellectual property is a global issue, this is a perfect example of where WIPO and WGIG need to come together to work on implementing regulations on both the Internet and intellectual property to ensure that rights are protected across the globe.

These issues collectively highlight the need for WIPO and WGIG to come together and work on policies and regulations on a global scale. Not only is the backing of the technology industry required to ensure compliance across the board (and to combat the temptation/capacity for illegal uses), but an international effort is required to ensure the protections of both copyright holders and the fair-using public consumers of information and technology.

For the most part, access to information in the developing world has distinct benefits. Sharing knowledge and copyrighted works will improve education across the globe. Some access structure for medical patents will make medicine available to sick and dying global citizens. However,
this sharing of information will also benefit rights holders. For instance, one scholar has argued that a greater release of information can be used to increase the market for related products.199 When students in a school in a Zimbabwean village access a digital library, a previously untapped market learns about authors and thinkers in other parts of the world. As remote nations become bigger world players in commerce, markets will open up for the protected works—books, medicines, movies, scientific articles.

Creative Commons200 is one example of an organization which is opening up the opportunities both for copyright holders and the fair-using public. Creative Commons has developed an alternate licensing structure for copyrighted materials which allows the author to explicitly preserve fair uses such as innovations and derivative works (or even free distribution) in a flexible structure observing both protections and freedoms.201 Creative Commons is quickly catching on amongst creators and authors, bringing more recognition and opening markets for those who are opting for these licenses.

With proper use of DRM, there is hope to achieve the sort of balance that would benefit both the protection concerns of the United States and the free spread of information in the developing world.202 WGIG, as the future of global Internet governance and the United Nations, is in a strong position to address these governance mechanisms in order to address the WIPO development agenda and satisfy the worries of the U.S. proprietary regime.

199. See Wagner I, supra note 4.
201. Wagner I, supra note 4 at 1032–33 (“Creative Commons seeks to assist owners in crafting ‘deeds’ to their works by drafting copyright licenses generally granting public access, but tailoring them to the particulars of the situation. Such efforts demonstrate the benefits that come with granting creators broad, flexible rights to control the uses of their inventions.”).
202. One example of such success is posted on the Creative Commons website. Pamela Jones, the founder of Groklaw, conducts complicated legal research and posts articles on her website. These articles are mostly licensed under Creative Commons and so the content has been mirrored at websites across the world. See Groklaw’s Pamela Jones, http://creativecommons.org/text/groklaw (last visited Apr. 4, 2007).
203. Williamson, supra note 181, at 323–24. (“In many senses, DRM schemes serve to enforce and protect the rights of all parties involved.”) (citations omitted); see also Joan Feigenbaum et al., Privacy Engineering for Digital Rights Management Systems, in SECURITY AND PRIVACY IN DIGITAL RIGHTS MANAGEMENT 76 (Tomas Sander ed. 2001).
C. Digital Archives—The Future Is Here

While DRM is perfected, global archives are popping up everywhere. Implementing these digital access mechanisms will mean that as remote countries become connected to the Internet, they will have access to a wealth of information as providers begin voluntarily making information available online.204 One example is the BBC’s recent decision to release news clips in an archival format, specifically for re-mixers and to encourage innovations with their own copyrighted material.205 Though the uses of these clips are limited to non-commercial uses, there is a clear delineation of the permissions to use the information to create at will outside of the commercial spectrum. This archive is already being used to foster creativity in the school settings in London through the London Children’s Film Festival.206 Children are able to download copies of the 1902 film The Little Match Seller and create their own original scores to be submitted for the festival.207

A more controversial example of the use of digital technology for archiving is the Google library archive project.208 Google has joined with national libraries to begin digitizing works which are in the public domain (where copyrights have expired) as well as books that are not yet in the public domain. Herein lies the controversy. Lawsuits have been filed over making these works alleging that this copying is “stealing” or piracy of books.209

Assessing Google’s activity and those by other companies is a difficult, but illustrative, example of the clash between traditional materials pro-

204. A whole host of publishers and other companies have made strides in the past few months to start digitizing books to take advantage of new technologies. These publishers are using the digital copies to maximize the exposure that their works get when users search for titles. Users will get to see digital excerpts from the books which will, hopefully, grab the interest of the reader. See Press Release, HarperCollins Publishers, HarperCollins Publishers Selects Newsstand, Inc. to Develop Global Digital Warehouse (Apr. 10, 2006), available at http://www.harpercollins.com/footer/release.aspx?id=445&b=&year=2006.


207. Id.


tected by traditional IP laws and the new digital world. No one would begrudge a small library in Nepal collecting public domain works to bring literacy to a village. However, the digitization makes the work far more susceptible to illegal copying—as the “proprietary model” supporters would argue. Instead, a digital archive is bringing resources of the developed world to remote developing nations on the same terms—a library which is instantly available as Internet infrastructure is laid down. It is through the development of digital rights management technology that this digital library could be accessed by readers in the hypothetical library in Nepal without violating the rights of the copyright holders.

WSIS has taken an important step towards resolution of these issues in its Action Line committee addressing access to information. The Golden Book shows other examples of global actors taking the problem into their own hands. The intellectual property concerns, nevertheless, are not being integrated into the process. By using these Internet governance mechanisms already in place and working within the structures established by the U.N. agencies in question—WGIG and WIPO—the rights of the accessing public and the copyright holders will both be protected in line with the “fair use” protections provided in traditional (and developing) intellectual property law.

V. CONCLUSION

The reality of working with behemoth government agencies is that government is slow and making changes at the global level is part of a painstaking process. This is increasingly impractical in a world where technology is developing faster everyday. Effective, efficient action is required with as much cooperation on the broad scale as possible in order to develop solutions that are practical and withstand the test of time, and protect and preserve rights across the board.

Global development depends on the application of technology solutions, such as DRM and public Internet archives, to age-old intellectual property conflicts. The flexibility of the Internet and its increasing availability around the world enables fast-paced answers to information problems. Implementing these solutions, as this Note has argued, will benefit developing nations by increasing access to information. Developed nations, and all intellectual property holders, will also benefit as the market for that information expands to encompass the entire globe. By bringing

210. See note 112 and accompanying text.
together the Internet governance policies developing in WSIS and WGIG, and the growing focus on development and intellectual property regulation in WIPO, the proper balance between information sharing across the globe and the protection of intellectual property rights-holders can be achieved.

*Kristin Delaney*
THE INTERNAL AFFAIRS RULE
AND THE APPLICABILITY OF U.S. LAW TO VISITING FOREIGN SHIPS

INTRODUCTION

Foreign ships have maintained a significant presence in American ports and waters since early in the nation’s history. This presence grew substantially during the last century, after many previously American vessels reflagged under flags of convenience. As a result, the vast majority of ships today calling on U.S. ports are foreign-flagged. For example, ninety-five percent of the passenger ships and seventy-five percent of the cargo vessels entering American waters are alien. These numbers underscore the desirability of developing a systematic approach for determining when domestic law applies to visiting foreign ships.

The Supreme Court recently tackled this issue in *Spector v. Norwegian Cruise Line Ltd.*, a case addressing whether Title III of the Americans with Disabilities Act (ADA) applies to foreign-flagged cruise ships in American waters. When the Court granted certiorari to resolve a pair of

---

1. For the purposes of this Note, foreign ships are simply those vessels whose nationality is not American. A vessel’s nationality, generally speaking, is determined by the country of her registry and under whose flag she sails. See *M/V Saiga (No. 2) (St. Vincent v. Guinea)*, 38 I.L.M. 1323, 1340 (Int’l Trib. L. of the Sea 1999).


4. A flag of convenience is a flag under which a vessel is registered in order to reduce operating costs and avoid the governmental regulations of the state of its beneficial ownership. Liberia and Panama are the most familiar flag of convenience states. *See Jane Marc Wells, Comment, Vessel Registration in Selected Open Registries, 6 MAR. LAW. 221, 221–27 (1981).*


conflicting rulings from the Fifth and Eleventh Circuits, some commentators believed that it would use the opportunity to clearly define when and how U.S. law applies to visiting foreign vessels. Unfortunately, this expectation went unfulfilled as much of the Court’s ruling, a plurality opinion with multiple subparts and voting blocs, has no binding precedential value.

Holding that the ADA’s basic requirements do apply to foreign cruise ships, a five-Justice majority eschewed as too “broad” the Fifth Circuit’s determination that the “clear statement rule” prohibits domestic legislation from applying to foreign vessels absent specific statutory language. However, a badly fractured Court authored four opinions and ultimately failed to agree on precisely how the rule should operate. Incorporating the “internal affairs rule” into its reading of the clear statement rule, a four-Justice plurality argued that the rule precluded the ADA and other generally applicable statutes from applying to the internal


12. The clear statement rule is a canon of statutory construction employed by American courts to resolve ambiguities in legislative language. In actuality, there is no one clear statement rule, but rather an array of such rules, all of which require a clear expression of congressional intent within the text of a statute before courts will interpret that statute in a way that encroaches on an area of traditional state authority, raises inconsistencies with international law, or impinges on intergovernmental immunities, just to name a few. William N. Eskridge & Philip P. Frickey, Quasi-Constitutional Law: Clear Statement Rules as Constitutional Lawmaking, 45 VAND. L. REV. 593, 598–611 (1992).

13. Spector, 545 U.S. at 130.


15. The internal affairs rule provides that foreign ships are not subject to the jurisdiction of the port state in matters touching only their internal order and discipline unless the peace or tranquility of the port state is disturbed. See infra note 36 and accompanying text.

16. Generally applicable statutes are laws whose legislative language contains words of universal application. Typical of most congressional legislation, these laws, if taken at
affairs of visiting foreign ships absent a clear expression of congressional intent. Furthermore, the plurality stated that the rule should be applied on a provision-by-provision basis, such that those statutory provisions not implicating a ship’s internal affairs would not trigger the rule and consequently would apply in full. Though the other five Justices did not share this vision of the clear statement rule, their disagreement amongst themselves ruined any chance that they might provide a majority standard. Thus, instead of definitively defining when U.S. law applies to visiting foreign ships, the Court left the question largely unanswered.

This Note examines the internal affairs rule and its interaction with the clear statement rule. Part I explores the jurisdictional issues inherent in applying domestic law to visiting foreign vessels and explains that international law resolves these issues through the internal affairs rule. Furthermore, Part I demonstrates that the United States adheres to the internal affairs rule, believing it to be a well-established international legal principle.

Part II examines how the Supreme Court has traditionally applied the internal affairs rule. Through consideration of the Court’s Jones Act and National Labor Relations Act precedent, this part aims to establish that the rule restrains the reach of legislation in some circum-

face value, would regulate any and all activities around the world. For example, in Lauritzen v. Larsen, 345 U.S. 571 (1953), Justice Jackson made the following observation of the Jones Act:

Unless some relationship . . . to our national interest is implied, Congress has extended our law and opened our courts to all alien seafaring men injured anywhere in the world in service of watercraft of every foreign nation -- a hand on a Chinese junk, never outside Chinese waters, would not be beyond its literal wording.

Id. at 577. Rather than give effect to the implausible results that follow from such literal statutory readings, American courts employ canons of statutory construction to establish the law’s proper reach. See id. at 577–79.

17. Spector, 545 U.S. at 130.
18. Id. at 138–39.
19. Commenting on Spector, one observer quipped, “[A]lthough the Court has rearranged the playing field, the rules of the game remain largely undefined.” Philip M. Berkowitz, ADA and Foreign Employers: New Guidance from the Court, N.Y.L.J., July 15, 2005, at 5.
20. Generally speaking, American courts recognize international law as constituting a kind of federal common law. See, e.g., The Paquete Habana, 175 U.S. 677, 700 (1900) (stating that “[t]he international law is part of our law” and holding that coastal fishing vessels are exempt from capture as prize of war under customary international law).
stances while having no such limiting effect in others. Part III returns to the Court’s decision in Spector, and discusses the plurality, concurring, and dissenting opinions. Part IV analyzes these opinions in light of Parts I and II, and concludes that the plurality provides the proper understanding of the clear statement and internal affairs rules. Finally, Part V synthesizes the main points of this Note and recapitulates the lesson learned from Spector.

I. JURISDICTIONAL CONFLICT AND THE INTERNAL AFFAIRS RULE

Any attempt by the United States to prescribe laws applicable to visiting foreign ships presents the risk of conflict between sovereigns. This conflict emanates from an overlap between two independent principles of jurisdiction, territoriality and nationality. The United States, as a port state, derives its authority to prescribe laws applicable to foreign vessels through the jurisdictional principle of territoriality. Stemming from a nation’s need to control activities within its geographic boundaries, territorial jurisdiction allows port states to regulate conduct that occurs or has effect within their waters. Vessels, on the other hand, are bound by the rules and regulations of their flag state through the jurisdictional

23. It is important to remark here that this Note is only concerned with laws containing words of universal application. See supra note 16. An express congressional statement that a statute does apply to visiting foreign ships effectively forecloses any question as to its applicability. For example, Congress amended the Seamen’s Act of 1915, Act of March 4, 1915, ch. 153, 38 Stat. 1164 (codified in part at 46 U.S.C. § 10313 (2006)), in 1920 to make it applicable to wage disputes between the owners and crewmembers of foreign ships docked in American ports. In Strathearn S.S. Co. v. Dillon, 252 U.S. 348 (1920), the Supreme Court upheld the applicability of the Act to a dispute between a British crewman and his British ship master for wages owed under a British employment contract for work aboard a British vessel. Given the Act’s express demarcation of its reach, the Court explained that any interpretation rendering the Act inapplicable to foreign ships and their crew would undermine Congress’s intent “to place American and foreign seamen on an equality of right.” Id. at 355.


25. See id; see also RESTATEMENT (THIRD) OF FOREIGN RELATIONS LAW § 402 cmt. b (1987) [hereinafter RESTATEMENT] (“Territoriality and nationality are discrete and independent bases of jurisdiction; the same conduct or activity may provide a basis for exercise of jurisdiction both by the territorial state and by the state of nationality of the actor.”).

26. A port state is a state that exercises control over a particular port. See McDorman, supra note 24, at 210.

27. RESTATEMENT § 402 cmt. h.

28. See id.
principle of nationality.\textsuperscript{29} By providing vessels with a comprehensive body of laws to govern their shipboard activities, nationality and the law of the flag play an essential role in maritime law.\textsuperscript{30} Ships entering foreign ports, therefore, find themselves subject to concurrent port state and flag state jurisdiction.\textsuperscript{31}

Though in principle port state jurisdiction is superior to flag state jurisdiction,\textsuperscript{32} “the legal certainty does not accurately reflect the tension.”\textsuperscript{33} Ships have a way of traveling from place to place. Thus if port states attempted to enforce their jurisdiction to the fullest extent upon every visiting vessel, then ships would be forced to navigate through an overwhelming regulatory patchwork.\textsuperscript{34} Similarly untenable, if every flag state could claim exclusive jurisdiction over its ships at all times, then port states would be completely defenseless against harmful vessel activities.\textsuperscript{35} Aiming for a reasonable solution to this jurisdictional dilemma, the major maritime nations developed the internal affairs rule, which provides that visiting foreign ships are not subject to port state jurisdiction in matters touching only upon the internal order and discipline of the ship unless those internal matters disturb the peace and tranquility of the port.\textsuperscript{36}

\begin{itemize}
\item \textsuperscript{29} See id. §§ 402(2), 502(2).
\item \textsuperscript{30} David F. Matlin, Note, \textit{Re-evaluating the Status of Flags of Convenience Under International Law}, 23 VAND. J. TRANSNAT’L L. 1017, 1021–22 (1991). While on the high seas, and until it enters another nation’s ports and internal waters, a ship is viewed conceptually as a floating extension of its flag state’s territory. See The Case of the S.S. Lotus (France v. Turkey), 1927 P.C.I.J. (Ser. A) No. 10, at 25; The Queen v. Anderson, [1868] L.R. 1 C.C.R. 161, 163 (U.K.) (“[A] ship, which bears a nation’s flag, is to be treated as part of the territory of that nation. A ship is a kind of floating island.”).
\item \textsuperscript{31} See \textit{RESTATEMENT} § 502 cmt. d (“The flag state’s jurisdiction also overlaps in some respects with the jurisdiction of the coastal state when the ship is in the territorial sea, contiguous zone, or a deepwater port of that state.”).
\item \textsuperscript{32} See 48 C.J.S. International Law § 23 (2006) (“[T]he jurisdiction of a nation to take enforcement action within its territory is absolute.”).
\item \textsuperscript{33} McDorman, supra note 24, at 211.
\item \textsuperscript{34} See \textit{Lauritzen}, 345 U.S. at 585 (observing that “there must be some law on shipboard, [and] that it cannot change at every change of waters”).
\item \textsuperscript{35} Noting the problems posed by this hypothetical situation, Chief Justice Marshall once remarked that “it would be obviously inconvenient and dangerous to society, and would subject the laws to continual infraction, and the government to degradation . . . [if foreign vessels] were not amenable to the jurisdiction of the country.” The Schooner Exchange v. McFadden, 11 U.S. (7 Cranch) 116, 144 (1812).
\end{itemize}
A. The Internal Affairs Rule: Two Theories and International Acceptance

While the finer points of enforcement policy may vary from nation to nation, a survey of the legal practices of the international community reveals that the internal affairs rule is widely accepted. Recently, in Re Maritime Union of Australia, the High Court of Australia recognized the internal affairs rule as a valid tenet of international law, and stated that the Australian Industrial Relations Commission should take the rule into account in deciding on remand whether certain Australian labor laws apply to the crews of foreign vessels. Similarly, the Federal Court of Canada ruled in Metaxas v. The Galaxias that whether severance pay was owed to several Greek sailors released from their employment aboard a Greek vessel in Vancouver was a question most appropriately answered by the law of the flag.

As regards the internal affairs rule, there are two competing theories to which nations subscribe, the French and the English. The French theory emerged from the consolidated cases of The Sally and The Newton. Both cases involved assaults by one seaman against another occurring on board American ships docked in France. The Counseil d'Etat began its analysis by remarking that “the rights of the [flag state] ought to be respected touching the internal discipline of the vessel in which the local

37. See, e.g., PHILIP C. JESSUP, THE LAW OF TERRITORIAL WATERS AND MARITIME JURISDICTION 144–91 (1927) (discussing the practices of Belgium, Brazil, Chile, England, France, Germany, Greece, Italy, Mexico, Norway, Peru, Portugal, Russia, and the United States); Sellers v. Maritime Safety Inspector, 1999] 2 N.Z.L.R. 44 (C.A.) (holding that the Maritime Safety Inspector could not, consistently with international law, require a Maltese ship, sailing from a New Zealand port, to carry radio and emergency equipment).

40. [1990] 2 F.C. 400 (Trial Div.) (Can.).
42. JESSUP, supra note 37, at 145.
44. Id. at 50.
authorities ought not to interfere unless . . . the peace and tranquility of
the port has been compromised." 46 Then, since in its opinion the assaults
failed to create a stir outside the ships themselves, the French court held
that the local tribunals had no jurisdiction to entertain criminal prosecu-
tions against the sailors. 47 The French theory thus asserts that port states
lack jurisdiction over the internal affairs of visiting foreign ships when
those matters do not disturb the port state’s interests. 48

Conversely, the English theory developed from the initial presumption
that port states possess absolute jurisdiction over visiting foreign ships. 49
This presumption provided the starting point for the court’s analysis in
The Queen v. Keyn, 50 a case involving a German captain arraigned on
criminal charges for running his vessel into an English ship while pass-
ing through English waters. 51 After deducing that any criminal conduct
that may have occurred would have taken place aboard the German ves-
sel, the court dismissed the indictment and held that the law of the flag
controlled. 52 The court explained its decision, in part, by declaring that
states may “choose[] to forego the exercise of her law over the foreign
vessel and crew, or exercise[] it only when they disturb the peace and
good order of the port.” 53 Accordingly, under the English theory, while
jurisdiction is fully vested in port states as a matter of right, it should not
be exercised over foreign vessels unless interests beyond those of the
ship and her crew are involved. 54

The French and English theories differ, therefore, only on the question
of whether port state jurisdiction over the internal affairs of foreign ves-
sels is yielded as a matter of right or discretion. Despite this abstract dif-
ference, there is no pragmatic distinction between the two theories since
both in practice decline jurisdiction when the interests of the port state
are not adversely affected. 55 It is this common practice that has solidified
into the customary rule of international law 56 that port states should not

47. Id.
48. JESSUP, supra note 37, at 154.
49. Id. at 169; see also Chung Chi Cheung v. The King, [1939] A.C. 160, 167 (U.K.)
(declaring that a port state’s jurisdiction is susceptible of no limitation not imposed by
itself).
51. Id. at 64.
52. Id. at 86.
53. Id. at 82.
55. CHURCHILL & LOWE, supra note 36, at 66; JESSUP, supra note 37, at 192.
56. Customary international law is derived from the consistent practice of states act-
ing out of a belief that international law requires them to act that way. See MARK W.
exercise jurisdiction over the internal affairs of visiting foreign ships unless those affairs disturb the peace of the port.  

B. The Internal Affairs Rule in the United States

The United States is generally regarded as a subscriber to the English theory. Thus, paralleling the English view, the Supreme Court has held that visiting foreign ships subject themselves to American jurisdiction as a condition of entry. Furthermore, this jurisdictional authority being absolute, the Court has taken the position that the United States is entitled to unqualifiedly enforce its laws against foreign vessels. Nevertheless, recognizing that international law compels a more moderate ap-

JANIS, AN INTRODUCTION TO INTERNATIONAL LAW 44–48 (4th ed. 2003). Indeed, consistent state practice is the “best evidence” that a rule has become part of customary international law. RESTATEMENT § 103 cmt. a.

57. It is worth mentioning here that this customary rule was codified in part in the United Nations Convention on the Law of the Sea, Dec. 10, 1982, 1833 U.N.T.S. 397 [hereinafter UNCLOS]. Article 27(1) of UNCLOS resolves conflicts with respect to criminal jurisdiction on board visiting foreign ships in the following way:

The criminal jurisdiction of the coastal State should not be exercised on board a foreign ship passing through the territorial sea to arrest any person or to conduct any investigation in connection with any crime committed on board the ship during its passage, save only in the following cases:

(a) if the consequences of the crime extend to the coastal State;

(b) if the crime is of a kind to disturb the peace of the country or the good order of the territorial sea;

(c) if the assistance of the local authorities has been requested by the master of the ship or by a diplomatic agent or consular officer of the flag State; or

(d) if such measures are necessary for the suppression of illicit traffic in narcotic drugs or psychotropic substances.

UNCLOS, art. 27(1).

58. See Charteris, supra note 43, at 58; JESSUP, supra note 37, at 191.

59. E.g., Patterson v. Bark Eudora, 190 U.S. 169, 179 (1903); United States v. Diekelman, 92 U.S. 520, 525 (1875).

60. Chief Justice Marshall delivered the most famous pronouncement of the nation’s complete and plenary authority over its waters in The Schooner Exchange v. McFadden, 11 U.S. (7 Cranch) 116, 136 (1812), where he declared that “[t]he jurisdiction of the nation within its own territory is necessarily exclusive and absolute . . . [and] is susceptible of no limitation not imposed by itself.”

61. See, e.g., Cunard S.S. Co. v. Mellon, 262 U.S. 100 (1923) (holding that visiting foreign vessels are barred from possessing any alcohol in American ports and waters, regardless of whether flag state law requires its ships to store such spirits, since the Eighteenth Amendment and the National Prohibition Act state a strong public policy against such possession).
approach to jurisdictional enforcement, the Court has stated that applications of domestic legislation to conduct occurring on board foreign ships should comply with the internal affairs rule.\footnote{McCulloch v. Sociedad Nacional de Marineros de Honduras, 372 U.S. 10, 19–21 (1963); \textit{see also} Lauritzen v. Larsen, 345 U.S. 571, 585–86 (1953); United States v. Rodgers, 150 U.S. 249, 260 (1893).} This was first explained by the Court over a century ago in \textit{Wildenhus’s Case}.\footnote{Mali v. Keeper of the Common Jail (\textit{Wildenhus’s Case}), 120 U.S. 1 (1887).} There, Chief Justice Waite made the following observation:

From experience, however, it was found long ago that it would be beneficial to commerce if the local government would abstain from interfering with the internal discipline of the ship, and the general regulation of the rights and duties of the officers and crew towards the vessel, or among themselves. And so by comity it came to be generally understood among civilized nations that all matters of discipline, and all things done on board, which affected only the vessel, or those belonging to her, and did not involve the peace and dignity of the country, or the tranquility of the port, should be left by the local government to be dealt with by the authorities of the nation to which the vessel belonged as the laws of that nation, or the interests of its commerce should require. . . . Such being the general public law on this subject, treaties and conventions have been entered into by nations having commercial intercourse, the purpose of which was to settle and define the rights and duties of the contracting parties with respect to each other in these particulars, and thus prevent the inconvenience that might arise from attempts to exercise conflicting jurisdictions.\footnote{Id. at 12.}

An exhaustive list of all ship matters considered “internal” has never been compiled. Notwithstanding this, it can be pointed out that the internal affairs rule’s reasoning has been invoked to relinquish jurisdiction in cases involving terms of employment and wages,\footnote{See Lopes v. S.S. Ocean Daphne, 337 F.2d 777 (4th Cir. 1964); The Albani, 169 F. 220 (E.D. Pa. 1909); The Becherdass Ambaidass, 3 F. Cas. 13 (D. Mass. 1871) (No. 1,203); Saunders v. The Victoria, 21 F. Cas. 539 (E.D. Pa. 1854) (No. 12,377); The Pacific, 18 F. Cas. 942 (S.D.N.Y. 1830) (No. 10,644); Thomson v. The Nanny, 23 F. Cas. 1104 (D.S.C. 1805) (No. 13,984); Willendson v. Forsoket, 29 F. Cas. 1283 (D. Pa. 1801) (No. 17,682). Where the complaining seaman has been badly mistreated, however, courts have exercised jurisdiction in the interests of justice. \textit{See} The Salomoni, 29 F. 534 (S.D. Ga. 1886); Weiberg v. The St. Oloff, 29 F. Cas. 597 (D. Pa. 1790) (No. 17,357). Jurisdiction has also regularly been exercised where the complaining crewmember is an American citizen. \textit{See} The Neck, 138 F. 144 (W.D. Wash. 1905); The Alnwick, 132 F. 117 (S.D.N.Y. 1904); The Falls of Keltie, 114 F. 357 (D. Wash. 1902).} collective bargain-
personal injury, and ship discipline. Being of considerable interest to the flag state and of little interest to the United States, American courts, through the internal affairs rule, have recognized the primacy of flag state law in these matters in order to avoid unnecessary conflicts. On the other hand, where the interests of the United States are significantly affected by internal ship matters, American courts have not shied away from exercising jurisdiction. Indeed, it is well-settled that jurisdiction over foreign vessels and their crew will be asserted when their activities offend “the peace or good order of the port either literally . . . or in some constructive sense.” This “port disturbance” exception has been construed broadly, allowing American authorities full discretion to determine for themselves what circumstances warrant an invasion of a foreign ship’s internal affairs. Intervention has been most common in cases involving criminal activity and customs or immigration violations.


68. See Ex parte Anderson, 184 F. 114 (D. Me. 1910). Where ship discipline rises to the level of torture, however, jurisdiction will be exercised in the interests of justice. See Bolden v. Jensen, 70 F. 305 (D. Wash. 1895).


70. See RESTATEMENT § 512 cmt. c (observing that “coastal states usually have little interest” in the internal affairs of foreign ships).


73. Id. at 66.

74. A broad interpretation of what constitutes a “port disturbance” is necessary, because “[c]ircumstances alter cases and a dispute which at times might have no effect on shore, at other times might have serious local consequences.” JESSUP, supra note 37, at 180.

75. For example, criminal convictions were upheld in five cases against Italian sailors who, upon learning of the U.S.’s entry into WWII, purposefully destroyed the propulsive machinery of their ships in order to scuttle them in their ports. See Polonio v. United
tions. Additionally, in their efforts to protect the peace of the port, American authorities have sometimes imposed obligations on foreign vessels which extend beyond their stay. Requiring that all oil tankers have double hulls, barring any and all possession of liquor during Prohibition, and enforcing antitrust laws on foreign shipping companies are three prominent examples.

Before moving on, it should be noted that the Court’s explanation in *Wildenhus’s Case* that the internal affairs rule was founded on considerations of comity has resulted in some confusion. Several commentators have understood this to mean that the rule is not a true tenet of international law, but rather a courtesy extended by the United States to other sovereign nations. A closer reading of Chief Justice Waite’s opinion, however, reveals that the Court did not leave the internal affairs rule to rest on comity alone, nor did it hold that treaties were necessary to confirm the rule’s existence. To be sure, the Court stated that the rule arose from comity originally and indicated that it might be better protected by


78. See Cunard S.S. Co. v. Mellon, 262 U.S. 100 (1923). In a decision strikingly similar to *Cunard S.S. Co.*, the Seychelles Supreme Court in *R. v. Fayolle*, 91 Int’l L. Rep. 384 (Sey. 1971), upheld the conviction of a foreign ship’s master under the Green Turtle Protection Regulations for possessing in port green turtle meat fished outside the territorial waters of Seychelles, reasoning that the Regulations stated an absolute policy against green turtle meat possession.


80. Comity can be defined as “a willingness to grant a privilege, not as a matter of right, but out of deference and good will.” BLACK’S LAW DICTIONARY 267 (6th ed. 1990).

81. See, e.g., Maitlin, supra note 30, at 1025.
treaties in the interest of preventing misunderstanding.\textsuperscript{82} Nevertheless, as the law stood in the mind of the Court, the internal affairs rule was existent and established in “the general public law.”\textsuperscript{83} Furthermore, this understanding of the internal affairs rule’s status as an international legal principle has been affirmed in subsequent Court decisions.\textsuperscript{84}

II. Application of the Internal Affairs Rule

Like other principles of statutory interpretation, the internal affairs rule is a canon of constructive caution.\textsuperscript{85} The rule, functionally speaking, presumes that legislation is not intended to regulate the internal affairs of visiting foreign ships unless there is a clear statement from Congress to the contrary.\textsuperscript{86} Furthermore, as the following examination of the Supreme Court’s Jones Act and National Labor Relations Act jurisprudence makes clear, the internal affairs rule does not take an all-or-nothing approach to the applicability of domestic law. Rather, while the rule may work to restrict a particular statute’s reach in one instance, it might not in another.

A. The Jones Act

The Jones Act allows injured seamen to bring actions for damages when their injuries are suffered in the course of employment and are due to the negligence of their employer.\textsuperscript{87} In \textit{Uravic v. F. Jarka Co.},\textsuperscript{88} an American stevedore\textsuperscript{89} was injured on board a German vessel while he

\textsuperscript{82} Wildenhus’s Case, 120 U.S. at 12.
\textsuperscript{83} Id. General public law and international law are synonymous. See \textit{Janis}, supra note 56, at 2.
\textsuperscript{84} See supra note 62 and accompanying text. The U.S. Government appears to have adopted this position as well, having relied on this understanding of the internal affairs rule to “protest[] the assertion of jurisdiction over controversies [involving American vessels], by local magistrates in the territory of a foreign State with which no adequate agreements had been concluded.” \textit{CHARLES CHENEY HYDE, INTERNATIONAL LAW CHIEFLY AS INTERPRETED AND APPLIED BY THE UNITED STATES} 739–40 (2d rev. ed. 1945).
\textsuperscript{85} See \textit{Garside}, supra note 71, at 790–91 (briefly sketching and comparing the internal affairs rule, the Charming Betsy canon, and the presumption against extraterritoriality).
\textsuperscript{87} 46 U.S.C. app. § 688(a). For a detailed discussion of liability and recovery under the Jones Act, see Brian J. Miles, \textit{The Standard of Care in a Seaman’s Personal Injury Action—Has the Jones Act Been Slighted?}, 13 TUL. MAR. L.J. 79 (1988).
\textsuperscript{88} 282 U.S. 234 (1931).
\textsuperscript{89} A stevedore is a laborer who loads and unloads vessels docked in port. \textit{THE AMERICAN HERITAGE DICTIONARY} 1701 (4th ed. 2000). Five years prior to \textit{Uravic}, in
was unloading it in New York harbor. This injury subsequently led to his death. Suit was brought under the Act by the stevedore’s administratrix who asserted the negligence of a fellow-servant as the cause of the injury. Speaking for a unanimous Court, Justice Holmes reasoned that the internal affairs rule was inapplicable under the circumstances since the stevedore’s activities went “beyond the scope of discipline and private matters that do not interest the territorial power.” Indeed, Holmes remarked that it would be “extraordinary” to deny American legal protections to citizens only “momentarily” engaged on board foreign ships. The decision of the New York court holding the Act inapplicable was accordingly reversed.

A separate set of facts yielded a different result in *Romero v. International Terminal Operating Co.* In *Romero*, a Spanish seaman brought suit against his Spanish employer after he was injured on board a Spanish ship docked in Hoboken, New Jersey. The seaman was loading a cargo of wheat onto the ship when a wire cable suddenly slipped, severing his left leg and causing multiple fractures to his right leg. Adhering to the internal affairs rule, the Court held that providing compensation for the injured seaman and regulating the liability of the shipowner were the concerns of the state of their common nationality, and that the United States had no interest in intruding its own policies into that relation.

---

90. *Urvic*, 282 U.S. at 238.
91. *Id*.
93. *Urvic*, 282 U.S. at 238. Interestingly, just eleven years earlier, Urvic’s cause of action would have failed at the outset since maritime law followed the fellow-servant rule, which barred recovery for injuries caused by other crewmembers. See, e.g., *The Osceola*, 189 U.S. 158 (1903). With the passage of the Jones Act in 1920, however, Congress abolished the fellow-servant rule as a defense to shipboard negligence liability. Joel K. Goldstein, *The Osceola and the Transformation of Maritime Personal Injury Law: Some Propositions About the Case and Its Propositions*, 34 Rutgers L.J. 663, 722 (2003).
95. *Id*.
99. *Id* at 356.
100. After the accident, the seaman was taken by ambulance to a hospital in Hoboken where he underwent treatment for the next eight months. Currie, *supra* note 68, at 1 (pulling these facts from the trial record and appellate briefs).
The fact that the injury occurred in American waters was dismissed as a “wholly fortuitous circumstance.” Consequently, unlike its determination in Uravic, the Court in Romero found that the circumstances of the case triggered the internal affairs rule and rendered the Jones Act inapplicable.

B. The National Labor Relations Act

The National Labor Relations Act (NLRA) was enacted to guarantee the right of workers to organize and to bargain collectively without fear of management reprisal. To make the NLRA effective, Congress established the National Labor Relations Board (NLRB), a quasi-judicial body empowered to resolve labor disputes. Such disputes erupted with recurrent regularity in ports throughout the United States during the 1950s, as labor unions, upset over depressed pay scales and working conditions, picketed foreign vessels and attempted to organize their crews.

Benz v. Compania Naviera Hidalgo, S.A. provided the Supreme Court with its first opportunity to consider the NLRB’s jurisdiction over labor relations aboard visiting foreign ships. In Benz, a Liberian vessel was docked in Portland, Oregon, when its crew suddenly went on strike demanding higher wages and improved working conditions. The crewmen appointed an American labor union as their collective bargaining representative, and the union and others promptly began picketing the vessel. After the shipowner brought suit in federal district court seeking to enjoin the union’s activities, the union claimed that only the NLRB could consider the controversy. The Court, however, held that the NLRB lacked jurisdiction under the internal affairs rule, remarking

101. See Romero, 358 U.S. at 384.
102. Id.
103. See id.
104. See National Labor Relations Act, 29 U.S.C. § 151 (setting forth Congress’s findings and declaration of policy with respect to the Act).
105. See id. §§ 153–156.
106. See Note, supra note 69, at 503.
108. Suit was brought under the Labor Management Relations Act, 29 U.S.C. §§ 141–197 (2006), an amendment to the NLRA.
110. Id. at 140.
111. Id. at 141. Adjudication of certain activities falling within the ambit of the NLRA are subject to the exclusive jurisdiction of the NLRB. See Stephen F. Befort, Demystifying Federal Labor and Employment Law Preemption, 13 LAB. LAW. 429, 430–34 (1998).
that “Congress did not fashion [the NLRB] to resolve labor disputes between nationals of other countries operating ships under foreign laws.” 112

The rule was similarly employed in McCulloch v. Sociedad Nacional de Marineros de Honduras 113 to find that the NLRB did not have authority under the NLRA to order an election for the unionization of foreign seamen recruited in Honduras to serve aboard Honduran vessels. In McCulloch, the National Maritime Union 114 petitioned the NLRB for certification as the bargaining agent for the crewmen. 115 The vessels were operated by a wholly owned subsidiary of an American corporation, but their labor relations were governed by several provisions of the Honduran Labor Code. 116 When the NLRB decided to assert jurisdiction and ordered an election to be held among the seamen, 117 both the shipowners and the Honduran labor union which claimed to represent the crewmen filed suit seeking an injunction. 118 The Court unanimously upheld their claims and found the NLRB lacked jurisdiction. 119 Rejecting the “balancing of contacts” test 120 employed by the NLRB, the Court reasoned that such a test would invariably “inquire into the internal discipline and order” of the ships in contravention of “the well-established rule of international law that the law of the flag state ordinarily governs the internal affairs of a ship.” 121

Distinguishing the factual elements that triggered the internal affairs rule in Benz and McCulloch, the Court upheld the applicability of the NLRA in International Longshoremen’s Local 1416 v. Ariadne Shipping

112. Benz, 353 U.S. at 143.
114. Founded in 1937, the National Maritime Union (NMU) organized seamen and waterfront workers, and was one of the most radical labor unions in the United States. The NMU lent support to various anti-colonial and international revolutionary struggles, and called for racial integration of the East Coast shipping industry. Ahmed A. White, Mutiny, Shipboard Strikes, and the Supreme Court’s Subversion of New Deal Labor Law, 25 BERKELEY J. EMP. & LAB L. 275, 311–17 (2004).
116. Id. at 14. Under these Honduran labor laws, only certain specified Honduran unions were permitted to act as bargaining representatives for seamen serving on board Honduran vessels. Id.
118. McCulloch, 372 U.S. at 12.
119. Id. at 21.
120. The balancing of contacts test employed by the NLRB consisted of weighing several factors, such as the nationality of the parties, the place of the dispute, and the beneficial ownership of the shipping corporation, to determine whether jurisdiction was appropriately asserted. See United Fruit Co., 134 N.L.R.B. at 288–90.
In *Ariadne*, American unions picketed several foreign ships docked in Florida to protest substandard wages paid to American longshoremen. The owners of the vessels attempted to enjoin the picketing in state court, but a unanimous Supreme Court held that the NLRA preempted the field and that the NLRB thus had exclusive jurisdiction to consider the matter. Noting that *Benz* and *McCulloch* each involved situations where NLRA regulation “would necessitate inquiry into the ‘internal discipline and order’ of a foreign vessel,” the Court remarked that the longshoremen’s “casual” involvement with the foreign ships in no way implicated their internal affairs.

### III. Spector v. Norwegian Cruise Line Ltd.

*Spector* was the first case in twenty years that presented the Court with a domestic statute arguably implicating the internal affairs of a visiting foreign ship. The Court resolved little in the case, however, as the Justices varied greatly in their readings of the aforementioned Jones Act and National Labor Relations Act cases.

#### A. Background

The plaintiffs in *Spector* were disabled individuals who purchased round-trip tickets for cruises aboard two Bahamian-flagged Norwegian Cruise Line (NCL) ships. The cruises departed from Houston, Texas and sailed to foreign ports of call. During their vacations, the plaintiffs discovered that many of the ships facilities—public restrooms, swimming pools, restaurants, elevators, preferred cabins—were inaccessible to them. Furthermore, the plaintiffs discovered that NCL main-

---

123. *Id.* at 196–97.
124. *Id.* at 200.
125. *Id.* at 198 (quoting *McCulloch*, 372 U.S. at 19).
126. *Id.* at 199–200.
127. Similar to *Benz* and *McCulloch*, the Court in *Windward Shipping (London) Ltd. v. American Radio Ass’n*, 415 U.S. 104 (1974), held that picketing of two foreign-flagged vessels by American unions to protest substandard wages paid to foreign seamen was an activity affecting the ships’ internal operations, and thus was not covered by the Labor Management Relations Act.
128. NCL is a Bermudian corporation offering cruise vacations departing from and returning to United States ports. NCL Corporation Ltd., Annual Report (Form 20-F), at 14 (Mar. 28, 2006).
129. *Spector*, 545 U.S. at 126.
130. *Id.*
tained several policies which were applicable to them but not to other passengers. For instance, NCL charged higher fares and special surcharges on account of the plaintiffs’ disabilities and required them to remain subject to removal from the ship should their presence endanger the “comfort” of the other passengers.132

Upset over their treatment, the plaintiffs filed suit claiming NCL discriminated against them in violation of Title III of the ADA.133 Title III prohibits discrimination based on disability in places of “public accommodation”134 and in “specified public transportation services.”135 To that end, Title III requires covered entities to make “reasonable modifications in policies, practices, or procedures” to accommodate disabled persons,136 and to remove physical barriers where “readily achievable.”137

While allowing many of the claims to go forward, the district court dismissed those pertaining to barrier removal on the grounds that the governing ADA compliance regulations did not specifically cover cruise ships.138 The Fifth Circuit affirmed in part and reversed in part, declaring that the clear statement rule precluded any application of federal law to visiting foreign ships without specific evidence of congressional intent.139 Finding nothing in either the ADA’s text or legislative history to indicate that Congress had thought about Title III’s application to foreign cruise ships, the Fifth Circuit held that Title III was wholly inapplicable.140

B. The Supreme Court’s Opinions

1. The Holding of the Court

The Supreme Court reversed the Fifth Circuit’s judgment, criticizing it as being “inconsistent with the Court’s case law and with sound princi-
plexes of statutory interpretation.” In particular, the five-Justice majority held that the Fifth Circuit’s articulation of the clear statement rule was too expansive, and that the rule could not be understood as blanketly applying to foreign ships in their entirety. The majority found that cruise ships fall within Title III’s definitions of “public accommodation” and “specified public transportation services.” Thus, said the majority, the ADA’s basic requirements on accommodation of disabled persons presumptively apply to all cruise ships operating in American waters, whether foreign or domestic. The majority then noted that any structural modifications which would bring a ship into non-compliance with its international obligations would not be “readily achievable” and therefore not be required by Title III.

Thus, while acknowledging that some of the plaintiffs’ claims were potentially barred, the majority rejected the Fifth Circuit’s determination that the ADA was altogether inapplicable to foreign cruise ships. More importantly for purposes of this discussion, the majority unambiguously held that the Fifth Circuit misinterpreted the clear statement rule. The majority could not agree, however, on how exactly the rule worked. Already divided on the issue of whether the clear statement rule disqualified any application of the ADA to foreign cruise ships, the Court diverged further as the majority splintered on the question of what limitations the rule imposed.

141. *Spector*, 545 U.S. at 130.
142. Justices Breyer, Ginsburg, Kennedy, Souter, and Stevens constituted the majority.
143. *Spector*, 545 U.S. at 130.
144. Id. at 129.
145. Id. at 132–33.
146. There is an open question here as to what exactly the majority would exclude under Title III’s “readily achievable” standard, since “international legal obligation” was not well defined. See id. at 135. A simple solution might be to say that an international legal obligation is the same as an obligation of international law. If indeed these terms are equivalent, then rules of customary international law, and therefore the internal affairs rule, would be included in the definition. See, e.g., *JANIS*, supra note 56, at 41–55 (explaining that international custom is a source of international law). It is unlikely that this is what the majority intended, however, since those parts of Justice Kennedy’s opinion dealing with the internal affairs rule did not command a majority vote. The majority probably intended international legal obligations to mean “treaty obligations,” since they cite as an example the International Convention for the Safety of Life at Sea, Nov. 1, 1974, 32 U.S.T. 47, 164 U.N.T.S. 113. See *Spector*, 545 U.S. at 135.
147. Id. at 135–36.
148. The dissenting Justices—O’Connor, Rehnquist, and Scalia—answered this question in the affirmative. For a discussion of their view, see infra Part III.B.4.
2. The Plurality Opinion

Three of the Justices from the majority—Kennedy, Souter, and Stevens—equated the clear statement rule with the internal affairs rule. Calling it “the internal affairs clear statement rule,” the plurality declared that the rule, absent a clear expression of congressional intent to do otherwise, would simply preclude those applications of the ADA which affect the internal affairs of visiting foreign ships.\textsuperscript{149} For instance, they reasoned that NCL’s policies of charging disabled persons higher fares and reserving the right to remove them “have nothing to do with [the] ship’s internal affairs,” so the rule would not prevent Title III from applying to those discriminatory policies in full force.\textsuperscript{150} On the other hand, the accessibility claims seeking removal of coamings\textsuperscript{151} and other structural barriers “likely would interfere with the internal affairs of foreign ships,” thus requiring a clear statement from Congress to achieve that end.\textsuperscript{152} Additionally, the plurality stated that the clear statement rule did not mandate an all-or-nothing approach to the applicability of a particular statute. Rather, the rule exempts from the reach of domestic legislation only those applications which implicate “the internal order and discipline of the vessel, rather than the peace of the port.”\textsuperscript{153}

3. The Concurring Opinion

The remaining two Justices from the majority—Breyer and Ginsburg—did not ascribe any further significance to the clear statement rule. They stated that the rule’s sole purpose lays in avoiding conflicts with international obligations, and since the majority’s interpretation of the ADA’s “readily achievable” standard fulfilled that purpose, nothing more was required.\textsuperscript{154} In their view, the extent to which Title III implicated the internal affairs of foreign ships was completely irrelevant.\textsuperscript{155} The concurring Justices concluded that domestic law should apply to visiting foreign vessels, even when it affects the ship’s internal affairs, so

\textsuperscript{149}. Spector, 545 U.S. at 138. Justice Kennedy stated that “[t]he relevant category for which the Court demands a clear congressional statement . . . consists not of all applications of a statute to foreign-flag vessels but only those applications that would interfere with the foreign vessel’s internal affairs.” \textit{Id.} at 132.

\textsuperscript{150}. \textit{Id.} at 133.

\textsuperscript{151}. Coamings are the raised edges around a ship’s doors and other openings designed keep water out. \textit{THE AMERICAN HERITAGE DICTIONARY} 353 (4th ed. 2000).

\textsuperscript{152}. Spector, 545 U.S. at 135.

\textsuperscript{153}. \textit{Id.} at 130. It should be noted here that Justice Thomas agreed with this assessment of the rule’s operation and joined in this part of the plurality’s opinion.

\textsuperscript{154}. \textit{Id.} at 144 (Ginsburg, J., concurring).

\textsuperscript{155}. \textit{Id.} at 145.
long as “there is good reason to apply our own law” and no potential for “international discord” exists.156

4. The Dissenting Opinion

The dissenting Justices, like the plurality, thought that the clear statement rule incorporated the internal affairs rule when the application of American law to visiting foreign ships was at issue.157 Indeed, they understood the internal affairs rule to be an established rule of international law.158 Despite this shared belief, the dissent disagreed with the plurality and majority on two points. First, the dissent stated that any application of domestic legislation implicating the internal affairs of a ship would trigger the clear statement rule, thus preventing that application.159 Consequently, the dissent declined to accept the majority’s invitation to construe Title III's “readily achievable” standard to avoid conflicts with international obligations.160 As they saw things, the clear statement rule achieved that result on its own.161 Second, parting ways with the plurality, the dissenting Justices thought that the ADA and other statutes could not be selectively applied under the rule.162 In their view, since some applications of Title III would affect the internal affairs of foreign cruise ships, the absence of a clear congressional statement rendered the entire statute inapplicable.163

IV. ANALYSIS

Spector’s holding with respect to the clear statement rule is rather limited. The Court clearly rejected the position taken by the Fifth Circuit that the rule applies to visiting foreign ships in toto. However, this is as far as the majority went. Left unresolved were questions regarding both the scope of the rule and its operation. First, an open issue remains as to whether the clear statement rule tracks the internal affairs rule, or

156. Id.
157. Id. at 149 (Scalia, J., dissenting) (“The plurality correctly recognizes that Congress must clearly express its intent to apply its laws to foreign-flag ships when those laws interfere with the ship’s internal order.”).
158. Id. at 150.
159. Id. at 155.
160. Id. at 153–54.
161. Id. at 154.
162. Id. at 156.
163. Id. at 157 (“Since some applications of Title III plainly affect the internal order of foreign-flag ships, the absence of a clear statement renders the statute inapplicable—even though some applications of the statute, if severed from the rest, would not require a clear statement.”).
whether it instead seeks only to avoid international conflicts. Second, assuming that the clear statement rule’s scope is defined by the internal affairs rule, the issue of whether the rule operates on a provision-by-provision or an all-or-nothing basis remains undecided. Both of these questions will be considered in turn, and it will be argued that the plurality’s position provides the proper standards.

A. The Scope of the Rule: Confirming the Internal Affairs Rule

Though not reflected in the Court’s holding, a majority of the Justices in Spector confirmed the vitality of the internal affairs rule. Seven Justices held the view that the clear statement rule requires specific congressional intent before domestic statutes will be construed to interfere with the internal affairs of visiting foreign ships. This interpretation of the clear statement rule’s scope is correct since it confers on flag states the deference over regulation of their ships’ internal matters long afforded them by both the United States and the international community.

1. Correcting the Concurrence

The idea of the “internal affairs clear statement rule” was not unanimously accepted, however, as the two concurring Justices declared that the clear statement rule’s scope went no further than avoiding conflicts with international legal obligations. Supporting their conclusion, the concurring Justices stated that Benz and McCulloch simply held that congressional statutes should not be construed so as to violate international obligations when other interpretations remain available. Reading the internal affairs rule into the clear statement rule, for them, then, was an over-inclusive means of avoiding such conflicts since doing so would preclude applying domestic legislation in cases where no conflict exists.

This position is problematic for several reasons. First, the concurring Justices’ conclusion rests on an incomplete interpretation of Benz and McCulloch. Those cases did indeed hold the NLRA inapplicable to the foreign ships in question for fear of provoking international discord.

164. Justices Kennedy, O’Connor, Scalia, Souter, Stevens, Thomas, and Chief Justice Rehnquist all held this view. See supra Parts III.B.2 and III.B.4.
165. See supra Part I.
166. See supra Part III.B.3; see also supra note 146 (discussing the ambiguity of the term “international legal obligation”).
167. Spector, 545 U.S. at 143–44 (Ginsburg, J., concurring).
168. The Court in Benz stated that it would not hold the NLRA applicable to the foreign ship’s labor relations without a clear congressional statement, since to do otherwise would “run interference in such a delicate field of international relations.” Benz, 353 U.S.
but hardly by paying lip service to the vessels’ internal affairs. Rather, Benz and McCulloch identified the NLRA’s intrusive effect on the ships’ internal labor affairs as being the root cause of the conflict,\(^{169}\) which is why McCulloch cited as authority the “rule of international law that the law of the flag state ordinarily governs the internal affairs of a ship.”\(^{170}\)

Second, in arriving at their conception of the clear statement rule, the concurring Justices appear to conflate two related canons of statutory construction, the internal affairs rule and the Charming Betsy canon. Derived from Chief Justice Marshall’s statement in \textit{Murray v. Schooner Charming Betsy}\(^ {171}\) that “an act of Congress ought never to be construed to violate the law of nations if any other possible construction remains,”\(^ {172}\) the Charming Betsy canon’s principal aim is to harmonize domestic law with international law.\(^ {173}\) The majority’s interpretation of Title III’s “readily achievable” standard to avoid conflicts with international legal obligations is a good example of the Charming Betsy canon at work.\(^ {174}\) However, while the plurality thought more was needed from the clear statement rule since foreign ships were involved, the concurring Justices did not. The plurality was right. Relying solely on the Charming Betsy canon under these circumstances is impractical given that port state law and flag state law are not always amenable to harmonization.\(^ {175}\)

When the two laws conflict, courts must sometimes abandon harmonization efforts and focus instead on determining which law will prevail. Here is where the internal affairs rule shows its worth, since its raison d’être is to answer that question.

As an aside, it is possible that the “errors” just discussed were an attempt by the concurring Justices to strip the internal affairs of visiting foreign ships of the partial immunity to which they are currently enti-

---

169. See supra Part II.B.
172. \textit{Id.} at 118.
174. See supra note 147 and accompanying text.
175. Such is the case whenever port state law and flag state law truly conflict, a frequent occurrence in foreign vessel actions. Indeed, in many of the Supreme Court’s cases involving the application of domestic law to visiting foreign ships, the governments of the vessels involved have filed briefs alerting the Court to the existence of a conflict of laws. See Symeon Symeonides, \textit{Maritime Conflicts of Law from the Perspective of Modern Choice of Law Methodology}, 7 MAR. LAW. 223, 224–25 (1982).
tled. While this “solves” the problems just mentioned, such an approach creates new problems of its own. First, pragmatically speaking, its excessive self-interestedness is counterproductive. To be sure, no one expects port states to sacrifice their own policies out of sheer altruistic concern for those of flag states, and international law is clear that flag state jurisdiction yields to port state jurisdiction. However, it is unjust and unfair for vessels to expect that their conditions might vary with each port at which they call. Moreover, this move towards territorial exclusivity will not only wreck havoc on the international shipping industry, but will also “invite retaliatory action from other nations.”

Second, from an international law perspective, this shift is illegal. Unless stripping foreign ships of their partial internal affairs immunity is attained by treaty or other agreement, one state’s unilateral disregard for that immunity is contrary to international law. That this act might signal the beginning of a new general practice among nations, thus eventually altering customary international law, will not save that act from the stigma of illegality. If such a shift in international law is indeed desired, then concerted action by formal agreement should be pursued rather than unilateral action in defiance of existing law.

176. That this may have been the concurring Justices’ ulterior motive is signaled by the advocacy of some commentators that the United States take such an approach. See, e.g., R. Tali Epstein, Comment, Should the Fair Labor Standards Act Enjoy Extraterritorial Application?: A Look at the Unique Case of Flags of Convenience, 13 U. Pa. J. Int’l Bus. L. 653 (1993) (arguing that stripping foreign ships of their partial internal affairs immunity will benefit the American labor force).

177. See supra note 32 and accompanying text.

178. See supra note 34 and accompanying text.


180. See JANIS, supra note 56, at 41–44 (describing the binding nature of customary international law).


182. That the United States once held this opinion and acted upon it is revealed by instructions sent from Washington to the American Minister in Madrid on October 28, 1852, during discussions arising out of the treatment accorded to an American vessel, The Crescent City, by the officials of Spain in Cuba. In those instructions it was said:

You will state that this government does not question the right of every nation to prescribe the conditions on which vessels of other nations may be admitted into her ports. That, nevertheless, those conditions ought not to conflict with the received usages which regulate the commercial intercourse among civilized nations. That those usages are well known and long established and no nation can disregard them without giving just cause of complaint to all other nations whose interests would be affected by their violation.
2. The Internal Affairs Rule and Modern Conflict of Laws Methodology

Another argument in favor of the clear statement rule incorporating the internal affairs rule is that the internal affairs rule is consistent with modern conflict of laws\textsuperscript{183} theory, “interest analysis”\textsuperscript{184} in particular. When port states seek to apply their laws to visiting foreign ships, true conflicts routinely arise. Such is the case whenever port state law and flag state law differ and the policies underlying each have a genuine claim to application.\textsuperscript{185} In this situation, interest analysis holds that port state courts should normally apply port state law.\textsuperscript{186} Because of the concurrent flag state interests, however, interest analysis also asserts that port state courts may take “a more moderate and restrained interpretation” of port state law.\textsuperscript{187} This means that port state courts may recognize the applicability of port state law, but, in deference to the contrary interest of the flag state, will decline to assert for the port state an interest in having its law applied.\textsuperscript{188}

This is all that the internal affairs rule requires. The rule presumes that port states do not intend to interfere with matters that are primarily of concern only to the ship and its flag state. If, however, those matters affect the peace of the port, then that presumption is effectively rebutted. Thus, the internal affairs rule does not seek to simply subserviate port state law to flag state law, but rather to mitigate the tension when the two conflict. This sensitivity to the substantive aspects of the two laws demonstrates the internal affairs rule’s compatibility with prevailing conflict of laws theory.

\textsuperscript{183} H.R. EXEC. DOC. NO. 86, 33d Cong., 1st Sess., at 24 (1853).

\textsuperscript{184} Conflict of laws is that branch of jurisprudence which deals with disputes subject to the conflicting laws of two or more states. BLACK'S LAW DICTIONARY 299–300 (6th ed. 1990).

\textsuperscript{185} Developed by Professor Brainerd Currie, interest analysis seeks “to ensure that the law which is applied in the majority of cases will be the one whose application in a particular context will serve the purposes for which that law was created.” Gregory E. Smith, Choice of Law in the United States, 38 HASTINGS L.J. 1041, 1047 (1987).

\textsuperscript{186} See id.

\textsuperscript{187} See Brainerd Currie, Notes on Methods and Objectives in the Conflict of Laws, 1959 DUKE L.J. 171, 178.

\textsuperscript{188} A judicial operation of this sort was famously performed in Bernkrant v. Fowler, 360 P.2d 906 (Cal. 1961), where Justice Roger Traynor, after observing that California would be constitutionally justified in applying its Statute of Frauds to deny a claim against the estate of a local domiciliary, chose not to attribute to the California legislature an intent to preclude enforcement of an oral agreement executed in Nevada for the benefit of Nevada domiciliaries when Nevada had no similar statute in accordance with its policy of vindicating the reasonable expectations of its people.
B. The Rule’s Operation: A Provision-by-Provision Argument

The debate within the Supreme Court over whether the clear statement rule applies on a provision-by-provision or all-or-nothing basis proved more contentious than that over the rule’s scope, as the Justices closely split four to three favoring the former approach over the latter.\(^{189}\) The plurality reasoned that since the NLRA was held applicable in *Ariadne* but not in *Benz* and *McCulloch*,\(^{190}\) the same provision-by-provision approach should be taken with respect to the ADA.\(^{191}\) Specifically, the plurality thought that the clear statement rule would restrain only those Title III provisions which touched upon the vessel’s internal affairs, leaving all the remaining provisions effective.\(^{192}\)

The dissenting Justices believed that the plurality’s approach encompassed an “utterly implausible” view of congressional intent.\(^{193}\) They stated that Congress could not have had separate intent with respect to each provision of Title III, and that Congress either did or did not have foreign ships in mind when it enacted the statute.\(^{194}\) Consequently, since Title III did not contain a clear statement as to its applicability to foreign vessels, the fact that some of its provisions implicated their internal affairs necessarily rendered the entire statute inapplicable.\(^{195}\) The dissent, therefore, took the plurality to task for attempting to force foreign ships into compliance with Title III requirements that were never intended to apply to them.\(^{196}\)

While the dissenting Justices’ argument is persuasive at a glance, closer inspection reveals that it rests on an inconsistency in reasoning. The dissent, like the plurality, threw its support behind the internal affairs rule, a rule that presumes Congress does intend its laws to regulate the non-internal affairs of foreign vessels. Thus admitting Congress’s intent to regulate the non-internal affairs of foreign vessels, the dissent cannot fairly argue that a provision-by-provision approach, whereby non-internal matters are automatically subject to regulation while internal matters require a clear statement, somehow fails to regard that intent.

---

189. *See supra* Parts III.B.2 and III.B.4. As an aside, the two concurring Justices, though not in favor of the internal affairs rule conception of the clear statement rule, noted that they agreed with the plurality’s provision-by-provision approach to the rule. *Spector*, 545 U.S. at 143 n.1 (Ginsburg, J., concurring).

190. *See supra* Part II.B.


192. *Id.* at 138–39.

193. *Id.* at 156 (Scalia, J., dissenting).

194. *Id.*

195. *Id.* at 157.

196. *Id.* at 156.
Rather than “delusional,” it is entirely rational to believe that Congress intends some provisions of its statutes to apply to visiting foreign ships while intending exemptions with regards to others. Moreover, the plurality’s adoption of the provision-by-provision approach adheres not only to previous Court practice with respect to this clear statement rule, but with other clear statement rules as well.

V. CONCLUSION

The clear statement rule is the canon of statutory construction employed by American courts to determine the applicability of U.S. law to visiting foreign ships. Due to the ever-increasing number of foreign vessels calling on American ports, there is a great need for a more definite and authoritative restatement of that rule. Indeed, while considering the applicability of the ADA to foreign cruise ships, the Court in Spector was well aware that its decision would have important implications for the applicability of other statutes as well. It is therefore unfortunate that the Justices could not come to an agreement on the rule’s scope or operation.

While the uncertainty over the proper standards to apply to the clear statement rule may not have been fully resolved in Spector, the case nevertheless provides an answer. By assimilating the internal affairs rule and the provision-by-provision approach into its articulation of the clear statement rule, the plurality adhered to the traditional American viewpoint and exercised due regard for international law. There is reason,

197. Id.

198. The dissenting Justices are likely correct that Congress did not “in fact” have individualized intent with respect to each provision of Title III. See id. However, it defies reality to expect Congress to have foreseen and considered every contingency arising under the statute. Given this situation, the determination that courts need to make is “what Congress would have wished if these problems had occurred to it.” Bersch v. Drexel Firestone, Inc., 519 F.2d 974, 993 (2d Cir. 1975) (Friendly, J).

199. See supra Part II.

200. For example, the Court took the same provision-by-provision approach when it applied the “presumption against extraterritoriality” to a couple of cases involving the Seamen’s Act of 1915. Compare Strathearn S.S. Co. v. Dillon, 252 U.S. 348 (1920) (holding that Section 4 of the Act does apply extraterritorially to invalidate wage advancements), with Sandberg v. McDonald, 248 U.S. 185 (1918) (holding that Section 11 of the Act does not apply extraterritorially to invalidate wage advancements).

201. In rejecting the Fifth Circuit’s articulation of the clear statement rule, the plurality noted the adverse consequences that such a broad reading of the rule would have for Title II of the Civil Rights Act of 1964, Pub. L. No. 88-352, 78 Stat. 241 (codified as amended in scattered section of 42 U.S.C.). See Spector, 545 U.S. at 132.
therefore, to expect that their opinion will be followed in future cases going forward.

Nathaniel Kunkle

---

* B.A. Johns Hopkins University (2001); J.D. Brooklyn Law School (expected 2007). I would like to thank the following people: the staff of the Brooklyn Journal of International Law for their assistance in the preparation of this Note; the staff of the Brooklyn Law School Library, especially Jean Davis and Victoria Szymczak for their expertise; my parents, Kevin and Shirley Kunkle, and brother, Jesse Kunkle, for their constant encouragement; and my wife-to-be, Krista Margies, for listening to tales of The Becherdass Ambaidass.
NATURAL DISASTERS AND THE RESPONSIBILITY TO PROTECT: FROM CHAOS TO CLARITY

I. INTRODUCTION

“The devastation was overwhelming.” . . . There were severely injured people everywhere, dead bodies wedged under fallen houses, people with deep cuts from the sheets of glass and metal that had churned through the rushing water like sharks’ teeth. Dive boats, under which the tsunami had passed as an unremarkable two-foot swell on the open water, came back piled high with bodies that had been swept out to sea. It was chaos, and there was no authority, no rescue agency, [and] no civil structure to deal with a catastrophe of this scale.¹

Natural disasters, either resulting from or exacerbated by natural phenomena,² such as earthquakes, volcanoes, tsunamis, hurricanes, and famines, are of increasing worldwide concern.³ Affecting nearly two-hundred million people per year over the past two decades⁴

² A United Nations (U.N.) study on disaster-preparedness classified disasters into two categories: man-made and natural. Man-made disasters, which are attributable to accidental, negligent, or deliberate human activity include: (1) civil disturbances, such as riots or demonstrations; (2) situations involving conventional, nuclear, biological, chemical, or guerilla warfare; (3) refugee situations; and (4) accidents involving transportation, collapse of buildings and dams, mine disasters, or technological failures, such as pollution, chemical leaks, or nuclear accidents. Natural disasters result from the effects of natural phenomena, which are characterized as: (1) meteorological, such as storms (cyclones, hailstorms, hurricanes, tornadoes, typhoons, and snowstorms), cold spells, heat waves, droughts, and famine; (2) topological, which include earthquakes, avalanches, landslides, and floods; and (3) biological, such as insect swarms and epidemics of communicable diseases. RUTH M. STRATTON, DISASTER RELIEF: THE POLITICS OF INTERGOVERNMENTAL RELATIONS 22–23 (1989). The categorical distinction between man-made and natural disasters is not rigid, since allegedly “natural” disasters are not purely “natural,” but result from multiple causes, in part affected by “mankind’s relationship with the environment.” PETER MACALISTER-SMITH, INTERNATIONAL HUMANITARIAN ASSISTANCE: DISASTER RELIEF ACTIONS IN INTERNATIONAL LAW AND ORGANIZATION 3 (1985); Jacqueline P. Hand, Disaster Prevention Presentation, From SCJIL Symposium 2003, 1 SANTA CLARA J. INT’L L. 147 (2003). This Note focuses on purportedly natural disasters, either caused or exacerbated by natural phenomenon, which generate emergency situations of grave human suffering and physical destruction.
and causing billions of dollars in property damage worldwide, natural disasters “know no country, [no] region, no boundaries.” All over the world, people have endured their devastating consequences, such as loss of life and livelihoods, damaged infrastructure, and economic costs.

Such devastation occurred on December 24, 2004 when the Indian Ocean tsunami killed nearly 240,000 people, displaced more than one-million people, and devastated the infrastructure of twelve countries in South Asia and East Africa. This transnational natural disaster devastated developing and vulnerable countries and demonstrated the need for humanitarian assistance from non-affected States (the international concern for disasters due to worldwide vulnerability from environmental degradation, climate change, and geological hazards).

4. Id. at 1.

5. U.N. OFF. FOR THE COORDINATION OF HUMANITARIAN AFF. [OCHA], IRIN Web Special: Disaster Reduction and the Human Cost of Disaster, June 2005, at 3, available at http://www.irinnews.org/webspecials/DR/default.asp (noting that the U.N. Bureau for Crisis Prevention and Recovery asserted that from 1980 to 2000, approximately seventy-five percent of the world’s population have suffered at least once from natural disasters. In 2003 alone, there were about 700 natural disasters, which resulted in nearly 75,000 deaths and caused $65 billion in damage.). The U.N. established the OCHA to strengthen the response and effectiveness of the U.N.’s humanitarian operations in complex emergencies and natural disasters.


9. In the context of natural disasters, vulnerability refers to “[t]he conditions determined by physical, social, economic, and environmental factors or processes that increase the susceptibility of a community to the impact of hazards.” Hyogo Framework for Action, supra note 3, at n.1; see also Strengthening Emergency Relief, supra note 8, ¶ 5 (noting that the December 26th tsunami struck vulnerable countries, such as Indonesia, Somalia, and Sri Lanka, which were already undergoing long-standing complex crises).

10. For the purposes of this Note, the term “humanitarian assistance” is used interchangeably with “humanitarian relief” and “disaster relief.” These terms refer to the provision of relief for suffering natural disaster victims. Disaster relief consists of three categories: (1) assistance in kind, which refers to commodities and materials, such as food, clothing, medicine, and hospital equipment; (2) financial contributions, which refer to cash expenses; and (3) services of trained personnel, which refer to the operational, ad-
tional community of States). Another notable disaster is Hurricane Katrina, which killed hundreds of people and ruined the lives of thousands more throughout the southern United States. Although Hurricane Katrina’s physical impact was confined within U.S. borders, its aftermath attracted global concern because the United States needed disaster assistance from the international community of States. These natural disasters have demonstrated the international community of States’ vital role in facilitating humanitarian assistance to disaster-affected States for catastrophic transnational and national disasters.

Generally, past natural disasters have demonstrated the international community of States’ willingness and generosity in providing relief to disaster-affected States. However, the provision of humanitarian assistance to disaster-affected States can be problematic. Since there are no international legal obligations on States regarding disaster relief, it could potentially follow that neither the international community of States nor disaster-affected States have any responsibilities concerning humanitarian assistance for natural disasters. The lack of international legal obligations pertaining to disaster response is troubling, particularly when disaster-affected States delay or prevent the provision of relief or when the international community of States inadequately or improperly provides humanitarian assistance. Such situations have arisen in past disaster relief operations due to international law’s disregard for the

11. For the purposes of this Note, “States” refer to Nation-states.
12. See discussion infra Part III.A.2.ii.
15. See discussion infra Part III.A.
16. MACALISTER-SMITH, supra note 2, at 56.
17. LYNN H. STEPHENS & STEPHEN J. GREEN, Conclusion: Progress, Problems and Predictions, in DISASTER ASSISTANCE: APPRAISAL, REFORM AND NEW APPROACHES 293, 295 (Lynn H. Stephens & Stephen J. Green eds., 1979) (noting that a common theme among reports and studies on natural disasters concerned “the problems created in effectively providing relief in natural disasters, when affected governments consciously delay or prevent the delivery of assistance”); see, e.g., discussion infra Part III.A.
responsibilities of all States concerning humanitarian assistance for natural disasters.\textsuperscript{19}

The Indian Ocean tsunami and Hurricane Katrina are perfect examples of catastrophic natural disasters that have exposed the imperfect state of international disaster relief. In particular, these disasters have revealed concerns pertaining to the provision of humanitarian assistance. They have raised issues concerning the general responsibilities of disaster-affected States and the international community of States regarding natural disasters, such as: (1) whether all States have a responsibility to warn of impending disasters; (2) whether the international community of States has a responsibility to unconditionally provide humanitarian assistance and whether disaster-affected States have a responsibility to accept needed disaster relief; and (3) whether all States have a responsibility to rebuild disaster-stricken communities.\textsuperscript{20} Issues relating to these responsibilities are important because of the severity of natural disasters worldwide\textsuperscript{21} and criticism regarding the inadequacy of disaster relief responses.\textsuperscript{22} International consensus on the legal responsibilities of all

\begin{itemize}
  \item \textsuperscript{19} MACALISTER-SMITH, supra note 2, at 56.
  \item \textsuperscript{20} STEPHEN GREEN, INTERNATIONAL DISASTER RELIEF: TOWARD A RESPONSIVE SYSTEM 29–30 (Thomas Quinn & Michael Hennelly eds., 1977) (explaining the current international disaster relief system as consisting of four major elements: (1) the U.N.; (2) private organizations; (3) donor governments; and (4) the international media. These elements comprise a complex and chaotic disaster relief system that responds ad hoc to a multitude of natural disasters differing in geographical locations and circumstances.); see discussion infra Part III.
  \item \textsuperscript{21} See OFF. OF THE HIGH COMMISSIONER FOR HUM. RTS. [OCHR], Fact Sheet No. 2 (Rev. 1): The International Bill of Human Rights (June 1996), available at http://www.unhchr.ch/html/menu6/2/fs2.htm [hereinafter The International Bill of Human Rights]; GREEN, supra note 20, at 20–21 (expressing concern over the geographical distribution of natural disasters that occurred in the 1970s, which primarily hit developing regions of the world, such as poor countries with significant political, social, and economic problems that are the least capable of single-handedly dealing with natural disasters).
  \item \textsuperscript{22} See, e.g., discussion infra Part III.A.2; see also David P. Forsythe, Diplomatic Approaches to the Political Problems of International Relief in Natural Disasters, in DISASTER ASSISTANCE: APPRAISAL, REFORM AND NEW APPROACHES, supra note 17, at 267, 268 (highlighting some problems of international disaster relief and recognizing the need for its improvement). Forsythe quotes a 1977 U.S. Department of State Action Memo:

\begin{quote}
The international “system” for disaster [relief] needs to be improved. It now suffers from bad coordination among national governments, voluntary agencies, and UN bodies, as well as from political resistance within some countries to admit disaster or to distribute relief fast and equitably.
\end{quote}
\begin{flushright}
Id.
\end{flushright}
\end{itemize}
States would help clarify the conceptual problems relating to State responsibilities with respect to disaster relief. International law should impose on all States certain responsibilities to foster international cooperation during relief operations and to ensure the effective provision of humanitarian assistance to disaster victims.

Part II of this Note resolves doctrinal issues concerning State sovereignty and international human rights, which are relevant to disaster relief operations. It describes the doctrine of State sovereignty and its evolution as a result of the emerging recognition of international human rights. It further explains two theories that have surfaced in the context of humanitarian intervention: “the forfeiture of sovereignty” and “the responsibility to protect.” These theories attempt to resolve the tension between sovereignty and international human rights. Part III addresses how the doctrine of “the responsibility to protect” can be applied to natural disasters. It highlights criticisms of prior disaster relief efforts and demonstrates instances where disaster-affected States and the international community of States embraced certain responsibilities—before, during, and after natural disasters occurred—despite the lack of any imposed legal obligations. It exemplifies how the responsibility to protect applies to natural disasters, but recognizes the difficulty in finding legal authority for this assertion. Part IV argues that recognizing the human rights of disaster victims within international human rights law is essential to promoting the responsibility to protect doctrine, and justifying its

23. J.W. Samuels, The Relevance of International Law in the Prevention and Mitigation of Natural Disasters, in Disaster Assistance: Appraisal, Reform and New Approaches, supra note 17, at 247–48 (stating that the approach by the U.N. Disaster Relief Organization and the International Red Cross Movement to establish a regulatory system for disaster relief neglects to mention the State’s obligation to provide and accept relief).

24. Id. at 250 (recognizing the need to establish the general principles of disaster relief before developing detailed law concerning its operational aspects).

25. While this Note focuses on humanitarian assistance for natural disasters, it is necessary to assess theories that attempt to assimilate sovereignty and human rights under the doctrine of humanitarian intervention because they are relevant to the discourse regarding humanitarian assistance during natural disasters.

26. Macalister-Smith, supra note 2, at 6 (noting that legal instruments, in the form of multilateral and bilateral declarations, resolutions, and treaties, have had an insufficient impact on international relief operations, and recognizing the necessity of evaluating the international humanitarian activities of States, government organizations, private agencies, and other international actors during disaster relief operations to assess the developing standards regarding the international community’s responsibility towards disaster victims).
applicability to natural disasters. It proposes codifying the rights of natural disaster victims to create legal obligations on all States with respect to disaster relief. Part V posits that an international disaster relief treaty, premised on the responsibility of all States to protect the rights of natural disaster victims, would foster international cooperation during disaster relief operations. This agreement would promote a more effective system of international disaster relief.

II. SOVEREIGNTY AND HUMAN RIGHTS: RESOLVING THE TENSION

A. The Evolving Doctrine of State Sovereignty

The doctrine of State sovereignty is “an almost sacred principle” that is a fundamental concept of international law. According to the doctrine of sovereignty, States are independent entities that can exercise supreme political authority over their territory. States can control movement across their borders, independently make foreign policy choices, and reject unwanted intrusion by other States. Under the traditional view of sovereignty, States may shape and determine their own policies with respect to the treatment of their citizens and control over their domestic affairs without interference from other States.

27. Richard H. Ullman, Introduction to International Disaster Relief: Toward a Responsive System, supra note 20, at 1, 3 (recognizing the viability of regarding the provision of disaster relief as a human right by focusing on the right of all people to receive satisfaction of their basic survival needs).

28. Green, supra note 20, at 12–14 (noting that the problem with international disaster relief operations is that they do not always adequately respond to the needs and interests of disaster victims, and suggesting that laws, such as international treaties, may improve the protection of disaster victims because such laws could address fundamental management problems during relief operations).


30. Richard N. Haass, Director, Policy Planning Staff, Remarks to the School of Foreign Service and the Mortara Center for International Studies: Sovereignty: Existing Rights, Evolving Responsibilities (June 14, 2003), available at http://www.state.gov/s/p/rem/2003/16648.htm (arguing that for over two centuries, the doctrine of sovereignty has fostered the emergence of representative governments, “the formation of international organizations, and the development of international law”).

31. Id.; David J. Scheffler, Toward a Doctrine of Humanitarian Intervention, 23 U. Tol. L. Rev. 253, 259–60 (1992); Gorman, supra note 10, at 58 (defining “sovereignty” as a basic principle that “confers on recognized States the right to conduct their domestic and foreign policies without interference from the outside. It calls for [S]tates to recognize their mutual territorial integrity and independence. It allows [S]tates to limit themselves, but not to be limited against their will by other [S]tates.”).

This traditional understanding of sovereignty has changed considerably due to decolonization following World War II, the establishment of the United Nations (U.N.), and the U.N. Charter’s introduction of the right to self-determination in 1951.33 These events have transformed the traditional view of sovereignty into a more modern view of sovereignty that acknowledges the sovereign status of States as conditional upon those States recognizing obligations to their people.34 Currently, international law recognizes the more modern view of sovereignty as “the people’s sovereignty rather than the sovereign’s sovereignty.”35

B. The Development of International Human Rights

The U.N. Charter obligates U.N. Member States (Member States) “to take joint and separate action in co-operation with the Organization for the achievement of the purposes set forth in Article 55,” which promotes respect for human rights and fundamental freedoms.36 Although the U.N. Charter has not expounded upon what these rights and freedoms entail, it has compelled Member States to promote the concept of human rights for all people.37


35. W. Michael Reisman, Sovereignty and Human Rights in Contemporary International Law, 84 AM. J. INT’L L. 866, 869 (1990); see also Scheffer, supra note 31, at 260 (explaining how the concept of sovereignty has evolved into one that is not solely premised on the power of States, but considers sovereignty powers in other international actors such as ethnic groups, refugees, displaced people, and regional and international organizations. The evolution of sovereignty has occurred, in part, as a result of the emergence of a multitude of international treaties and conventions fostering the protection of human rights at the expense of limiting national sovereignty.). Furthermore, the emergence of the U.N. as an international diplomatic authority has played a huge part in fueling the evolving concept of sovereignty among the international community. Mithi Mukherjee, Justice, War, and the Imperium: India and Britain in Edmund Burke’s Prosecutorial Speeches in the Impeachment Trial of Warren Hastings, 23 LAW & HIST. REV. 589, 593 (2005).


On December 10, 1948, the U.N. General Assembly (General Assembly) confirmed its pledge to human rights by adopting the Universal Declaration of Human Rights (UDHR). The UDHR provides an exhaustive list of civil and political rights, including: the right to life, liberty, and security; the right not to be held in slavery; and the right to freedom of movement and residence. Scholars have argued that the UDHR, though adopted as a non-binding resolution, has been accepted as customary international law, thus representing legal obligations developed from the general acceptance and consistent practice of States.

In addition to the UDHR, the General Assembly adopted a series of U.N. Human Rights Covenants in 1966, which include the International Covenant on Civil and Political Rights (ICCPR), the International Covenant on Economic, Social and Cultural Rights (ICESCR), the Optional Protocol to the ICCPR, and the Second Optional Protocol to the

38. The General Assembly approved the UDHR upon the recommendation of the U.N. Human Rights Commission. The UDHR was the first worldwide endeavor to identify human rights standards. GORMAN, supra note 10, at 124.


40. See, e.g., Filartiga v. Pena-Irala, 630 F.2d 876, 882 (2d Cir. 1980) (recognizing that certain rights, such as the right to be free from torture, have become part of customary international law under the UDHR); see also GORMAN, supra note 10, at 124 (contending that although the UDHR is not legally binding, the UDHR has established human rights norms that many States have incorporated into their constitutions or statutes, thereby granting the UDHR domestic legal significance); Reisman, supra note 35, at 869 (asserting that the UDHR has acquired customary international law status and its ideals are upheld by several regional agreements).

41. See Statute of the I.C.J. art. 38(1)(b) (stating that “a general practice accepted as law” is international custom); MACALISTER-SMITH, supra note 2, at 53 (1985) (explaining that the “creation of a customary rule requires both the existence of a general practice of States and a second constitutive element which is the opinio juris, or the acceptance by States of the practice as law”); Michael Y. Kieval, Note, Be Reasonable! Thoughts on the Effectiveness of State Criticism in Enforcing International Law, 26 Mich. J. Int’l L. 869, 872–73 (2005) (explaining that customary international law results from established norms that have evolved from State practice).

42. The ICCPR reiterated the following rights: the freedom of movement, the right to be recognized as persons before the law, the freedom of religious and political expression, and the right to nondiscrimination. GORMAN, supra note 10, at 113.

43. The ICESCR emphasized the right to have a “healthier, safer, and more satisfying life.” Such a right includes: to be free from hunger, to subsistence, to work, to have safe and healthy working conditions, to join unions, to social security, to primary education, and to undertake in cultural life. Id. at 113–14.
ICCPR. The adoption of these covenants codified individual human rights principles.

Collectively, the U.N. Charter, the UDHR, and the U.N. Human Rights Covenants, comprise an International Bill of Human Rights, which provides a human rights framework for the international community. While the acknowledgment and significance of specific human rights may vary among individual States, the general respect for the principle of international human rights has made it a universally accepted norm under customary international law.

C. Resolving the Tension: The Forfeiture of Sovereignty and the Responsibility to Protect

The escalating concern over catastrophic natural disasters and the lack of clarity pertaining to States’ responsibilities to disaster-stricken people necessitates assessing the assimilation of two principles: sovereignty and human rights. Despite worldwide concern over natural disasters, the international community is still fundamentally comprised of sovereign States whose primary concern is to safeguard their sovereignty. Herein arises the dilemma. When disaster-affected States are

44. The International Bill of Human Rights, supra note 21.
45. GORMAN, supra note 10, at 124 (explaining that the adoption of the ICCPR, Protocol to the ICCPR, and the ICESCR were efforts to codify and enforce human rights principles).
47. See Alex Y. Seita, Globalization and the Convergence of Values, 30 CORNELL INT’L L. J. 429, 447 n.60 (1997).
50. Samuels, supra note 23, at 263 (noting the misunderstanding and need for resolution regarding issues arising from the obligation of governments to provide relief and the obligation of other governments to receive relief).
51. See, e.g., Tampere Gives Birth to a New Convention, supra note 6, at 39, 49 (noting that certain countries raised concerns about the possible loss of sovereignty with the ratification of the Tampere Convention on the Provision of Telecommunication Resources for Disaster Mitigation and Relief Operations (Tampere Convention), an international treaty lowering regulatory barriers to facilitate telecommunications equipment and personnel in the aftermath of a disaster); see also discussion infra Part IV.B for more information about the Tampere Convention.
either unable or unwilling to provide adequate relief to disaster victims, the international community of States must be able to provide humanitarian assistance without compromising the sovereignty of disaster-affected States. The examples proffered throughout this Note demonstrate that this is not always the case.

Many theories have emerged to reconcile State sovereignty with human rights in justifying humanitarian intervention: the right of foreign States and international organizations to take intervening and coercive action, specifically military action, in a State’s domestic affairs if its citizens are being subject to treatment that “shocks the conscience of mankind.” Two theories: “the forfeiture of sovereignty” and “the responsibility to protect” deserve recognition because they justify humanitarian assistance during natural disaster relief operations.

According to the forfeiture of sovereignty, a State temporarily forfeits its sovereignty when it allows gross violations of human rights to occur, thus contravening its citizens’ rights.

[W]here the government is not in control or the controlling authority is unable or unwilling to create the conditions necessary to ensure rights, and gross violations of the rights of masses of people result, sovereignty in the sense of responsible government is forfeited and the international community must provide the needed protection and assistance.

52. Samuels, supra note 23, at 248–49.
53. GORMAN, supra note 10, at 48 (defining “humanitarian intervention” as a controversial legal doctrine referring to the right of a [S]tate . . . to intervene in the domestic affairs of another [S]tate to prevent it from persisting in flagrant abuse of its own population . . . [such as] persistently violat[ing] the rights of its citizens . . . in a way that ‘shocks the conscience of mankind’

55. Id. at 212; Michael L. Burton, Note, Legalizing the Sublegal: A Proposal for Codifying a Doctrine of Unilateral Humanitarian Intervention, 85 GEO. L. J. 417, 435–36 (1996) (noting that “a [S]tate is endowed with a defeasible right of sovereignty, contingent upon some minimum standard of treatment of its subjects. Should [S]tate action fall below this threshold—for example, by flagrantly violating the human rights of those within its borders—the [S]tate forfeits its sovereignty entirely and thus becomes subject to external intervention.”).
This theory maintains that a State’s primary obligation is to protect its citizens. When the State cannot—thereby allowing severe human rights violations to ensue against its people—the State’s sovereignty is deemed forfeited. The forfeiture of sovereignty views a State’s sovereign status as conditional upon its ability to protect the human rights of its people. Under this theory, intervention from the international community of States is acceptable when a State is perceived to have forfeited its sovereignty by violating human rights.

The second theory is the responsibility to protect. The International Commission on Intervention and State Sovereignty (ICISS) proposed this theory, which is a more positive version of the forfeiture of sovereignty because it reconciles, rather than surrenders, sovereignty with responsibility. Such reconciliation becomes evident from assessing the development of the concept of “State responsibility” and its emergence among the international community.

The responsibility to protect developed from the State responsibility concept, which the International Court of Justice (I.C.J.) established in the Corfu Channel Case in 1949. The I.C.J.’s judgment recognized in-

57. Ruddick, supra note 56, at 462.
58. Id.
59. Id.
60. See generally THE RESPONSIBILITY TO PROTECT, supra note 34.

[T]o wrestle with the whole range of questions—legal, moral, operational and political—rolled up in this debate, to consult with the widest possible range of opinion around the world, and to bring back a report [to] help the Secretary-General and everyone else find some new common ground.

63. The Corfu Channel Case, (U.K. v. Alb.), 1949 I.C.J. 4 (April 9). In this case, two British cruisers and two destroyers were traveling in the North Corfu Strait, which was considered Albanian waters. One of the destroyers, the Saumarez, struck a mine. The other destroyer, the Volage, came to the assistance of the Saumarez and, while towing it, struck another mine. This accident resulted in damage to both destroyers, the deaths of forty-five British officers and sailors, and injury to forty-two other people. The I.C.J. held Albania liable for its failure to warn of the mines, about which it should have known, located in its territory. Id. at 12–13, 118.
ternational obligations towards humanity that exist during peacetime and during war.\textsuperscript{64} Subsequently, the U.N. acknowledged the importance of State responsibility\textsuperscript{65} from the International Law Commission’s (ILC)\textsuperscript{66} Draft Articles on the responsibility of States for internationally wrongful acts (ILC’s Draft Articles),\textsuperscript{67} which brought State responsibility to the international community’s attention.\textsuperscript{68} The ILC’s Draft Articles state, in pertinent part, “[e]very internationally wrongful act [or omission] of a State entails the international responsibility of that State.”\textsuperscript{69} Under this rationale, since respect for human rights is a State obligation under customary international law,\textsuperscript{70} it is a logical inference that States have an international responsibility not to allow violations of human rights to occur in their territories. Otherwise, they have committed an internationally wrongful act. Therefore, international human rights law has challenged the traditional understanding of State sovereignty by reinforcing a concept of sovereignty that imposes on States the responsibility to protect human rights.\textsuperscript{71}


\textsuperscript{66} The ILC is composed of thirty-four members who are elected by the General Assembly to “encourag[e] the progressive development of international law and its codification.” International Law Commission: Introduction, \textit{http://www.un.org/law/ilc/introfra.htm} (last visited Mar. 27, 2007); U.N. Charter art. 13(1).

\textsuperscript{67} See G.A. Res. 56/83, Annex, \textit{supra} note 65.

\textsuperscript{68} G.A. Res. 56/83, \textit{supra} note 65, at 2, ¶ 3.

\textsuperscript{69} G.A. Res. 56/83, Annex, \textit{supra} note 65, at 2, art. 1.

\textsuperscript{70} See \textit{supra} note 40 and accompanying text.

The ICISS report further elaborated on the notion of State responsibility by presenting the concept of the responsibility to protect:

[T]he idea that sovereign [S]tates have a responsibility to protect their own citizens from avoidable catastrophe—from mass murder and rape, from starvation—but that when they are unwilling or unable to do so, that responsibility must be borne by the broader community of [S]tates.72

The report redefines the view of humanitarian intervention from the right to intervene into the responsibility to protect,73 thus refocusing humanitarian intervention on “the rights of affected populations and the obligations of outsiders to help.”74 The ICISS report elaborates two basic principles of the responsibility to protect:

[First,] State sovereignty implies responsibility, and the primary responsibility for the protection of its people lies with the [S]tate itself. [Second,] where a population is suffering serious harm, as a result of internal war, insurgency, repression or [S]tate failure, and the [S]tate in

A State cannot pretend absolute sovereignty without demonstrating a duty to protect people’s rights . . . . When Governments fail to protect human rights . . . the international community can intervene and exercise an extraterritorial duty to protect people at risk. The principle of sovereignty is not denied by such intervention; it refocuses from being an absolute control over certain territory to being a responsibility to govern in a certain manner. The sovereignty of States . . . depends on their duty when governing to respect human beings. The sovereignty of States means the sovereignty of people, not of leaders.

Id.; Katja Luopajärvi, Is There an Obligation on States to Accept International Humanitarian Assistance to Internally Displaced Persons Under International Law?, 15 INT’L J. REFUGEE L. 678, 684 (2003) (highlighting Mr. Francis Deng’s, the Representative of the U.N. Secretary-General on Internally Displaced Persons, support for the State responsibility concept. Deng argues that responsible sovereign powers ensure respect for and protection of fundamental human rights, and under exceptional circumstances when governments cannot fulfill that responsibility, the international community has an obligation to step in and provide a remedy for those who are suffering.).

72. THE RESPONSIBILITY TO PROTECT, supra note 34, at viii.

73. Gareth Evans, The Responsibility to Protect: Rethinking Humanitarian Intervention, 98 AM. SOC’Y INT’L. L. PROC. 78, 81 (2004) (explaining that the responsibility to protect is an obligation owed by all sovereign States to their citizens, which must be assumed by the international community of States if sovereign States fail to live up to that responsibility); Levitt, supra note 61, at 155 (noting that the ICISS report focuses on the obligations of States, as opposed to the rights of States).

74. Levitt, supra note 61, at 155 (quoting Thomas G. Weiss, To Intervene or Not to Intervene? A Contemporary Snap-Shot, CANADIAN FOREIGN POL’Y 141, 146 (2002)).
question is unwilling or unable to halt or avert it, the principle of non-intervention yields to the international responsibility to protect.\textsuperscript{75}

The ICISS report explains the responsibility to protect as embracing three responsibilities:

1. **The responsibility to prevent**: to address both the root causes and direct causes of internal conflict and other man-made crises putting populations at risk.

2. **The responsibility to react**: to respond to situations of compelling human need with appropriate measures, which may include coercive measures like sanction and international prosecution, and in extreme cases military intervention.

3. **The responsibility to rebuild**: to provide, particularly after a military intervention, full assistance with recovery, reconstruction and reconciliation, addressing the causes of the harm the intervention was designed to halt or avert.\textsuperscript{76}

The responsibility to protect encompasses three State responsibilities: the responsibility to prevent, to react, and to rebuild, which the U.N. Secretary-General Kofi Annan urged Member States to embrace “as a basis for collective action against genocide, ethnic cleansing and crimes against humanity.”\textsuperscript{77} While the ICISS report discusses the responsibility

---

75. **The Responsibility to Protect**, supra note 34, at xi. The ICISS report also reads:

The foundations of the responsibility to protect, as a guiding principle for the international community of States, lie in: obligations inherent in the concept of sovereignty; the responsibility of the Security Council, under Article 24 of the UN Charter, for the maintenance of international peace and security; specific legal obligations under human rights and human protection declarations, covenants and treaties, international humanitarian law and national law; the developing practice of States, regional organizations and the Security Council itself.

*Id.*

76. **The Responsibility to Protect**, supra note 34, at xi.


I believe that we must embrace the responsibility to protect, and, when necessary, we must act on it. This responsibility lies, first and foremost, with each individual State, whose primary raison d’être and duty is to protect its population. But if national authorities are unable or unwilling to protect their citizens, then the responsibility shifts to the international community to use diplomatic,
to protect under the doctrine of humanitarian intervention—military intervention during situations of gross human rights violations—it would be logical to apply this concept to humanitarian assistance—relief efforts provided for disaster-stricken countries. Applying the responsibility to protect to natural disasters is a novel idea, as much of the legal scholarship regarding State responsibility discusses State responsibility to intervene under the doctrine of humanitarian intervention and State responsibility for international man-made disasters. 

---

humanitarian and other methods to help protect the human rights and well-being of civilian populations.

*Id.* at 35, ¶ 135.

78. See Scheffer, *supra* note 31, at 264. Scheffer writes:

The classical definition of ‘humanitarian intervention’ is limited to those instances in which a nation unilaterally uses military force to intervene in the territory of another [S]tate for the purpose of protecting a sizable group of indigenous people from life-threatening or otherwise unconscionable infractions of their human rights that the national government inflicts or in which it acquiesces.

*Id.* The concept of State responsibility has recently been addressed in the context of internally displaced persons, which are:

[P]ersons or groups of persons who have been forced or obliged to flee or to leave their homes or places of habitual residence, in particular as a result of or in order to avoid the effects of armed conflict, situations of generalized violence, violations of human rights or natural or human-made disasters, and who have not crossed an internationally recognized State border.


79. See generally McClatchey, *supra* note 64 (noting that the international community addressed the concept of State responsibility in the aftermath of two well-known man-made environmental disasters: the Chernobyl explosion in the former Soviet Union and the Sandoz spill in Switzerland); Am. Soc’y of Int’l L. [ASIL], *International Responsibility for Manmade Disasters*, 81 AM. SOC’Y INT’L L. PROC. 320 (1990) (providing a summary of a discussion among legal scholars regarding three aspects of government
This Note proposes that all States, disaster-affected States and the international community of States, have a responsibility to humanity when natural disasters occur. Applying the responsibility to protect to natural disasters would highlight the following responsibilities befalling on all States: (1) to warn people and nations potentially affected by an impeding or occurring disaster; (2) to unconditionally provide essential disaster relief and accept it if needed; and (3) to ensure sustainable reconstruction and rehabilitation of disaster-affected areas. Further clarification of how these responsibilities play out in the context of past natural disasters demonstrates the plausibility of applying the responsibility to protect to the provision of humanitarian assistance in natural disaster situations.

III. WHEN DISASTER STRIKES

A. Empirical Examples: The Responsibility to Protect in Natural Disasters

State sovereignty is fundamental to international disaster relief operations. When natural disasters strike, disaster-affected States have the authority to manage all aspects of the disaster’s aftermath by either requesting or refusing aid from the international community of States. Disaster-affected States may reject assistance for a variety of political reasons, such as, “the embarrassment and dependency implicit in asking the outside (usually the West) for help,” which consequently exposes the disaster-affected States’ incompetence in effectively responding to responsibility for transnational man-made disasters: (1) State responsibility to prevent disasters; (2) State responsibility to mitigate damages after a disaster has occurred or is imminent; and (3) State responsibility for the payment of damages); Sudhir K. Chopra, Multinational Corporations in the Aftermath of Bhopal: The Need for a New Comprehensive Global Regime for Transnational Corporate Activity, 29 Val. U. L. Rev. 235 (1994) (explaining the factual situations of the industrial disasters that occurred in Bhopal, India; Seveso and Meda, Italy; the former Soviet Union; and Basel, Switzerland. Chopra assesses the adequacy of the legal responses to these disasters and recognizes the need for an international legal regime to regulate multinational corporate activity to prevent future industrial disasters.).

80. MACALISTER-SMITH, supra note 2, at 55 (1985) (stating that respect for the sovereignty of States and protection of their freedom of action are essential tenets applicable to relief actions).

81. Id. at 56 (noting that under international law, there is neither a legal obligation on States to offer disaster relief to other countries nor an obligation on States to respond to requests for help).

natural disasters. Political motivations significantly influence how disaster-affected States choose to respond to disaster relief offers. For example, political factors may delay disaster-affected States’ acknowledgement that a natural disaster has occurred, as well as the full extent of its effects, thereby hindering the provision of relief to disaster victims. Herein lies the problem. Ensuring the adequate protection of natu-

83. Macalister-Smith, supra note 2, at 72 (explaining that similar to other emergency situations, natural disasters draw attention to a disaster-affected State’s competence and authority in responding to the adverse effects of such disasters within its territory).

84. Freudenheim, supra note 82, at 228. Freudenheim identifies four major political problems: (1) domestic politics, wherein “domestic politics in the disaster-stricken country obstruct formal acknowledgement of the disaster, and/or obstruct disaster preparedness, relief and rehabilitative measures;” (2) domestic corruption, wherein “corruption in the disaster-stricken country leads to inflated assessments of the disaster damage, and misallocation of relief goods;” (3) rejection of aid, wherein “the government of the disaster-stricken country rejects offers of governmental foreign aid for political reasons;” and (4) international politics, wherein “international politics obstruct disaster preparedness, relief or rehabilitation.” Id.; see also Ullman, supra note 27, at 5 (listing political reasons for why those in power may impede disaster relief: to prolong incumbency; to cover up government incompetence; to conceal disagreeable domestic conditions that could hinder tourism, trade, and foreign financial transactions; and to undermine domestic opposition).

85. Macalister-Smith, supra note 2, at 73; Ullman, supra note 27, at 1 (explaining that the governments of disaster-stricken States may be unwilling to acknowledge, to their own people and the international community of States, the magnitude of a disaster’s effects on their territory for both domestic and international political reasons. Governments exhibiting this conduct intensify the devastation and suffering of natural disaster victims.).

86. See Freudenheim, supra note 82, at 241 (noting that “[u]nfortunately, without the government’s acknowledgement that it needs help, the foreign aid mechanisms of other governments and international agencies (with the important exception of the voluntary agencies) are immobilized”); see also Green, supra note 20, at 12 (noting that disaster-affected governments may make politically motivated decisions that are harmful to disaster victims. In effect, these decisions have “denied the existence of a disaster[,] . . . steered relief to one ethnic group rather than another[,] . . . given aid on the basis of perceived diplomatic benefits rather than according to the needs of the victims, [and allowed governments to profit] enormously from the sale of relief goods.”); Zama Coursen-Neff, Preventative Measures Pertaining to Unconventional Threats to the Peace Such as Natural and Humanitarian Disasters, 30 N.Y.U. J. Int’l L. & Pol. 645, 650–51 n.19 (1998) (noting that mass human suffering is not always less political in natural disasters than in conflict situations. “The cholera epidemic in Guinea (1969), the tidal wave and floods in East Pakistan (1970), the early and middle stages of the drought in the Sahel (1968-1972), the Ethiopian famine and cholera epidemic (1973), the famines in India and Haiti (1975), and the Philippine earthquake (1976) are but a few of the disasters in which politically prompted official resistance prevented or delayed relief operations.”).
ral disaster victims may be in peril because international law does not impose any legal obligations on States regarding disaster prevention, relief, and recovery. Therefore, when disaster-affected States alone cannot adequately protect their disaster-stricken populations—when they are unable or unwilling to provide relief—then intervention from the international community of States in the form of humanitarian assistance becomes necessary to protect the livelihoods of natural disaster victims.

Although there are no established, universal guidelines regarding States’ actions during times of natural disasters, past disasters have highlighted certain responsibilities of disaster-affected States and the international community of States before, during, and after natural disasters occur. Applying the responsibility to protect to natural disaster situations illuminates these responsibilities, which include: (1) the responsibility to prevent, which is the responsibility of disaster-affected States and the international community of States to warn people and other regions that may be affected by an impending natural disaster; (2) the responsibility to react, which is the responsibility of disaster-affected States to accept needed humanitarian relief from the international community of States, particularly when they are unable or unwilling to effectively provide relief within their territory, and the concurrent responsibility of the international community of States to provide humanitarian assistance to disaster-affected States by offering relief goods, funding, ser-

87. Id. at 13–14 (noting that “the international disaster relief system does not always serve well the interests of disaster-stricken countries and their people”).
88. Macalister-Smith, supra note 2, at 56.
89. Id. at 52 (explaining that when States alone cannot ensure effective and adequate protection of disaster victims within their territory, international actions are often taken to ensure the sufficient protection of disaster-stricken populations).
90. Samuels, supra note 23, at 263–64 (noting that the legal system for ensuring the practice of disaster relief assistance between States is weak. However, the progress made by the U.N. Disaster Relief Organization (UNDRO) and the International Red Cross Movement regarding the codification of disaster relief practices provides some optimism for developing international disaster relief laws.); see generally International Disaster Response Laws, supra note 78 (The IFRCRCs initiated the International Disaster Response Law Project (IDRL Project) to analyze the various international laws, principles, and practices relating to international disaster response. The IFRCRCs recognized that unlike the detailed and well-recognized provisions for protection and assistance under international humanitarian law during times of armed conflict, there is a lack of an identifiable systematic source of law for humanitarian assistance during times of natural disasters. The IDRL Project is a valuable step towards improving the effectiveness of disaster relief operations.); Macalister-Smith, supra note 2, at 165 (noting the important role of the UNDRO in strengthening and improving the coordination of humanitarian assistance).
91. See discussion infra Part III.A.1–3.
The Responsibility to Protect

1. The Responsibility to Prevent

The first prong of the ICISS’ three-fold responsibility is the responsibility to prevent: the “[p]revention of deadly conflict and other forms of man-made catastrophe.” The context of natural disasters, since prevention requires warning, it would be logical to interpret the responsibility to prevent as encompassing the responsibility to warn. The responsibility to warn would include taking preventative measures to mitigate the consequences of natural disasters. The duty to warn of an impending disaster is not a new concept, particularly in the context of man-made disasters, such as the Chernobyl explosion and the Sandoz spill in 1986. Neither of the offending States, the former Soviet Union nor Switzerland, faced legal consequences for their failure to notify adversely affected neighboring States. However, these incidents have raised awareness among legal scholars concerning the responsibility of disaster-
affected States to notify other States of impending disasters causing transnational harm.  

Additionally, prior international case law has addressed the duty to warn. Some scholars have broadly interpreted the I.C.J.’s ruling in the Corfu Channel Case as providing that all States have a general duty to warn other States of potential or impending harm.  

Other scholars have limitedly construed the I.C.J.’s holding in contending that the duty “not to allow knowingly [their] territory to be used for acts contrary to the rights of other [S]tates” imposes a duty to warn only on disaster-affected States, where the danger is located within their territory.  

Regardless of these variant views’ interpretation of the I.C.J.’s opinion, the Corfu Channel Case serves as a building block for the emergence of the duty to warn concept.  

Due to prior international man-made disasters and case law addressing these events, the responsibility of all States to prevent, or mitigate, harms by warning of impending disasters has become well-recognized. The Indian Ocean tsunami demonstrates how this responsibility to warn materializes in a natural disaster.

i. Indian Ocean Tsunami

On December 26, 2004, an earthquake measuring 9.0 on the Richter scale triggered a massive and powerful tsunami across the Indian Ocean.  

The Indian Ocean tsunami disaster created one of the most complex coordination and logistical challenges the disaster response system of the United Nations has ever had to manage. The timing and scale of the event required quick and flexible coordination efforts in a variety of countries and contexts, and its global reach led to a proliferation of relief actions and actors and garnered assistance from public, private and government sources at the highest levels.
As a repercussion of the tsunami disaster, European survivors and relatives of the victims brought a class action suit against the Pacific Tsunami Warning Center, the U.S. National Oceanic and Atmospheric Administration (NOAA), the Accor Group (Accor), a French hotel chain with beach resorts in Thailand, and the Kingdom of Thailand for their failure to establish or properly use warning systems for natural disasters. The Complaint alleges, inter alia, that Thailand had a duty to notify countries of the tsunami because it possessed adequate information of the deadly waves and the waves’ direction prior to the disaster’s occurrence. Specifically, the Petitioners argue that Thailand should have alerted the affected countries through media or other telecommunications means and evacuated Thailand’s beach areas.

This alleged duty to warn applies to natural disasters. There is a considerable amount of scholarship suggesting that governments, businesses, and individuals have a duty to prepare for hazardous events, many of which are highly unpredictable, such as tsunamis, meteor strikes, and earthquakes. In the aftermath of the Indian Ocean tsunami, the interna-

While such global attention and support has been both welcome and constructive, the high-intensity environment it has created has raised expectations of performance and accountability.

Id. at 6, ¶ 17.

102. Id. at 2, ¶ 2.

103. Id. at 2, ¶ 3.


105. The Complaint, ¶ 19–21. The Complaint also contends that although the NOAA registered the Indian Ocean earthquake, it failed to timely alert the affected countries of the impending tsunami. Id. ¶ 17. Further, the Complaint alleges that Accor knew or should have known that its resort in Khao Lak, Thailand was located in an area that was susceptible to earthquakes and tsunamis, and had a duty to notify its guests of impending dangers and take the proper and necessary precautions, such as evacuating the beach and moving people to higher ground. Id. ¶ 23. For the purposes of this Note, an analysis of the Complaint solely focuses on the alleged duty of the Kingdom of Thailand—to warn other countries of the approaching tsunami. This Note does not assess the sufficiency of the Plaintiffs’ other arguments in the Complaint.

106. The Complaint lists the following suggested notification devices: radio, television, electronic notices, emergency broadcasts, and air raids. Id. ¶ 21.

107. See Evan R. Seamone, The Duty to “Expect the Unexpected”: Mitigating Extreme Natural Threats to the Global Commons Such as Asteroid and Comet Impacts with the Earth, 41 COLUM. J. TRANSNAT’L L. 735, 758 (2003) (justifying the duty to warn of natu-
tional community recognized the lack of disaster preparedness from not having a proper warning system in place, which resulted in the failure to mitigate harm to the people in disaster-stricken regions. In response to the tsunami disaster, the U.S. Agency for International Development (USAID) launched the Indian Ocean Tsunami Warning System (IOTWS) program, an effort to develop “integrated early warning and mitigation systems” to assist the Indian Ocean region in detecting and preparing for natural disasters, such as tsunamis and other coastal hazards. The IOTWS program involves the participation of several U.S. agencies, the Intergovernmental Oceanographic Commission of the U.N. Educational, Scientific, and Cultural Organization, numerous donor nations, and national governments located in the Indian Ocean region. The goal of the IOTWS is “to build up the human and institutional infrastructure to make sure these [warning] systems are interoperable and sustainable for years to come.”

---

[Tsunamis] may occur in a particular geographic area at an indefinite upcoming time. Some coastal locations might not be affected by tsunamis for hundreds or thousands of years. Other localities, like the Indian Ocean, are the tsunami equivalent to San Francisco for earthquakes or Florida for hurricanes. However, since all of these acts of nature are reasonably foreseeable, . . . that fact establishes a duty to use reasonable care to decrease the risks of a catastrophe. Id.


110. While donor nations or donor governments repeatedly provide the majority of disaster relief assistance, their relief aid is often directed through the U.N. or private agencies. GREEN, supra note 20, at 29.

111. USAID Press Release, supra note 108.

112. Id. (explaining that a top priority in the post-tsunami reconstruction effort was to work with the participating entities to establish a fully functional warning system in the Indian Ocean).
Considering the international community’s response to the tragic tsunami disaster—developing a tsunami warning system in the Indian Ocean—and the recognition of the duty to warn in prior international case law and past international man-made disasters, there is considerable evidence that the international community has acknowledged the duty to warn concept. Thus, in the context of natural disasters, it is conceivable that disaster-affected States and the international community of States have a duty to warn of impending or occurring natural disasters, which encompasses the responsibility to prevent, or mitigate, harm afflicted to disaster-affected people and countries.\(^{113}\)

2. The Responsibility to React

The second prong of the ICISS’ three-pronged responsibility is the responsibility to react, which denotes that “when a [S]tate is unable or unwilling to redress the [adverse] situation, then interventionary measures by other members of the broader community of [S]tates may be required.”\(^{114}\) The duty to intervene has principally been discussed in the context of humanitarian intervention, during which States take armed intervening measures in another State, without its consent, to redress the treatment of its population who are being subject to gross violations of human rights.\(^{115}\) While there is currently debate over the controversial...

\(^{113}\) Formulating an argument for imposing on all States a duty to warn of natural disasters logically begs the question of whether States should be held liable for their failure to warn. While this Note does not delve into the aspects of States’ liability for failure to adhere to their proposed responsibilities, it is arguably not in the best interest of the international community to hold States strictly liable because of the value of international diplomacy over adverse political implications arising from a finding of liability. \textit{Cf.} McClatchey, \textit{supra} note 64, at 676–78 (explaining that international law prefers to use a negligence standard for State responsibility for transboundary harm, rather than a strict liability standard. Under a traditional tort analysis of negligence for transboundary harm, the complaining State must prove that: (1) the offending conduct is attributable to the defendant State; (2) the offending State breached an international duty; (3) a causal connection exists between the conduct and the injury; and (4) material damages took place.). \textit{Id.} This Note does not inquire into the intricacies of States’ liability for failure to adhere to their responsibilities regarding natural disasters, but posits that implementing diplomatic measures and political incentives, which focus on shared interests and foster international cooperation among States, would more effectively influence the conduct of States in adhering to their responsibilities. \textit{Id.} at 680.

\(^{114}\) \textit{THE RESPONSIBILITY TO PROTECT, supra} note 34, ¶ 4.1.

\(^{115}\) A.P.V. Rogers, \textit{Humanitarian Intervention and International Law}, 27 HARV. J.L. & PUB. POL’Y 725, 730 (2004) (noting that humanitarian intervention is generally thought of as the “use of force without the authorization of the Security Council to protect sections of a State’s population from gross and persistent human rights abuses”); Ravi Mahalingam, Comment, \textit{The Compatibility of the Principle of Nonintervention with the...
doctrine of humanitarian intervention,\textsuperscript{116} this Note does not delve into the complexities of the humanitarian intervention dispute. Rather, this Note posits that the responsibility to react can conceivably apply to catastrophic natural disasters, during which there also exists grave human suffering. Such responsibility would be to unconditionally provide and accept needed humanitarian assistance when disaster-affected States are unwilling or unable to provide disaster relief.

International case law has addressed the responsibility to provide and accept humanitarian assistance. The I.C.J. in \textit{Nicaragua v. U.S.}\textsuperscript{117} articulated this responsibility:

There can be no doubt that the provision of strictly humanitarian aid [such as food, clothing, and medicine] to persons or forces in another country, whatever their political affiliations or objectives, cannot be regarded as unlawful intervention, or as in any other way contrary to international law.\textsuperscript{118}

In its opinion, the I.C.J. justified neutral offers of humanitarian aid, which is offered in accordance with the purposes declared by the Red
Cross: “to prevent and alleviate human suffering,” “to protect life and health,” “to ensure respect for the human being,” and to “be given without discrimination to all in need.” Based on these stated objectives, it is reasonable to conclude that international law supports humanitarian assistance for natural disaster victims, provided that the relief offered is for purely humane purposes.

Several past natural disasters, in particular North Korea’s famine and Hurricane Katrina, have highlighted the responsibility of disaster-affected States to accept needed humanitarian assistance and the concurrent responsibility of the international community of States to unconditionally offer humanitarian aid.

i. North Korea’s Famine

In the 1990s under the Pyongyang regime, a famine in North Korea killed approximately one-million people, about five percent of its population. This famine continues today as North Korea is now facing its second decade of food shortages. There exists disagreement over the causes of this famine. North Korea claims it arose primarily from natural disasters, such as the floods of 1995, which devastated its rice-growing regions. Conversely, other views regard the floods as a significant exacerbating factor that diverted blame away from the North Korean government, and ultimately speculate that the causes of the famine are political. Regardless of the cause, the effects of the famine are
devastating as close to two-million people have died and children continue to suffer from chronic malnutrition. These figures show that North Korea was not only slow to recognize the extent of its famine problem, but also slow to provide adequate relief for its famine-stricken population.

North Korea’s inadequate response became evident at the height of the famine in the mid-1990s when it allowed humanitarian assistance from the international community of States, but kept tight restrictions over the entry of relief goods into many areas of the country. North Korea also refused “to allow swift and equitable distribution” of the aid to its people. Furthermore, North Korea’s recent request to end all humanitarian relief has received much criticism. Deputy Foreign Minister Choe Su-hon asserted that because of North Korea’s good harvest, North Korea has enough food and no longer needs food aid. However, putting a stop to relief efforts such as the U.N. World Food Program (“WFP”),

because of its isolation, the particular cause of its famine are unknown.). This Note neither probes into an analysis of opposing views regarding North Korea’s famine crisis, nor seeks to determine which view was correct, be it floods or political agenda. Rather, this Note focuses on North Korea’s inadequate response to its famine crisis. Terry, supra note 125, at 90.


128. Haggard & Noland, supra note 122, at 20 (noting that the Public Distribution System—the rationing system implemented to ensure food security—was unable to provide even the minimal amount of food needed for human survival during the mid-1990s).


132. N. Korea Says Asked UN to End All Humanitarian Aid, supra note 131 (quoting Deputy Foreign Minister Choe Su-Hon, “[H]umanitarian assistance cannot last too long . . . [W]e have very good farming this year. Our government is prepared to provide the food to all our people.

133. The U.N. established the WFP in 1961. Its purpose is to provide food on an emergency basis to disaster-stricken and famine-affected regions. The WFP also provides technical support to developing countries to help improve their food production, storage, and distribution systems. GORMAN, supra note 10, at 83–84.
an international food aid agency that has nourished approximately 6.5 million North Koreans, raises concerns about North Korea’s ability to provide adequate relief to its famine-stricken population.  

Governments of disaster-affected States have rejected relief offers in the past. During the 1976 Guatemalan earthquake and the 1976 Philippine earthquake and tidal wave, humanitarian aid to both disaster-affected States appeared to “have strings attached.” Likewise, Mr. Choe Su-hon asserted that food aid to North Korea also had strings attached and argued that the international community of States, particularly the United States, “politicize[d] humanitarian assistance, [by] linking it to the human rights issue.” The U.S. State Department has adamantly rejected North Korea’s accusations, arguing that the United States’ decisions to provide humanitarian assistance were not based on political agenda, but rather on need: the need of the country involved, competing needs elsewhere, and ensuring that the provision of aid goes to the neediest. This need-based rationale provides a sensible approach for all

134. See sources cited infra note 140.
135. Freudenheim, supra note 82, at 242 (explaining that in the aftermath of the 1976 Guatemalan earthquake, the Guatemalan government rejected a British offer of aid, not because the aid was unnecessary, but rather because of a “political dispute between the two countries over the neighboring country of Belize.” Freudenheim notes that Guatemala’s acceptance of the British aid “would have been politically embarrassing.” Additionally, in the aftermath of the Philippine earthquake and tidal wave in 1976, President Ferdinand Marcos refused American aid because the Philippine government viewed the aid as “aid for concessions in their treaty negotiations.” Freudenheim notes that while the Philippine government declared that no foreign aid was necessary in its relief operation, it consequently stated that UNICEF aid and aid from other U.N. agencies did not classify as foreign aid.). Id.
136. Id. Freudenheim further notes the concern that political considerations of national governments, donor governments, and international agencies may influence the provision of disaster relief, stating, “[t]he effects of this political reality, which can be as disastrous to the victim as any earthquake, drought, or epidemic, should be of urgent concern to all those interested in improving the international and national response to natural disaster situations.” Id. at 244.
137. See N. Korea Says Asked UN to End All Humanitarian Aid, supra note 131.
138. Id. In the World Disasters Report of 2000, the IFRCRCS quoted Margareta Wahlström, the Under-Secretary General for Disaster Response and Operational Coordination, on the effects of continuing aid to North Korea in the wake of political upheaval. Wahlström states:

The [aid] system might be utilized but . . . it is for a good purpose because you cannot create stability in this part of the world without creating a bridge. The humanitarian agencies, be it the UN, the Red Cross or NGOs . . . have made an incredible contribution to creating that bridge unconditionally. The conditions we have imposed . . . belong to the humanitarian agenda . . . . [W]e have not said that in order to give food we need something else from you.
States to follow regarding the provision and acceptance of international disaster relief aid. Therefore, North Korea’s justification for rejecting humanitarian assistance should depend on whether its harvest meets the needs of its people.

North Korea’s harvest has eased the food shortage. However, the international community still believes that North Korea needs humanitarian assistance, particularly to provide food for those most in need: children, pregnant women, urban poor, and elderly. Current conditions in North Korea also show that its obstruction of needed humanitarian aid is hurting, rather than benefiting its population. Therefore, because of North Korea’s unwillingness to protect the well-being of its people, the international community of States has a responsibility to provide unconditional humanitarian assistance to protect the people of North Korea. Accordingly, North Korea should have a responsibility to accept the needed humanitarian relief.

ii. Hurricane Katrina

From August 25 to August 31, 2005, Hurricane Katrina, varying in strength from a category one to category five hurricane, swept across the southern United States striking Florida, Louisiana, Mississippi, and Alabama. Homeland Security Secretary Michael Chertoff described
the hurricane as an “ultra-catastrophe,” ravaging through four states, triggering a flood due to failed canal levees that swamped New Or-
leans, the nation’s thirty-fifth largest city, and resulting in a break-
down of civil order. As of September 15, 2005, Hurricane Katrina had caused “$125 billion dollars in damage/costs, making [it] the most ex-

pensive natural disaster in U.S. history.” Additionally, Hurricane Katrina devastated the lives of thousands of people. Ordinary citizens and other countries criticized the United States’ response to this hurri-
cane as a disaster in itself because of the lack of preparedness and the many days it took for relief aid and personnel to reach New Orleans and other hurricane-devastated areas. President George W. Bush acknowledged the United States’ inadequate response to the hurricane saying, “Katrina exposed serious problems in our response capability at all levels of government and to the extent the federal government didn’t fully do its job right, I take full responsibility.”

A congressional investigation revealed that the U.S. government at all levels failed to adequately prepare for and respond to this tragedy. For


144. It is also possible to view the breached canal levees, which flooded about eighty percent of New Orleans, as a failure of the responsibility to prevent, or mitigate, harm resulting from Hurricane Katrina. Investigation is pending regarding whether the failure of the levees was due to poor engineering, faulty construction, or lack of proper mainte-


147. Id.

148. Juan Forero & Steven R. Weisman, U.S. Allies, and Others, Send Offers of Assist-
ance, N.Y. TIMES, Sept. 4, 2005 (noting that the circulation of televised and printed im-
ages of destruction, suffering, and death from Hurricane Katrina provoked disbelief and an outpouring of sympathy and assistance from the international community); World Shocked by U.S. Response to Storm, GAINESVILLE.COM, Sept. 4, 2005, http://www.gainesville.com/apps/pbcs.dll/article?AID=/20040904/WIRE/50904039/1117/news (noting the world’s shock and criticism over such a wealthy and powerful country’s response to the storm).


150. Id.

example, in the aftermath of the hurricane’s destruction, it took several days to get help to the thousands of people left behind without food and water. Those who sought safety in shelters such as the Superdome stadium were subject to grossly degenerate conditions.\textsuperscript{152} Moreover, as aerial shots of New Orleans showed, hundreds of school and city buses were flooded with water—buses that could have been used in the evacuation process.\textsuperscript{153} The rest of the United States and the world watched these horrific images on television, questioning why the United States was unable to respond effectively to the disastrous effects, as well as criticizing the United States for its lack of any competently executed evacuation plan, organized relief effort, or central command to coordinate relief.\textsuperscript{154}

Despite criticism regarding the United States’ inability to effectively provide relief to its disaster-stricken regions, other countries expressed their sympathy and generosity towards disaster victims, as offers of over $1 billion in cash and supplies poured in from nearly ninety-five nations,\textsuperscript{155} which included several countries that are adversaries or typically aid recipients, not aid donors.\textsuperscript{156} The United States graciously accepted

\footnotesize{\textbf{HURRICANE KATRINA,}} at 1–5 (2005), \textit{available at} \url{http://katrina.house.gov/full_katrina_report.htm}.

\textsuperscript{152} Evan Thomas, \textit{The Lost City: What Went Wrong: Devastating a Swath of the South, Katrina Plunged New Orleans into Agony. The Story of a Story—and a Disastrously Slow Rescue}, \textit{NEWSWEEK}, Sept. 12, 2005, \textit{available at} \url{http://www.msnbc.msn.com/id/9179587/site/newsweek/?g=1} (describing the nightmarish conditions at the Superdome: the air conditioning failed; the lights went out, leaving the stadium only dimly lit; the hurricane broke holes through the roof; the availability of bottled water was scarce; the stench of unwashed bodies and overflowed toilets; the occasional sounds of gunshots; and the occurrences of rapes).


\textsuperscript{154} Thomas, \textit{supra} note 152.

\textsuperscript{155} \textit{U.S. Receives Aid Offers from Around the World}, \textit{CNN.COM}, Sept. 4, 2005, \textit{http://edition.cnn.com/2005/US/09/04/katrina.world.aid}; \textit{Offers of Aid from Around the World}, \textit{CNN.COM}, Sept. 5, 2005, \textit{http://www.cnn.com/2005/US/09/05/katrina.world.aid} (highlighting examples of documented foreign aid, which include: Nigeria’s pledge of $1 million; Japan’s offer of $200,000 to the American Red Cross and $300,000 in emergency supplies; India’s $5 million donation to the American Red Cross and willingness to donate medicine; Afghanistan’s offer of $100,000; Sri Lanka’s donation of $25,000 to the American Red Cross; Mexico’s offer of $1 million and delivery of fifteen truckloads of water, food, and medical supplies; Israel’s offer to provide medical assistance in the form of personnel, equipment, and medicines; and Qatar’s offer of $100 million).

much of the disaster relief aid, provided that it was matched with need and offered in the spirit of humanitarian assistance, and was thankful for the generosity expressed by the international community. Although the United States accepted most of the offered relief, there were complaints from the international community of States concerning the “bureaucratic entanglements” that hindered the shipment of relief supplies to the United States. However, it is reasonable to perceive that such delay was due to temporary technical and logistical problems, as well as the novelty of coordinating the acceptance of such a considerable amount of international aid to the United States. Similar to North Korea’s response to its famine, international criticism also arose from the United States’ rejection of aid from certain countries. This rejection has been

157. Slavin, supra note 156 (quoting U.S. State Department official Harry Thomas, who was in charge of coordinating foreign aid offers, “[t]he worst thing we could do is to take things, have them . . . sit on the ground and not be utilized, to have something rot or not get to people quickly”).

158. A Foreign Aid Twist: U.S. Gets, Others Give, supra note 156 (noting Secretary of State Condoleezza Rice’s efforts in profusely thanking other countries for the humanitarian assistance offered to the United States); César G. Soriano, Foreign Aid Flows in for Katrina Victims, USATODAY.COM, Sept. 5, 2005, http://www.usatoday.com/news/world/2005-09-05-aid-katrina_x.htm (quoting Secretary of State Condoleezza Rice’s deep appreciation for the foreign offers, “[p]eople have said that America has been so generous . . . in other places, and now it is time to be generous to America”); Forero & Weisman, supra note 148 (quoting Secretary of State Condoleezza Rice, “[r]ecently we have seen the American people respond generously to help others around the globe during their times of distress, such as during the recent tsunami . . . [t]oday we are seeing a similar urgent, warm and compassionate reaction”).


160. Id.

161. Forero & Weisman, supra note 148 (noting that the United States seemed initially unprepared for the outpouring of aid, which surpassed the humanitarian aid offered after the 9/11 attacks); Slavin, supra note 156 (quoting Natalie Loiseau, press counselor at the French Embassy in Washington, D.C., “this is the first time the United States has [had] to welcome foreign aid, so no one has had this job (of facilitating foreign aid)”).

162. Mary Murray, Katrina Aid from Cuba? No Thanks, Says U.S., MSNBC.COM, Sept. 14, 2005, http://msnbc.msn.com/id/9311876 (explaining that the United States asserted it would not need Cuba’s medical brigade because of a “robust response from the American medical community”); Wolfson, supra note 13 (noting that the United States rejected Iran’s offer of ten million barrels of crude oil because it was conditioned on lifting economic sanctions).
justified because certain offers were either not needed or conditional, not solely for the protection of disaster victims.\textsuperscript{163}

Nevertheless, the outpouring of hurricane relief to the United States further supports the responsibility of the international community of States and disaster-affected States regarding humanitarian assistance. Hurricane Katrina demonstrates that when disaster-affected States, regardless of their wealth and power, are unable to effectively provide adequate relief to its disaster-stricken populations, the international community of States has a responsibility to unconditionally offer aid and disaster-affected States have a responsibility to accept it, provided that the aid is needed and in the spirit of humanitarian assistance.

3. The Responsibility to Rebuild

The third prong of the ICISS’ three-pronged responsibility is the responsibility to rebuild, which implies that:

\begin{quote}
[B]ecause of a breakdown or abdication of a [S]tate’s own capacity and authority in discharging its ‘responsibility to protect’—there should be a genuine commitment to helping to build a durable peace, and promoting good governance and sustainable development.\textsuperscript{164}
\end{quote}

In the context of natural disasters, the responsibility to rebuild refers to the responsibility of disaster-affected States and the international community of States to cooperatively work towards “[e]nsuring [the] sustainable reconstruction and rehabilitation”\textsuperscript{165} of communities that have suffered extensive destruction caused by natural disasters. The responsibility to rebuild coincides with the U.N. Declaration on the Right to Development (“Right to Development”):

\begin{quote}
[A]n inalienable human right by virtue of which every human person and all peoples are entitled to participate in, contribute to, and enjoy economic, social, cultural and political development, in which all human rights and fundamental freedoms can be fully realized.\textsuperscript{166}
\end{quote}

The Right to Development imposes on all Member States the duty to improve the well-being of their populations, by ensuring “access to basic resources, education, health services, food, housing, employment, and the fair distribution of income.”\textsuperscript{167} The right to development, as it per-

\begin{footnotes}
163. \textit{Id.}
164. \textsc{The Responsibility to Protect}, \textit{supra} note 34, ¶ 5.1.
165. \textit{Id.} ¶ 5.2.
167. \textit{Id.} art 2, ¶ 3; art. 8, ¶ 1.
\end{footnotes}
tains to natural disasters, is the responsibility of all States to rebuild disaster-stricken communities by assisting in the post-disaster reconstruction phase and providing essential resources for disaster victims. The earthquake in Pakistan is an example of a natural disaster where all States recognized their responsibility to rebuild.\textsuperscript{168}

\textit{i. Pakistan’s Earthquake}

On the morning of October 8, 2005, a devastating earthquake, measuring 7.6 on the Richter scale,\textsuperscript{169} struck Northern Pakistan, triggering landslides and hundreds of continuous aftershocks.\textsuperscript{170} As of November 20, 2005, the earthquake had killed approximately 86,000 people, including more than 17,000 children, most of whom died as a result of being crushed by collapsing concrete roofs of poorly constructed schools;\textsuperscript{171} injured over 100,000 people; and caused destruction to infrastructure and housing, leaving an estimated 500,000 families homeless.\textsuperscript{172} Immediately after the earthquake, the Pakistani army and aid workers struggled to overcome logistical challenges that led to difficulties providing relief for disaster victims.\textsuperscript{173} During the weeks following the earthquake, conditions in Pakistan worsened as the estimated death toll increased and the provided aid to victims remained grossly insufficient.\textsuperscript{174}

\textsuperscript{168} Although not discussed more fully in this Note, another example of the failure to adhere to the responsibility to rebuild was apparent with the intense criticism that arose over the reconstruction efforts after Hurricane Katrina. While President Bush pledged the federal government’s commitment to “stay as long as it takes to help citizens rebuild their communities and their lives,” many state and local officials cited several stalled bills and policy changes, including measures to finance hurricane protection, revive small businesses, and compensate uninsured victims, all of which were vital to rebuilding New Orleans and encouraging evacuees to return to the city. James Dao, \textit{Louisiana Sees Faded Urgency in Relief Effort}, N.Y. TIMES, Nov. 22, 2005.


\textsuperscript{171} \textit{Id.}

\textsuperscript{172} \textit{Id.}


\textsuperscript{174} \textit{Id.} (noting that the estimated death toll climbed from 10,000-15,000 people on the first day to more than 50,000 people two weeks later. Additionally, funds from the international community, both governments and private organizations, were extremely slow to materialize.). Comparing the relief efforts in Pakistan with that of the Indian Ocean tsunami demonstrate the grossly inadequate international responses to catastrophic natural
In its relief efforts, Pakistan focused on two goals: (1) to provide adequate shelter to the 200,000 people in the high altitude villages, which would be cut off by snow; and (2) to improve the sanitation system of hundreds of refugee camps. The U.N. reaffirmed these goals in its relief operation’s reconstruction efforts, which involved providing adequate shelter, particularly for those communities in high altitudes, and rebuilding houses to withstand future natural disasters. However, insufficient funding significantly hindered the U.N. relief effort. Because the funding needed to rebuild the earthquake-devastated areas of Pakistan did not reach the requested level, U.N. Secretary General Kofi Annan called upon the entire international community, including governments, the private sector, and individuals, for the relief supplies, resources, and funding needed to adequately rebuild Pakistan’s communities.

Along with the U.N. relief effort, the U.N. High Commissioner for Refugees (UNHCR), an organization whose mandate is to assist refugees fleeing war and persecution and is not normally involved in natural disaster relief efforts, was entrusted to manage over 500 camps consisting of nearly 187,000 people. Moreover, the United States, through USAID and the U.S. military, remained committed to providing humanitarian disasters. "Quake ‘Is UN’s Worst Nightmare,’" BBC NEWS, Oct. 20, 2005, http://www.news.bbc.co.uk/go/pr/fr/-/1/hi/world/south_asia/4358902.stm (highlighting that while ninety-two countries provided humanitarian assistance for the Indian Ocean tsunami, only fifteen to twenty countries have responded to the earthquake in Pakistan).

175. Sengupta & Rohde, supra note 170.


177. Id. (quoting U.N. Secretary-General Kofi Annan’s plea for help, “[w]e need more resources, not just for emergency relief, but also for recovery and reconstruction . . . . We are going to do what we tend to call ‘recovery plus.’ Not just build what was there before, but build in a manner that can withstand . . . if another disaster struck.” Further noting that in its latest update, the “UN Office for the Coordination of Humanitarian Affairs (OCHA) reported that the UN’s $550 million flash appeal was still less than thirty percent funded, with only five percent of shelter needs and nine percent of water and sanitation requirements so far met.”); Sengupta & Rohde, supra note 170 (noting that dangers persisted as the U.N. relief operation remained grossly underfinanced).

178. Urging Greater Generosity, Annan Arrives in Pakistan for Quake Donor Conference, supra note 176 (quoting U.N. Secretary-General Kofi Annan, “I would expect the world, those with capacity, to be generous and to give, and give willingly and I am not just speaking to governments, I am also speaking to the private sector and individuals who have been generous in other situations. I call on all of them to help us here as well”).

assistance to the earthquake-devastated region during both the immediate emergency response phase and long-term reconstruction phase. In a joint news conference with Pakistani Foreign Minister Mian Khurshid Mahmood Kasuri, Secretary of State Condoleezza Rice stated:

[T]he United States will want to support[,] along with the international community[,] the people of Pakistan as they try to rebuild . . . . [S]o as we think about the immediate needs we will also start with the Pakistani Government to look to the future . . . . [O]ur thoughts are with you [the people of Pakistan] . . . in your hour of need . . . . [W]e will be with you not just today, but also tomorrow as you try to rebuild.\footnote{181}

The combined actions of Pakistan and the international community of States, through their support of the relief efforts of organizations such as the U.N. and the UNHCR, demonstrate the responsibility of disaster-affected States and the international community of States to work together, beyond the immediate relief phase, to rebuild disaster-devastated regions.

B. Problem: Authority for the Responsibility to Protect in Natural Disasters

These past natural disasters have demonstrated a regard among all States for the responsibility to prevent harm to disaster-affected populations, to react to the devastation, and to rebuild destroyed communities.\footnote{182} However, the issue remains of finding the basis for this collection of responsibilities under international law. The responsibility to protect is justified as arising from legal obligations, primarily imposed by international treaties and customary international law.\footnote{183} Since there are no in-

\footnote{180. USAID, US Continues Massive Humanitarian Response to Pakistan Earthquake, available at http://www.usaid.gov/pl/ (last visited Dec. 11, 2005) (noting that USAID has organized nine airlifts of emergency relief supplies consisting of 45,000 blankets, 1,570 winterized tents, plastic sheeting, water purification units, emergency kits, and over 4,800 metric tons of food. The U.S. military has delivered over 2,400 metric tons of relief supplies and evacuated 10,000 casualties.).}


\footnote{182. See discussion supra Part III.A.}

\footnote{183. Jack L. Goldsmith & Eric A. Posner, A Theory of Customary International Law, 66 U. CHI. L. REV. 1113, 1116 (1999) (recognizing two main forms of international law: treaty and custom); Coursen-Neff, supra note 86, at 693–702 (explaining in depth the two requirements for creating an obligation on States under customary international law: the opinio juris requirement—States’ acceptance that a certain rule is obligatory, and the State practice requirement—the general practice of States reflecting that the rule is obligatory); MACALISTER-SMITH, supra note 2, at 6 (noting that under the principles of}
ternational legal instruments that set forth the obligations of all States regarding natural disasters, it is logical to justify the responsibility to protect under customary international law.

Under customary international law, there must be general agreement among all States to be bound to the responsibility to protect and extensive and uniform State practice of this responsibility. While the international community of States and disaster-affected States, especially during the Indian Ocean tsunami, North Korea’s famine, Hurricane Katrina, and Pakistan’s earthquake, have demonstrated an interest to be bound to the responsibility to protect, this interest may not rise to the level of a legal duty since State practice of the responsibility to protect victims of natural disasters is likely insufficient under customary international law standards. Thus, justifying the responsibility to protect under customary international law may encounter difficulty meeting the State practice requirement.

IV. TOWARDS RECOGNIZING THE RIGHTS OF DISASTER VICTIMS

Within the international community, it has been widely recognized that in the event of a natural disaster, “States have the primary responsibility to protect the people and property on their territory.” When a disaster

international law, international instruments (rules, conventions, declarations, and resolutions), State practice, and international organizations, all of which address disaster relief actions, are relevant to determining “the potential extent of the role of law within global relief policy”).

184. See supra text accompanying note 90; Coursen-Neff, supra note 86, at 704 (noting that “[i]nternational law has progressed to the point where there is recognition of responsibility to disaster victims, but this recognition has not yet become legally binding”).

185. North Sea Continental Shelf (F.R.G. v. Den./F.R.G. v. Neth.), 1969 I.C.J. 3, 43, ¶¶ 74, 77 (Feb. 20) (stating that the “[t]he frequency, or even habitual character of [State] acts is not in itself enough [to constitute State practice under customary international law],” but that there must also be a sense of legal duty for the acts).

186. See discussion supra Part III.A.

187. See supra text accompanying note 90.

188. See Coursen-Neff, supra note 86, at 697 (noting that current “State practice thus lags behind the developing principle of responsibility to disaster victims”).

189. INT’L STRATEGY FOR DISASTER REDUCTION [ISDR], Hyogo Declaration, World Conference on Disaster Reduction (Jan. 18–22, 2005), at 2, ¶ 4, available at http://www.unisdr.org/wcdr/intergover/official-doc/L-docs/Hyogo-declaration-english.pdf. The goal of the ISDR is to build disaster-resilient communities through increasing public awareness of disaster risk and vulnerability, implementing disaster reduction policies and actions through the commitment of public officials, fostering the creation of partnerships among entities to expand risk reduction networks, and improving scientific knowledge about disaster reduction. ISDR: Mission and Objectives,
initially strikes, the first seventy-two hours are the most critical. During this initial relief phase, the responsibility of disaster-affected States is triggered, as the local and national response—in the form of civilian volunteers, local governments, and non-governmental organizations—provide the most immediately available relief to disaster victims. However, when disaster-affected States are unable or unwilling to provide relief to victims of natural disasters, help from the international community of States is required. Typically, the international community of States responds by unconditionally providing the humanitarian relief needed to protect the livelihoods of disaster victims. This emerging global recognition of the responsibility of all States to provide protection to natural disaster victims implies that these victims have certain rights deserving of protection.

Since it may be problematic to find sufficient State practice of the responsibility to protect natural disaster victims under customary international law standards, recognizing and codifying the rights of disaster victims within international human rights law would impose on all States a legal duty to protect these victims, thus justifying the responsibility to protect.

A. The Relationship Between Human Rights and Natural Disaster Victims

Acknowledging the applicability of human rights to natural disaster victims would expand the scope of international human rights and reinforce the regime of international human rights law. If natural disaster


191. See id.

192. See, e.g., supra Part III.A.2.

193. Id.

194. See Coursen-Neff, supra note 86, at 704 (noting that political factors, which envelop almost every natural disaster relief operation, may permanently obstruct State practice of international disaster relief from emerging into customary law).

195. Samuels, supra note 23, at 250. Arguments for applying human rights law to the area of refugees have also been made. Acknowledging the rights of refugees as human
victims possess rights to protection and all States recognize their legal obligation to protect those rights, then the international response to natural disasters would improve because all States would possess a sense of legal duty to protect the rights of disaster victims regardless of their political concerns. Recognizing and codifying the relationship between human rights and disaster victims would impose on all States a legal responsibility to protect natural disaster victims by: (1) preventing and mitigating adverse effects from disasters; (2) reacting to the needs of disaster victims by unconditionally providing and accepting needed humanitarian assistance; and (3) rebuilding disaster-stricken communities.

B. Legal Codification of the Rights of Natural Disaster Victims

In a General Assembly Resolution entitled Setting International Standards in the Field of Human Rights (General Assembly Resolution 41/120), the General Assembly reaffirmed the primary importance of the UDHR, the ICCPR, and the ICESCR to the framework of international human rights. The General Assembly recognized the value of further rights would recognize a legal right to asylum, thus strengthening the existing fundamental right to life and liberty and expanding the recognition of the UDHR, the ICCSR, and the ICESCR as applicable to all human beings, including refugees. Henkin, supra note 71, at 119. Since applying international human rights law to refugees would strengthen and expand the existing fundamental human rights framework, it logically follows that applying international human rights to natural disaster victims would have the same effect on the international human rights regime.

196. See generally Forsythe, supra note 22, at 276–79 (recognizing the political obstacles of international disaster relief and proposing four models of diplomatic approaches to cope with these political problems: (1) model one embraces the principle of State sovereignty and focuses on solving disaster relief problems by “improving traditional diplomacy between nations;” (2) model two is a “laisssez-faire vision of international disaster relief that curtails government involvement in society;” (3) model three is a “transnational vision of international disaster relief” that allows intergovernmental and nongovernmental organizations to “act for individuals rather than as instruments of governments;” and (4) model four envisions the drafting of a multilateral “treaty, or legal, conception of international disaster relief” that regulates relief). Under this categorical distinction, the proposed solution in this Note advocates for a combination of certain aspects of models one and four—reinforcing State sovereignty as the responsibility to protect the rights of natural disaster victims and codifying this principle under international law.

197. See Coursen-Neff, supra note 86, at 701–02 (explaining that the relationship between human rights and disaster victims would impose three legal obligations on States: (1) “the obligation to assist another in time[s] of natural disaster;” (2) “the obligation to prepare for disaster relief within [their] own territory and to take preventative measures in order to minimize the suffering resulting from natural disasters;” and (3) “the obligation to accept relief for [their] people from other [S]tates after the occurrence of a natural disaster, if [their] own resources are inadequate”).

developing the regime of international human rights law by discovering specific areas that call for international action with respect to human rights.\textsuperscript{199} Article 4 of General Assembly Resolution 41/120 sets forth guidelines for Member States and the U.N. bodies to follow when proposing new human rights instruments.\textsuperscript{200} Under Article 4, to establish a human rights instrument for natural disaster victims, the instrument must meet the following requirements:

1. \textit{be} consistent with the existing body of international human rights law;
2. \textit{be} of fundamental character and derive from the inherent dignity and worth of the human person;
3. \textit{be} sufficiently precise to give rise to identifiable and practicable rights and obligations;
4. \textit{provide}, where appropriate, realistic and effective implementation machinery, including reporting systems; \textit{and}
5. \textit{attract} broad international support.\textsuperscript{201}

The first requirement for establishing a human rights instrument for natural disaster victims is that the proposed rights of disaster victims be consistent with the current international human rights framework.\textsuperscript{202} In meeting the first requirement, the UDHR supports principles that are relevant to humanitarian assistance, that: (1) all States should prepare for disasters by taking preventative and precautionary measures to minimize suffering and destruction; (2) disaster-affected States should accept needed and neutral relief aid from the international community of States if their resources are inadequate; and (3) the international community of States and disaster-affected States should assist each other during emergencies.\textsuperscript{203} More specifically, Article 25 of the UDHR provides that:

\begin{quote}
Everyone has the right to a standard of living adequate for the health and well-being of himself and of his family, including food, clothing, housing and medical care and necessary social services, and the right to security in the event of unemployment, sickness, disability, widowhood, old age or other lack of livelihood in circumstances beyond his control.\textsuperscript{204}
\end{quote}

\textsuperscript{199.} Id.
\textsuperscript{200.} Id. ¶ 4.
\textsuperscript{201.} Id.
\textsuperscript{202.} Id. ¶ 4(a).
\textsuperscript{203.} MACALISTER-SMITH, \textit{supra} note 2, at 64.
\textsuperscript{204.} UDHR art. 25(1).
In the event of natural disasters, the UDHR recognizes that all States have a responsibility to protect the victims to ensure their right to humanitarian relief to improve the standard of living.

Along with the UDHR, Article 11 of the ICESCR also encourages the cooperation of all States to promote and ensure the right to humanitarian assistance, stating:

The States Parties to the present Covenant recognize the right of everyone to an adequate standard of living for himself and his family, including adequate food, clothing and housing, and to the continuous improvement of living conditions. The States Parties will take appropriate steps to ensure the realization of this right, recognizing to this effect the essential importance of international co-operation based on free consent.\(^{205}\)

It is logical to interpret “continuous improvement of living conditions” and “international cooperation” to mean that all States have a responsibility to work together towards improving the well-being of natural disaster victims. Specifically, this continuous cooperation among disaster-affected States and the international community of States should encompass the continuum of responsibilities outlined in the responsibility to protect.\(^{206}\) While neither the UDHR nor the ICESCR explicitly mention natural disasters, these covenants are nevertheless applicable to disaster victims because the devastating effects of natural disasters significantly affect the protection of the rights enshrined in these covenants.\(^{207}\) Therefore, the right of natural disaster victims to receive protection is consistent with the International Bill of Human Rights, thus satisfying the first requirement for establishing a human rights instrument.

The second requirement is that an international human rights instrument pertaining to natural disaster victims “derives from the inherent dignity and worth of the human person.”\(^{208}\) Natural disasters wreak grave destruction on communities, killing and injuring many people and destroying homes and infrastructure.\(^{209}\) This suffering is similar to that borne by victims of armed conflict and other gross violations of human rights who also experience death and displacement.\(^{210}\) Because the suffering of natural disaster victims parallels that of victims deserving of

---

206. See discussion supra Part III.A.
207. Samuels, supra note 23, at 248.
208. G.A. Res. 41/120, supra note 198, ¶ 4(b).
209. See, e.g., discussion supra Part III.A.
210. See sources cited supra note 78 and accompanying text.
humanitarian intervention, it is reasonable to surmise that disaster-affected States’ refusal of needed humanitarian assistance for natural disaster victims also “shock[s] the conscience of mankind.” Moreover, General Assembly Resolution 45/100 specifically pertains to humanitarian assistance for natural disaster victims and regards neglect of these victims as constituting a “threat to human life and an offence to human dignity.”

Therefore, affording human rights protections to natural disaster victims satisfies the second requirement because such protection derives from the recognition of their inherent dignity and worth.

The third requirement for establishing an international human rights instrument for natural disaster victims is that the instrument be precise enough to demonstrate “identifiable and practicable rights and obligations.” Applying the ICISS’ responsibility to protect theory to natural disasters reveals the rights and obligations of all States regarding disaster victims. The responsibility to protect identifies three obligations of disaster-affected States and the international community of States with respect to victims of natural disasters. Thus, under the rationale of the ICISS’ theory of the responsibility to protect, the third requirement is met because the responsibility to protect demonstrates “identifiable and practicable rights and obligations” of all States to protect natural disaster victims.

The fourth requirement is that a human rights instrument for natural disaster victims provides “realistic and effective implementation machinery, including reporting systems.” The current implementation methods for protecting the human rights of disaster victims exist through international treaties, which establish agreements on ways to implement effective disaster relief operations.

---

211. Scheffer, supra note 31, at 270–71 (arguing that a new definition of humanitarian intervention should encompass responses to natural disasters, such as earthquakes, floods, famine, volcanic eruptions, and man-made disasters because victims of these disasters suffer harms similar to those suffered by victims of oppressive governmental regimes who endure gross human rights abuses).
213. G.A. Res. 41/120, supra note 198, ¶ 4(c).
214. See discussion supra Part III.A.
215. G.A. Res. 41/120, supra note 198, ¶ 4(c).
216. Id., ¶ 4(d).
217. See generally INTERNATIONAL DISASTER RESPONSE LAWS, supra note 78, at 24–39 (noting that the IFRCRCS collected and analyzed various international disaster response treaties, which consisted of multilateral and bilateral treaties and U.N. resolutions. These treaties and resolutions addressed a broad range of disaster relief issues, relating to: (1) relationships between requesting and responding States; (2) responsibility and coordination of humanitarian assistance; (3) instructions for emergency relief teams; (4) access of
the Tampere Convention, which came into effect January 8, 2005 after unanimous adoption by seventy-five countries and ratification by thirty countries.218 The formation of the Tampere Convention resulted from the recognition of national regulations that hinder the import of telecommunications equipment, thereby causing delay in the provision of relief and ultimately loss of life.219 Its provisions describe the procedures for effectively deploying telecommunications resources to assist with disaster mitigation and relief operations.220 The Tampere Convention eases the restrictions imposed on telecommunications assistance by waiving certain regulatory barriers.221 The implementation of this international treaty recognizes the rights of natural disaster victims by ensuring an effective response from the telecommunications field.222 Another example is the aforementioned IOTWS program, which is a conglomeration of international agencies and national governments working towards establishing a tsunami warning system in the Indian Ocean.223 The Tampere Convention and the IOTWS are just some of the methods designed to provide effective protection of disaster victims, which thus demonstrates satisfaction of the fourth requirement.

The final requirement for establishing an international human rights instrument for natural disaster victims is that the instrument attracts “broad international support.”224 With the recent catastrophic natural disasters and the intense international criticism over disaster relief operations,225 proposing an international human rights instrument for natural disaster victims would garner support from several countries. Furthermore, a report by U.N. Secretary-General Kofi Annan highlights the world’s vulnerability to natural disasters and recognizes the need for collective ef-

219. Id.
220. Id.
221. Id. (explaining that the Tampere Convention: (1) waives various regulatory barriers that inhibit the import, coordination, and use of telecommunications equipment; (2) exempts relief agencies from taxation and duties; and (3) grants privileges and immunities to the NGO staff).
222. Id. (quoting Jan Egeland, U.N. Emergency Relief Coordinator and Operational Director of the Tampere Convention, “OCHA aims to ensure the best response to disasters to prevent loss of life and help survivors”).
223. See discussion supra Part III.A.1.i.
224. G.A. Res. 41/120, supra note 198, ¶ 4(e).
225. See discussion supra Part III.A.
forts to address their devastating effects. Therefore, this last requirement is met because it seems promising that launching a proposal for an international human rights instrument for natural disaster victims would receive extensive international support.

Implementing an international human rights instrument for natural disaster victims requires some sort of legal framework detailing its implications, which would be clarified once all States enter into an international agreement founded upon this principle. The recent catastrophic natural disasters demonstrate the need for an international treaty clarifying the responsibilities of all States regarding natural disasters.

An international disaster relief treaty is not a novel idea, but one that the Convention Establishing an International Relief Union (the Convention) pioneered in the 1930s. The Convention attempted to establish a legal framework for humanitarian assistance to natural disaster victims: the International Relief Union (IRU). The Convention founded the IRU on two principles: respect for territorial sovereignty and nondiscrimination in disaster assistance. Members of the Convention appointed an Executive Committee to control and organize relief operations through the services of the International Committee of the Red Cross and the League of Red Cross Societies. Article 2 of the IRU outlined the Convention’s mission:

- to furnish first aid, in the form of funds, resources, and assistance, to victims of natural disasters;
- to coordinate the efforts of disaster relief organizations;

227. Coursen-Neff, supra note 86, at 702 (recognizing that even if the rights of disaster victims were recognized as human rights norms, the implications of such rights on the role of disaster-affected States and the international community of States would remain unclear without a legal framework detailing their implications).
228. Id.; see Green, supra note 20, at 68.
229. In Larger Freedom, supra note 77, at 49–50, ¶¶ 202–08 (recognizing that events, in particular the Indian Ocean tsunami, have escalated the demands of the international humanitarian response system and demonstrated the need for more predictability in response capacity, funding, and right of access and security for humanitarian relief workers).
230. Forsythe, supra note 22, at 286.
231. Macalister-Smith, supra note 2, at 18. The League of Nations took up the IRU project in 1922. Forty-three member States attended the Conference for the Creation of an International Relief Union in Geneva and adopted the Convention and Statute of the IRU. Thirty States adhered to the Convention, which became effective on December 27, 1932. Id. at 19.
232. Id.
233. Id. at 19–20.
to encourage the study of preventative measures against disasters; and

to induce all people to render mutual international assistance.234

The IRU’s objectives seemed sensible. Although as a practical matter, it failed to provide the means for supplying relief to disaster victims, which brought about its demise.235 Later attempts to revive the IRU in the 1940s also failed.236 Nevertheless, the IRU initiative raised awareness concerning the need for an international legal framework that clarifies the responsibility of all States to provide humanitarian assistance to natural disaster victims.237

V. CONCLUSION

The proposed solution to the issue of providing adequate relief for catastrophic natural disasters attempts to clarify the responsibilities of all States to protect disaster victims.238 The suggested scheme entails: (1) applying the responsibility to protect to natural disasters; (2) recognizing and codifying the rights of disaster victims within the regime of international human rights law; and (3) establishing an international treaty that imposes on all States the responsibility to protect natural disaster victims.

During the past decade, the world has endured the wrath of natural disasters, which have caused substantial destruction to numerous countries and severe devastation to mass amounts of people.239 Now is the time to revive the vision originally embarked upon by the IRU.240 Due to the

235. Macalister-Smith, supra note 2, at 20–21 (noting that the IRU’s failure was due to inadequate funding and its premature expression of universal solidarity aimed at benefiting disaster victims).
236. Id. at 95.
237. Id. at 20–21 (noting that the creation of the IRU, an organization focused primarily on providing humanitarian assistance for natural disaster victims, raised awareness of the issue of international disaster relief among the international community).
238. Green proposes a noteworthy approach to establish international responsibility (a shared responsibility among States) regarding natural disasters. First, the proposal calls for delegating “the authority and responsibility” of representing the needs of disaster victims “to one [independent] agency,” which must be agreed upon by the international community (State governments and international and private organizations). Second, the proposal emphasizes the need for the international community to establish a formal mechanism that recognizes international responsibility for humanitarian assistance during natural disasters. Green, supra note 20, at 65–76.
239. See discussion supra Part III.A.
240. See Green, supra note 20, at 15–17 (noting that while the occurrence of natural disasters has not increased in frequency, but remained relatively stable, the worldwide
emerging principle of the responsibility to protect and the universal recognition of international human rights, an international agreement recognizing how these ideals apply to natural disasters would improve humanitarian assistance. Such an agreement would hopefully obligate all States to defer to their responsibility to protect disaster victims rather than their political agendas, thus de-politicizing international disaster relief. Establishing an international treaty that clarifies the responsibilities of all States to protect natural disaster victims would require a great deal of international diplomacy. Achieving such international cooperation would be difficult to accomplish in the near future. However, increased awareness among all States concerning their responsibility to protect natural disaster victims would improve the imperfect system of international disaster relief in our vulnerable world.

Tyra Ruth Saechao*

human cost of natural disasters has gradually escalated, as disasters have become more complex, causing destruction to diverse regions and populations. Green further notes the need to consider natural disasters not as an “isolated” problem, but as a “multidimensional” occurrence that compels different approaches and responses in handling their consequences. Id. at 77. In 1989, the General Assembly declared the 1990s to be the International Decade for Natural Disaster Reduction, which was an attempt to “reduce through concerted international action, especially in developing countries, the loss of life, property damage, and social and economic disruption caused by natural disasters” and improve disaster prevention, mitigation, and response. G.A. Res. 236, U.N. GAOR, 44th Sess., U.N. Doc. A/RES/44/236/Annex (1989); Coursen-Neff, supra note 86, at 656, n.42; Levitt, supra note 61, at 176 (expressing the imperative with which the international community should recognize the responsibility to protect by inquiring, “[h]ow many more millions of people must die before the international community [recognizes the responsibility to protect as] a global imperative?”).

241. McClatchey, supra note 64, at 679–80 (noting that scholars have recently begun focusing on the value of international diplomacy, rather than the merits of international law, to reinforce all States to recognize their responsibilities to humanity).

242. Id. at 678–79 (recognizing the view that despite the binding force of international law, States will continue to act in their own best interests and not that of their people, especially when matters concerning their sovereignty arise).

* B.A., University of Michigan, Ann Arbor (2004); J.D., Brooklyn Law School (expected June 2007). I would like to dedicate this Note to my parents, Tula and Ruth Saechao, and my brother, Troy Saechao, whose love, support, and encouragement have kept me going through law school. I would like to thank Professor Claire Kelly and Professor Jayne Ressler for their invaluable guidance and advice; Lea La Ferlita and Sujeet Rao for reviewing and critiquing my earlier drafts; Aaron Warshaw, Jennifer McGovern, and the editorial staff of the Brooklyn Journal of International Law for their editing assistance; and my friends for providing many pleasant distractions. All errors are my own.
BALANCING ACT: WILL THE EUROPEAN COMMISSION ALLOW EUROPEAN FOOTBALL* TO REESTABLISH THE COMPETITIVE BALANCE THAT IT HELPED DESTROY?

I. INTRODUCTION

Some people think that [European] football is a matter of life and death. I don’t take that attitude. I can assure them it is much more serious than that.¹

This sentiment reflects the special place that European football holds for so many. The sport, which was first organized by London schools in 1863, has united Europe for nearly 150 years.² Today, fifty-two European nations have national European football associations, most of which organize professional leagues within their respective nation.³

Aside from the sport’s social significance, European football has become a prominent part of Europe’s entertainment industry. The commercialization of European football is often traced to the 1980s when television recognized European football as valuable content, and advertising revenue began to flow into the sport.⁴ In 2006, Deloitte & Touche estimated that across Europe, professional football generated £7.8 billion in total revenue.⁵ It seems clear, however, that the sport was not prepared for such economic growth. With the commercialization of European football came rising operating costs, most noticeably in the form of

---

* For purposes of this Note, the term “European football” refers to the sport known as soccer in the United States. The term is used here to refer specifically to the professional game within Europe.

¹. This famous quote was first uttered by Bill Shankly, former manager of the English club Liverpool FC. Jens Pelle van den Brink, EC Competition Law and the Regulation of European Football, 7 SPORTS LAW. J. 105, 105 (2000).


player salaries. These higher stakes have put small-market clubs\(^6\) out of business, while many other clubs struggle to turn a profit.\(^7\) The few clubs that are financially stable generally operate in large markets in Europe. As the financial gap between large and small-market clubs grows, the sport itself has begun to suffer. Today, large-market clubs dominate European football in leagues across all of Europe, leaving the majority of clubs with no hope of remaining competitive. This lack of competitive balance in European football is the focus of this Note.

Within the professional sporting world it is generally accepted that there must be a competitive balance among teams in order to preserve the integrity of sporting competition, the interest of fans, and in turn, commercial success. In any sport, if a few elite teams are able to collect all of the best players so that other teams cannot provide reasonable competition on the field, results will become predictable, and spectators, sponsors, advertisers, and broadcasters will all lose interest.\(^8\) Thus, when the unrestricted market for players fosters a competitive imbalance among teams, it is within the interest of sporting associations, leagues, and organizations to enforce restrictions so as to restore a competitive balance.\(^9\)

To some extent, European law has enhanced European football’s growing lack of parity. More than a century ago the transfer system was created as a means to instill competitive balance among clubs in European football.\(^{10}\) This system remained largely in place until the 1995 landmark case Union Royale Belge des Societes de Football Association ASBL v. Jean-Marc Bosman, where the European Court of Justice (ECJ)\(^{11}\) held

---

6. Throughout this Note, the term “club” shall be used to refer to European football teams.


9. Id.

10. James G. Irving, Red Card: The Battle over European Football’s Transfer System, 56 U. MIAMI L. REV. 667, 668 (2002); see also infra Part II.B.

11. The European Court of Justice (ECJ) is an institution of the European Community. The ECJ, comprised of judges from each European Community member state, functions to ensure that Community law is interpreted and enforced uniformly by all member states. The ECJ has the power to settle disputes between Community member states, Community institutions, businesses, and individuals, as well as to make preliminary rulings on interpretation of European Community law at the request of national courts. The ECJ also includes Advocate Generals who provide detailed recommendations or “advisory opinions” to the court before it rules on novel issues of law. See Treaty Establishing the European Community (consolidated text), arts. 220–45, 2002 O.J. (C 325) 33 [hereinafter EC Treaty].
that the transfer system violated Article 48 of the Treaty of Rome. Subsequently, the European Commission (EC) negotiated rule changes bringing the transfer system in line with the Bosman decision and with European competition law. Unfortunately, these changes seem to have undermined the transfer system’s effectiveness. The resulting lack of parity among European football clubs has predictably led to a small group of elite clubs that dominate competition perennially. In turn, leagues in countries such as England and Italy are experiencing large drops in attendance. The current poor state of European football has prompted high-ranking people within the sport to call for limits on player salaries as a means of reestablishing competitive balance among clubs. However, questions remain as to whether a salary cap or other restrictive measures would be a violation of European competition law. Will the EC allow European football to reestablish the competitive balance that it helped destroy?

The purpose of this Note is to evaluate whether or not restricting the amount of money European football clubs can spend on player salaries in an effort to revitalize competitive balance would violate European competition law. Part II of the Note traces how the ECJ and EC have contributed to the sport’s lack of competitive balance. Part III looks at salary caps and other policies meant to institute balance. Part IV reviews European competition law and governance of sport in Europe. Part V analyzes whether salary restrictions would violate European competition law, and whether salary restrictions could be exempt from EC regulation. Part VI is a conclusion.


13. The European Commission (EC) is the executive arm of the European Community. As such, it proposes legislation to parliament, manages and implements Community policies and budgets, and enforces European Community law. See EC Treaty, supra note 11, arts. 211–19. For purposes of this note, it is significant that the EC generally regulates and enforces EC competition law. See Irving, supra note 10, at 672.

14. See Irving, supra note 10, at 688–723; see also infra Part II.D.

15. Stratis Camatsos, European Sports, the Transfer System and Competition Law: Will They Ever Find a Competitive Balance?, 12 SPORTS LAW J. 155, 178 (2005); see also infra Part II.E.


17. See Nick Harris, Whelan Calls for a Top Flight Salary Cap, INDEPENDENT (U.K.), Sept. 22, 2005, available at http://sport.independent.co.uk/football/news/article314199.ece; see also infra Part III.
II. THE PROBLEM: A LACK OF COMPETITIVE BALANCE IN EUROPEAN FOOTBALL

A. The Structure of European Football

In order to understand European football and its relationship with the law, a basic understanding of the structure of the sport would be helpful given that it differs from the American professional sports model. At the top of the hierarchy is the Federation Internationale de Football Association (FIFA) which governs football at the world level. FIFA is divided into confederations which govern each continent. The European confederation is the Union des Associations Europeennes de Football (UEFA). UEFA is comprised of the national associations of each country in Europe. National associations must comply with UEFA regulations and decisions, and UEFA in turn, is subject to FIFA regulations. National associations govern virtually all aspects of football within their respective countries, including enforcement of FIFA and UEFA regulations, organization of leagues, the relationship between leagues and clubs, and the relationship between clubs and players.

Professional European football leagues are organized by nation and are generally structured quite differently from American leagues. The European approach features a divisional hierarchy within each league, and a system of promotion and relegation. First division clubs play at a higher level than second division clubs, which play at a higher level than third division clubs, and so on. In all divisions, at the end of the season, a limited number of the worst performing clubs (usually between one to four clubs) are relegated to the immediately lower division, while the

19. Id.
20. Id.
21. Id.
22. Id.
23. Id.; see, e.g., The Football Association—The Organisation, http://www.thefa.com/TheFA/TheOrganisation (outlining the role of the English national association, the FA).
25. See id. Taking England as our example, the top division in England is called the “Premier League” (or “Premiership”), which features the top twenty clubs in England. Below the Premier League is the English second division known as “Championship” which consists of the next best twenty-four clubs. Below Championship is “League One,” and below that is “League Two”—both of which also feature twenty-four clubs. See BBC Sport—Football, http://news.bbc.co.uk/sport1/hi/football/default.stm (last visited Jan. 20, 2007).
same number of top performing clubs in each division are promoted to the immediately higher division.\footnote{26. See Szymanksi & Valletti, supra note 24. Returning to the English system as our example, at the end of the season, the three worst performing clubs in the Premier League are relegated to Championship, and the top three clubs in Championship are promoted to the Premier League. See Premiership Standings—2005/06, http://soccernet.espn.go.com/tables?league=eng.1&cc=5901. Promotion and relegation also occur between Championship and League One, and between League One and League Two.}

Aside from professional leagues, UEFA organizes pan-European competitions that bring together the top clubs from each national league, including the UEFA Champions League—the winner of which is considered the champion of Europe.\footnote{27. See UEFA—Champions League Format, http://www.uefa.com/competitions/ucl/Format/index.html (last visited Jan. 25, 2007).} These competitions generally follow a format closer to the American model where clubs are split into groups and play only those clubs within their group.\footnote{28. See, e.g., id.} At the end of the group stage, the winner and runner-up of each group advance to a series of knock-out rounds, similar to playoffs in American sports, where the winner advances and the loser is eliminated.\footnote{29. See, e.g., id.} The last club remaining is the winner of the competition.

\textbf{B. The Transfer System}

The importance of comparatively level teams for the success of sporting leagues was recognized as far back as the late nineteenth century when the transfer system was first created in England.\footnote{30. See Irving, supra note 10, at 668.} The purpose of the system was to control player movement so that wealthier clubs could not buy the best players away from smaller clubs without compensation.\footnote{31. Id. at 669.} Over time, the transfer system was adopted by leagues across all of Europe.\footnote{32. See id.} Under the rules of the transfer system, at the end of a season, each club produced a list of players to be retained for the next season (retention list) and a list of players who were available for transfer to another club (transfer list).\footnote{33. Id.} Players that were not included on the transfer list had no right to demand a transfer.\footnote{34. Id.} Players that were on the transfer list could be purchased by another club for a “transfer fee” set by
Although the transfer system has been modified over time, this fundamental structure has remained unchanged. It is significant that players whose contracts had expired were not exempt from the transfer system. Transfer rules were set out by both FIFA and UEFA, and incorporated into the rules of national associations which had some discretion as to how they were enforced. As a result, club control over players who were no longer under contract varied from country to country within Europe. Ultimately, however, UEFA and FIFA rules required that a transfer fee be paid to the former club at some point, in order to sign a player whose contract had expired. By allowing clubs to receive transfer fees for players, whether under contract or not, the transfer system enabled small-market clubs to either retain their top players or be compensated for transferring them, and thus compete with big market clubs. In short, the transfer system helped impose competitive balance throughout European football.

C. The Bosman Decision

In December of 1995, Jean-Marc Bosman, a little-known football player from Belgium, challenged the legality of the transfer system. The result was a decision from the ECJ that would fundamentally change the transfer system and the governance of sport in Europe.

Jean-Marc Bosman played for FC Liege, a Belgian first division club, and his contract was set to expire in June of 1990. On April 21, 1990, FC Liege offered Bosman a new contract that reduced his salary from 120,000 Belgian francs (BFR) per month to BFR 30,000 per month. Bosman rejected the offer and was placed on a transfer list with a transfer fee of more than BFR 11 million. This fee was calculated in accor-

35. Id.
36. Camatsos, supra note 15, at 157; see also Irving, supra note 10 (outlining in detail the development of the transfer system from its inception through Bosman).
38. Compare Irving, supra note 10, at 672 (noting that by 1978 England’s transfer system allowed a player whose contract had expired to sign with a club of the player’s choice even if his new club and his former club could not agree on a transfer fee, in which case transfer tribunals would determine the fee), with Bosman, 1995 E.C.R. at I-5047 (noting that the Belgian transfer system, like most European systems, allowed clubs to set a transfer fee for a player whose contract had expired, and such a player could not sign with a new club until the new club agreed to pay the transfer fee).
42. Id.
dance with the transfer rules as set out by the Union Royale Belge des Societes de Football Association ASBL (URBSFA), Belgium’s national association. By June 1, no clubs had showed interest in Bosman given his high transfer fee, so according to URBSFA rules, a “free” transfer period began during which a club could negotiate a mutually agreeable transfer fee with FC Liege. Once the free transfer period began, Bosman arranged a contract with US Dunkerque, a second division French club, at BFR 100,000 per month, plus a BFR 900,000 signing bonus. US Dunkerque and FC Liege then agreed on a one-year transfer of Bosman for BFR 1,200,000, with an option to purchase a full transfer after one year. However, FC Liege, concerned about US Dunkerque’s solvency, withheld required paperwork and the transfer never took effect. On July 31, FC Liege suspended Bosman for the entire 1990–1991 season, as was its right under URBSFA transfer rules.

Ultimately, after four years of lawsuits, the Belgian national courts referred two questions to the ECJ for a preliminary ruling. One of the questions focused on whether Articles 48, 50, 85, and 86 of the Treaty of Rome prohibited “a football club from requiring and receiving payment of a sum of money upon the engagement of one of its players who has come to the end of his contract by a new employing club.” Many within Europe felt that the very survival of football was at stake.

43. Id.
44. See id. at I-5046, I-5050.
45. Id. at I-5050–I-5051.
46. Id. at I-5051.
47. Id.
48. Id.
49. Id. at I-5056.
50. Article 48 of the Treaty of Rome (Treaty Establishing the European Economic Community) is now embodied in Article 39 of the Treaty Establishing the European Community. Article 39 guarantees freedom of movement for workers within the European Union. See EC Treaty, supra note 11, art. 39.
51. Articles 85 and 86 of the Treaty of Rome (Treaty Establishing the European Economic Community) are now embodied in Articles 81 and 82 (respectively) of the EC Treaty. For the text of Article 81 see infra note 126. For the text of Article 82 see infra note 118.
52. Bosman, 1995 E.C.R. at I-5056. The other question concerned whether or not foreign players could be restricted from European Community clubs and competitions. See id. This question is, however, outside the scope of this Note. See Lindsey Valaine Briggs, UEFA v. The European Community: Attempts of the Governing Body of European Soccer to Circumvent EU Freedom of Movement and Antidiscrimination Labor Law, 6 CHI. J. INT’L L. 439 (2005), for a discussion of this issue before, during, and after the Bosman decision.
53. See Irving, supra note 10, at 681 (detailing the sentiment of the football world in anticipation of the ECJ’s decision).
After determining that the transfer system constituted an obstacle to the freedom of movement of workers, the ECJ held that

Article 48 of the Treaty precludes the application of rules laid down by sporting associations, under which a professional footballer who is a national of one Member State may not, on the expiry of his contract with a club, be employed by a club of another Member State unless the latter club has paid to the former club a transfer, training or development fee.\footnote{Bosman, 1995 E.C.R. at I-5073.}

In reaching its conclusion, the ECJ accepted the importance of “maintaining a balance between clubs by preserving a certain degree of equality and uncertainty as to results. . . .”\footnote{Id. at I-5071.} However, the ECJ found that the transfer system did not adequately maintain a competitive balance between clubs because it did not prevent the richest clubs from acquiring the best players.\footnote{Id.} Furthermore, the ECJ held that there were alternative means of imposing a competitive balance that did not impede freedom of movement of workers, including instituting a salary cap and/or revenue-sharing policies.\footnote{Id. at I-5072.} Note that the ECJ here is referencing the advisory opinion of Advocate General Lenz. Recall that as part of the ECJ’s structure, Advocate Generals issue advisory opinions before the ECJ rules on a novel issue. See EC Treaty, \textit{supra} note 11. In his opinion on the \textit{Bosman} case, Advocate General Lenz set forth that the possibility of a salary cap as a means of preserving competitive balance without restricting freedom of movement of workers meant that the transfer system was in violation of the Article 48 of the Treaty of Rome. See \textit{Bosman}, 1995 E.C.R. at I-5017.

Ironically, after expressing support for a salary cap, the ECJ declined to apply competition law under Articles 85 and 86 of the Treaty of Rome, since the transfer system was already found to be in violation of Article 48.\footnote{Id.}

\textbf{D. European Commission Rule Changes}

The European football world, though not happy with the ruling in \textit{Bosman}, took comfort in the fact that the ECJ’s decision invalidated the transfer system only as it applied to players whose contracts had expired.\footnote{Irving, \textit{supra} note 10, at 688.} Almost ninety percent of transfer revenue came from transfers of players who were still under contract, and FIFA and UEFA aimed at keeping what remained of the transfer system intact.\footnote{Id. at 689.} However, the \textit{Bosman} decision, as highly public as it was, put the football world under

\begin{itemize}
\item \footnote{Bosman, 1995 E.C.R. at I-5073.}
\item \footnote{Id. at I-5071.}
\item \footnote{Id.}
\item \footnote{Id. at I-5072. Note that the ECJ here is referencing the advisory opinion of Advocate General Lenz. Recall that as part of the ECJ’s structure, Advocate Generals issue advisory opinions before the ECJ rules on a novel issue. See EC Treaty, \textit{supra} note 11. In his opinion on the \textit{Bosman} case, Advocate General Lenz set forth that the possibility of a salary cap as a means of preserving competitive balance without restricting freedom of movement of workers meant that the transfer system was in violation of the Article 48 of the Treaty of Rome. See \textit{Bosman}, 1995 E.C.R. at I-5017.}
\item \footnote{Id.}
\item \footnote{Irving, \textit{supra} note 10, at 688.}
\item \footnote{Id. at 689.}
\end{itemize}
a microscope, and the EC, which had not generally involved itself in the 
sports industry prior to *Bosman*, was now suspicious of the entire transfer 
system.\(^6^1\) The EC was beginning to question whether transfer fees might 
violate both Article 48 free movement of workers\(^6^2\) and European com-
petition law.\(^6^3\) In April of 1998, after the much publicized and high-
priced transfer of Brazilian striker Ronaldo from FC Barcelona to Inter 
Milan, the EC warned FIFA that it would face official action if it did not 
amend its transfer rules.\(^6^4\) Lengthy, complex, and trying negotiations 
involving the EC, FIFA, UEFA, and the players’ union (FIFPro) ensued.\(^6^5\) 
Finally, in 2001, six years after the ECJ decided *Bosman*, FIFA 
adopted a new transfer system.\(^6^6\) A brief summary of the more signifi-
cant rule changes follows.

One key rule provides that transfers can only take place during two 
limited periods of the year (transfer windows), one during the summer, 
and the other mid-season.\(^6^7\) Furthermore, a player can only be transferred 
once in a single season.\(^6^8\)

In an effort to protect smaller clubs which usually develop younger 
players, new rules require purchasing clubs to compensate the selling 
club for the training of a player under the age of twenty-three.\(^6^9\) The 
amount of the compensation fee is to be calculated according to a codi-
ified formula.\(^7^0\) However, no compensation will be paid for transfers of

---

\(^{61}\) Id.

\(^{62}\) See id.

\(^{63}\) *EU Warning to FIFA on Rules of Transfer*, STATESMAN (India), Apr. 3, 1998.

\(^{64}\) Id. The entire football world paid close attention to the transfer of Ronaldo who, 
at barely 20 years old, was considered the best player in the world. In 1997, Ronaldo Luis 
Nazario de Lima, or “Ronaldo” as he is known internationally, became the first player to 
win back-to-back FIFA World Player of the Year awards. See *A Night of Records and 
events/playergala/index_E_1996.html. Considered by many to be the best Brazilian 
scorer since Pelé, Ronaldo went on to win the award again in 2002. See *Ronaldo, Hamm 
MR_A/48852_E.html.

\(^{65}\) See Irving, *supra* note 10, at 688–722 (tracing, in depth, the negotiations from 
*Bosman* through to the 2001 agreement).

\(^{66}\) Id. at 716–17; FIFA, *Regulations for the Status and Transfer of Players*, July 5, 
Status_Transfer_EN.pdf.

\(^{67}\) Id. at ch. 3, art. 5(2).

\(^{68}\) Id.

\(^{69}\) Id. at ch. 7, art. 15.

\(^{70}\) Id. at ch. 7, art. 16.
players over the age of twenty-three.\textsuperscript{71} To further protect smaller developmental clubs, when a player is transferred mid-contract, five percent of the compensation paid to the selling club will be distributed among all clubs that trained the player between the ages of twelve and twenty-three.\textsuperscript{72}

Additionally, a number of rules address contract stability.\textsuperscript{73} The most significant of these rules allows players under twenty-eight years old to breach a contract unilaterally after the first three years, while players over twenty-eight years old can unilaterally breach after the first two years.\textsuperscript{74}

\textbf{E. The Lack of Competitive Balance in European Football}

Though the ECJ and the EC sought to protect a competitive balance among football clubs, the \textit{Bosman} decision and the subsequent transfer system rule changes have only widened the gap between large and small-market clubs.\textsuperscript{75} The \textit{Bosman} decision liberalized player movement and shifted power from clubs to players.\textsuperscript{76} This, in turn, has triggered a dramatic increase in player salaries.\textsuperscript{77} Adding to the problem is the new rule allowing players unilaterally to breach their contracts, which forces clubs to pay top players astronomical salaries to prevent them from changing sides.\textsuperscript{78} As a result, top players pool together on big-market clubs.\textsuperscript{79}

The rule creating transfer windows has also come under heavy scrutiny. Ironically, transfer windows both restrict freedom of movement and restrain competition, and thus seem to conflict with the \textit{Bosman} decision and European competition law.\textsuperscript{80} Furthermore, transfer windows allow big-market clubs to hold onto their money so that they can buy all of the top players available when the windows are open, rather than allowing the market to dictate when and where players’ services are most valued.\textsuperscript{81} It is therefore more likely that small-market clubs will be pushed out of the market for top players when the windows are open.

\textsuperscript{71} \textit{Id.} at ch. 7, art. 20.
\textsuperscript{72} \textit{Id.} at ch. 9, art. 25.
\textsuperscript{73} See \textit{id.} at ch. 8.
\textsuperscript{74} See \textit{id.} at ch. 8, art. 21(1)(a) & (1)(b).
\textsuperscript{75} See Camatsos, \textit{supra} note 15, at 178.
\textsuperscript{76} \textit{Id.} at 173.
\textsuperscript{77} \textit{Id.}
\textsuperscript{78} See \textit{id.} at 173–74.
\textsuperscript{79} See \textit{id.} at 174–75.
\textsuperscript{80} \textit{Id.} at 170.
\textsuperscript{81} \textit{Id.}
The rule prohibiting transfer fees for players over the age of twenty-three also effectively diminishes competitive equality. Big-market clubs can choose to exclusively pursue players twenty-four or older, and as a result never have to compensate small-market clubs for training.\textsuperscript{82}

As a result of \textit{Bosman} and the EC-mandated rule changes, a select few clubs have dominated their respective leagues in recent years. In England’s Premier League, four clubs—Arsenal, Chelsea, Liverpool, and Manchester United—have finished in the top five spots the last four seasons. In two of those four seasons they were the top four clubs, and in the current 2006–2007 season, they once again occupy the top four spots.\textsuperscript{83} Similarly, in Italy’s Serie A, three clubs—Juventus, Inter Milan, and AC Milan—have finished in the top four spots in each of the past four seasons—finishing in the top three in three of those four seasons.\textsuperscript{84} In Germany, Bayern Munich has won the championship three out of the last four years, finishing second in 2003–2004.\textsuperscript{85} In France, Lyon has been champion for four straight seasons, and is currently the leader in the 2006–2007 season.\textsuperscript{86} Finally, in Holland, PSV Eindhoven has been

\textsuperscript{82} \textit{Id.} at 171. It should be noted that the 2001 rule changes do not apply to amicable transfers. Thus, when the player, and both the selling and purchasing club agree on a transfer fee, large sums will continue to change hands regardless of a player’s age. \textit{See} Irving, \textit{supra} note 10, at 724.


\textsuperscript{84} In the final standings from the 2005/06 Serie A season, Juventus and AC Milan do not appear in the top two spots (though they finished with the two best records) due to penalties the league assessed for their involvement in a match-fixing scandal. The extent to which their respective records were due to their underhanded arrangement rather than their play on the field will never be known. \textit{See} ESPN SoccerNet—Italian Serie A Standings—2005/06, http://soccernet.espn.go.com/tables?league=ita.1&season=2005&column=none&order=false&cc=5901, (for 2004/05 standings follow “2004/05” hyperlink; for 2003/04 standings follow “2003/04” hyperlink; for 2002/03 standings follow “2002/2003” hyperlink) (last visited Jan. 25, 2007).


Such complete lack of parity has clearly had a negative effect on the sport. A recent Italian match between giants Juventus and Inter Milan drew a crowd of less than half the capacity of Juventus’ home stadium in Turin.\footnote{See Juventus Deserve Bigger Crowds Says Moggi, supra note 16.} This has caused the president of Italy’s national association to call for a reduction in ticket prices.\footnote{Italian Clubs Need to Cut Prices, REUTERS (Rome), Oct. 5, 2005, http://soccernet.espn.go.com/news/story?id=344995&cc=5901.} England’s Premier League is facing similar problems, and has drawn up a working group to deal with dwindling crowds.\footnote{See Premier League Probes Crowd Slump, supra note 16.} A member of the working group has pointed to “predictable results” as a possible cause for the attendance drop.\footnote{Id.} So extreme is the lack of balance in England that seven matches into the 2005–2006 season an Irish betting agency had already declared Chelsea the winner of the league with thirty-one matches left to play, and began paying out on single bets made on the defending champions.\footnote{Bookmaker Pays Out on Chelsea Title Victory, REUTERS (London), Sept. 29, 2005, http://soccernet.espn.go.com/news/story?id=344274&cc=5901.} Clearly something must be done to impose balance among clubs and restore uncertainty as to results—something to make football matches meaningful again. But what?
III. A SOLUTION PROPOSED: THE SALARY CAP

In September of 2005, after watching Chelsea dominate league play over the last season and a half of the Premier League, a small-market club chairman, Dave Whelan of Wigan Athletic, publicly announced the need for a salary cap to restore meaningful competition to the league.96 He went on to claim that four other club executives would support a cap,97 and later claimed that elite clubs Arsenal and Manchester United would possibly support one as well.98 Shortly thereafter, Whelan proved to be correct when Arsene Wenger, the manager of Arsenal, went public with his support for salary restrictions.99 But perhaps the clearest sign that European football is ready to consider salary limits came in the form of an article written by Joseph Blatter, the president of FIFA. In it, Blatter claims that European football has become a “society of have and have nots,” and hints at setting limits on player salaries.100 Blatter concludes by announcing that he has commissioned a FIFA task force to deal with the excessive spending that is “suffocating” the game.101

---

96. Harris, supra note 17.
97. Id.
98. Gordon Tynan, Arsenal and United May Back Cap, Says Whelan, INDEPENDENT (U.K.), Sept. 23, 2005, http://sport.independent.co.uk/football/news/article314430.ece. While Whelan was expressing support for a salary cap in England’s Premier League, in order for a cap to be successful it would probably have to be implemented by UEFA and apply to all of Europe. A cap in one European country alone would result in a movement of top players from that country to other countries whose leagues do not have a salary cap and can thus offer higher salaries. Furthermore, this phenomenon could easily be characterized as affecting trade between Community member states, and thus European competition law would still likely apply. See infra note 135 and accompanying text. As a result, though this Note considers a salary cap for all of European football, its basic inquiry would be equally relevant to a salary cap in one European country.
101. Id. Interestingly, these recent calls for salary caps are not the first time restrictions on player salaries have been considered by European football. In 2002, an organization of eighteen of the most powerful clubs in Europe known as G-14 agreed on a “cost control” measure whereby G-14 clubs would not spend more than 70% of their total revenue on player salaries. See Top Clubs Agree Salary Cap, BBC SPORT (U.K.), Nov. 5, 2002, http://news.bbc.co.uk/sport1/hi/football/europe/2402329.stm. Though the media referred to it as a “salary cap,” the agreement only applied to G-14 members. Furthermore, the agreement was a response to the widespread financial problems clubs were facing at the time—it was not meant to foster competitive balance. See id. Effectively, the G-14 agreement has been replaced by UEFA’s licensing system which includes financial criteria also aimed at ensuring that clubs are solvent and credible, rather than dealing with issues of competition. See UEFA, UEFA Club Licensing System Manual at 82, March,
Though it is not clear what solutions FIFA’s task force may suggest, based on success in other sports—particularly league sports here in the United States—a salary cap would be an effective mechanism for revitalizing European football’s lacking competition. All four major pro-


For more information on the G-14, see G-14, Who We Are, http://www.g14.com/main.php (follow “Who We Are” hyperlink) (noting that the purpose of G-14 is to give its 18 member clubs a voice in the development of international club football) (last visited on Jan. 25, 2007); G-14, Who We Are—Basics, http://www.g14.com/main.php (follow “Basics” hyperlink) (noting that the G-14 is mainly concerned with issues of “EC employment legislation, sponsorship and TV rights, intellectual copyright, format and management of club and national team competitions”) (last visited Jan. 25, 2007).

102. It is perhaps surprising to learn that salary caps are in effect in other sports within Europe. Several rugby leagues within Europe, including the Zurich Premiership, French rugby, and the Super League have salary caps ranging from £1.7 million to £2.25 million per season. Additionally, in European football, lower English divisions—League One and League Two—have voluntary salary caps whereby clubs refrain from spending more than 60% of their total revenue on player salaries. Richard Mooney & Marjorie Holmes, A Question of Sport: Does US Treatment of Football Offer Any Solutions for the UK?, COMPETITION L. INSIGHT, Aug. 9, 2005, at 7; Harris, supra note 17. The reason that such leagues have caps and have not been investigated by the EC is that they draw media attention and engage in economic activity on a much lower scale than does football at its highest national divisions. As such, a salary cap in European football’s “top-flight” would surely draw the attention of the EC. See Mooney & Holmes, supra.

103. Salary caps in U.S. league sports do not conflict with U.S. antitrust law because in the United States, unlike in Europe, contracts between leagues and players are collectively bargained by the league and players’ unions. The National Labor Relations Act provides that a majority of employees in a unit may elect a representative of all employees in such unit for purposes of collective bargaining. 29 U.S.C. § 159 (1959). Salary caps that are included in collective bargaining agreements (CBAs) have been deemed permissible since invalidating a cap in such cases would unravel the entire CBA, and ultimately frustrate the purpose of collective bargaining. See Wood v. Nat’l Basketball Ass’n, 809 F.2d 954, 961–62 (2d Cir. 1987).

104. Taking U.S. league sports as our model once again, there are policies other than the salary cap that are widely used to promote competitive balance. One such alternative is revenue sharing. Revenue sharing plans generally require teams to put a limited amount of their local revenue into a league pool which is then distributed so that revenues are channeled to teams with below league-average revenues. Major League Baseball (MLB), the National Football League (NFL), and the National Hockey League (NHL) all have revenue sharing plans. See MLB, 2003–2006 Basic Agreement, Sept. 30, 2002, art. XXIV, available at http://mlbplayers.mlb.com/pa/pdf/cba_english.pdf [hereinafter MLB CBA]. On November 10, 2006, shortly before publication of this Note, MLB and the players association (MLBPA) came to terms on a new CBA that will be in effect for the 2007–2011 seasons. See Press Release, MLB, MLB, MLBPA Reach Five-Year Labor Accord (Nov. 24, 2006), available at http://mlb.mlb.com/NASApp/mlb/news/press_releases/press_release.jsp?ymd=20061024&content_id=1722380&vkey=pr_mlb&f
As of the publication of this Note, the MLB 2007–2011 CBA has not been made publicly available by MLB. Therefore, for purposes of this Note, reference will be made to the MLB 2003–2006 CBA which is no longer in effect. All of the portions of the MLB 2003–2006 CBA that are discussed herein, including the competitive balance tax, revenue sharing, and the amateur draft, are still in effect under the MLB 2007–2011 CBA, though some changes to each policy have been made.; NFL, *Collective Bargaining Agreement Between the NFL Management Council and the NFL Players’ Association*, Mar. 8, 2006 (as amended), art. XXIV, § 11, available at http://www.nflpa.org/pdfs/Agents/CBA_Amended_2006.pdf [hereinafter NFL CBA]; NHL, *Collective Bargaining Agreement Between National Hockey League and National Hockey League Players’ Association*, July 22, 2005, art. 49, available at http://www.nhl.com/cba/2005-CBA.pdf [hereinafter NHL CBA]. However, revenue sharing in the United States is fostered by the strength of professional leagues, and the close relationship between teams and players as evidenced by collective bargaining. As leagues in Europe tend to be looser unions—thus the system of promotion and relegation—where clubs are more independent of one another, revenue sharing is not likely to prove workable. Furthermore, revenue sharing is in some ways more extreme than a salary cap insofar as a cap limits a team’s spending, while revenue sharing actually takes a piece of an owner’s profits and redistributes it. Given the financial troubles European football is currently experiencing, the few owners that do turn a profit would be unlikely to agree to a policy that takes a cut of their profit and gives it to a competing club. That said, in Europe, broadcast rights to live football matches are generally sold “collectively” by the league, which then distributes the proceeds amongst the clubs. This can be seen as a form of revenue sharing. This practice is often referred to as “collective selling” or “joint selling.” Unfortunately for those who stress competitive balance on the field, joint selling is largely frowned upon by the EC as anti-competitive off the field since it limits media coverage of live matches, and prevents clubs from competing in the sale of rights to matches (oh, the irony!). See Ivo Van Bael & Jean-François Bellis, *Competition Law of the European Community* 1441–43 (4th ed. 2005) for a discussion of the EC’s views on joint selling.

Another alternative balancing mechanism to a salary cap employed by U.S. leagues is the player draft. Drafts regulate how new players enter a league, and are generally weighted so that the order in which teams draft is directly related to performance from the prior season. Poorly performing teams are allowed to draft ahead of better performing teams, meaning they can select (and thus control) the top talents entering the league. The system thus promotes competitive balance. See Lewis et al., supra note 8, at 855–56. MLB, the NBA, the NFL, and the NHL all have player drafts. See MLB CBA, supra, Attachment 24, at 202; NBA, *Collective Bargaining Agreement*, art. X, July 29, 2005, available at http://www.nbpa.com/cba_articles.php [hereinafter NBA CBA]; NFL CBA, supra, art. XVI; NHL CBA, supra, art. 8. By contrast, in European football, new players can be contracted by any club, and thus often sign with the highest bidder. Since the clubs with the most money to offer are often the clubs that traditionally win, this practice fosters a competitive imbalance. Still, it is unlikely that a player draft could be instituted in Europe without a major reorganization of European football. Unlike U.S. league sports, the market for players in European football spans multiple leagues in multiple countries. Organizing one draft for all of Europe is impractical, while organizing a draft in only one or two leagues would likely motivate young players to play in leagues that do
fessional sports leagues in the United States have some form of salary cap or cap-like policy in their current collective bargaining agreements (CBA). Of those four leagues, there is perhaps no better example of how the salary cap can balance competition, and in turn enhance the popularity and profitability of a sport, than the NFL. In the twenty-one-year period from 1972–1993, seven of the twenty-eight NFL franchises won the Super Bowl a combined total of twenty times, making the NFL a league of dynasties. Since the NFL instituted the salary cap in 1993, nine different teams have won the thirteen Super Bowls. During the “salary cap era” the NFL has become the most profitable sport in the United States, and, to the enjoyment of many fans, has also become synonymous with parity.

not have drafts where they are free to sign with the highest bidder, and have some control over where they play.

Since these alternative methods of promoting competitive balance do not seem practical for European football to adopt, this Note only considers mechanisms that limit spending on player salaries such as a salary cap.

105. They are Major League Baseball (MLB), the National Basketball Association (NBA), the National Football League (NFL), and the National Hockey League (NHL).

106. MLB has a “competitive balance tax”. See MLB CBA, supra note 104, art. XXIII. Note that the “competitive balance tax” remains a part of the current MLB 2007–2011 CBA. See Press Release, MLB, MLB, MLBPA Reach Five-Year Labor Accord, supra note 104. The NBA, NFL, and NHL all have salary caps. See NBA CBA, supra note 104, art. VII; NFL CBA, supra note 104, art. XXIV; NHL CBA, supra note 104, art. 50, § 50.5.


108. The Cowboys (three), Dolphins (two), 49ers (four), Giants (two), Raiders (three, twice in Oakland and once in Los Angeles), Redskins (three), and Steelers (four) won all but one Super Bowl from 1972–1993. The Bears in 1985 were the only team to win once and never repeat. See Super Bowl Recaps, http://www.superbowl.com/history/recaps (last visited Jan. 26, 2007).


110. Only the Broncos (twice), Cowboys (twice), and Patriots (three times) have won the Super Bowl more than once since the salary cap was introduced. See Super Bowl Recaps, supra note 108.

The basic idea of a salary cap is rather simple: limit the amount of money a team can spend on player salaries. However, in U.S. leagues, salary caps are often complex policies that reflect the profitability of the sport, as well as the balance of power between the league and the players’ union, and thus vary from league to league. This Note will consider two types of salary caps as possible solutions for European football: the “hard cap” and the “luxury tax.” The NFL institutes a so-called hard cap, whereby no team can spend more than the designated limit on total player salary at any time. The hard cap is the most restrictive form of team salary cap, and the most effective at striking competitive balance. It completely separates financial resources from success on the field.

A luxury tax, by contrast, is technically not a salary cap, but a cap-like restriction, since it does not impose an absolute ceiling on player salary.

112. In U.S. sports leagues, the designated limit that a team can spend is usually equal to a percentage of the average team revenue. By contrast, those who have talked about having a cap in Europe often support an agreement whereby each club will not spend more than a fixed percentage of its own revenue as opposed to the average revenue. See Wenger Supports Top-Flight Salary Limit, supra note 99; Top Clubs Agree Salary Cap, BBC SPORT (U.K.), supra note 101. It is easy to see how such an agreement is self-serving of big-market clubs and would do little to effectively create a competitive balance. A limited percentage of the revenue of big-market clubs, which tend to take in more revenue, will almost always afford them much greater spending room than that same percentage of revenue of small-market clubs.

113. This Note does not consider other types of caps, including the “soft cap.” Soft caps set a limit on spending, however, teams are allowed to spend over that limit to some extent. The NBA has a soft salary cap whereby teams can use league-approved “cap exceptions” to spend over the cap limit. See NBA CBA, supra note 104, art. VII, § 6. The NBA allows nine exceptions which teams often use. Soft caps are not considered here because they generally require defined situations when a team is allowed to spend over the cap limit, and the policy considerations required to define such cases in European football are beyond the scope of this Note.

114. See NFL CBA, supra note 104, art. XXIV, § 4; id. art. XXV.

115. A distinction can be made between “individual caps,” which limit the amount of money a team can pay an individual player, and “team caps,” which limit the amount of money a team dedicates to total player salaries. Both the NBA and NHL have individual as well as team salary caps. See NBA CBA, supra note 104, art. VII, § 5; NHL CBA, supra note 104, art. 50, § 50.6. Individual player caps are more restrictive than team salary caps since under a team cap there is no limit as to what a team owner can pay an individual player so long as the amount paid on total player salaries does not exceed the cap limit. Since an individual cap accomplishes a similar end, yet is more restrictive than a team cap, it is more likely to be struck down by European competition law, and thus individual caps are not considered here.
Under MLB’s luxury tax system, a spending threshold is set, and teams can choose to spend over the threshold, but will be taxed a fixed percentage for every dollar so spent.\footnote{116} Such policies keep spending lower, but do not totally restrict owners’ freedom to spend, and thus bigger market teams will still usually maintain a slight advantage.

Though it seems that a salary cap (hard cap or luxury tax) can effectively deal with European football’s lack of competitive balance, and there is support for such a measure among prominent figures within the sport, doubts remain as to whether the EC would view a salary cap as a violation of European competition law.\footnote{117} The next two parts of this Note explore this question. Part IV will review relevant European law, while Part V will apply the law to both a hard cap and a luxury tax, and contrast outcomes.

IV. EC COMPETITION LAW AND SPORTS

A. Articles 81 and 82

Articles 81 and 82 of the EC Treaty govern competition in pan-European markets. We will first examine Article 82, which prohibits abuse of a dominant position within the European market that may affect trade between member states.\footnote{118} Then we will take up the more complex

\footnote{116. See MLB CBA, \textit{supra} note 104, art. XXIII, § B. Interestingly, MLB has named its luxury tax the “competitive balance tax.” The tax proceeds are used by the league to pay for player benefits, to fund projects to promote baseball in developing countries and areas where baseball is not widely played, and to reinforce MLB’s Industry Growth Fund. \textit{See} MLB CBA, \textit{supra} note 104, art. XXIII, § H.}

\footnote{117. This concern is based on the fact that European football clubs are competitors in the market for players, and thus a salary cap would be seen as an agreement among competitors.}

\footnote{118. Article 82 reads as follows:

Any abuse by one or more undertakings of a dominant position within the common market or in a substantial part of it shall be prohibited as incompatible with the common market in so far as it may affect trade between Member States. Such abuse may, in particular, consist in: (a) directly or indirectly imposing unfair purchase or selling prices or other unfair trading conditions; (b) limiting production, markets or technical development to the prejudice of consumers; (c) applying dissimilar conditions to equivalent transactions with other trading parties, thereby placing them at a competitive disadvantage; (d) making the conclusion of contracts subject to acceptance by the other parties of supplementary obligations which, by their nature or according to commercial usage, have no connection with the subject of such contracts.

\textit{EC Treaty, supra} note 11, art. 82. For a comprehensive explanation of Article 82 of the EC Treaty, see \textit{VAN BAEI & BELLIS, supra} note 104, at 115–32.}
Article 81. Case law defines “dominant position” under Article 82 as “a position of economic strength enjoyed by an undertaking [119] which enables it to prevent effective competition being maintained on the relevant market by giving it the power to behave to an appreciable extent independently of its competitors, customers and ultimately of its consumers.” [120] In *Bosman*, Advocate General Lenz, in his advisory opinion, [121] applied Article 86 of the Treaty of Rome [122] to the transfer system. [123] Ultimately Lenz concluded that the transfer system did not violate Article 86 because it restricted competition between clubs insofar as they can contract players, and players are not “competitors, customers or consumers” in relation to clubs. [124] Similarly, a salary cap is a restriction of competition between clubs in the market for players, and since it would not affect the relationship between clubs and “competitors, customers or consumers,” it follows that there would be no violation of Article 82 of the EC Treaty.

The real challenge to the legality of a salary cap in European football comes from Article 81, which prohibits agreements between undertakings that restrict competition and affect trade between member states. [125]

---

119. For purposes of Articles 81 and 82, an “undertaking” is defined quite broadly as “every entity engaged in an economic activity regardless of the legal status of the entity and the way in which it is financed.” See *Case C-41/90, Höfner & Elsner v. Macrotron GmbH*, 1991 E.C.R. I-1979, ¶ 21.


121. Recall that as part of the structure of the ECJ, Advocate Generals provide “advisory opinions” to the court before it rules on novel issues. *See supra* note 11.

122. Article 86 of the Treaty of Rome is now embodied in Article 82 of the EC Treaty. *See supra* note 51.


124. *Id.* at I-5038–I-5039.

125. Article 82 reads as follows:

The following shall be prohibited as incompatible with the common market: all agreements between undertakings, decisions by associations of undertakings and concerted practices which may affect trade between Member States and which have as their object or effect the prevention, restriction or distortion of competition within the common market, and in particular those which: (a) directly or indirectly fix purchase or selling prices or any other trading conditions; (b) limit or control production, markets, technical development, or investment; (c) share markets or sources of supply; (d) apply dissimilar conditions to equivalent transactions with other trading parties, thereby placing them at a competitive disadvantage; (e) make the conclusion of contracts subject to acceptance by the other parties of supplementary obligations which, by their nature or according to commercial usage, have no connection with the subject of such contracts. 2. Any agreements or decisions prohibited pursuant to this Article shall be automatically void. 3. The provisions of paragraph 1 may, however, be declared inapplicable in the case of:—any agreement or category
Article 81 accomplishes its purpose in three parts: 81(1) lays out conduct that is prohibited as anti-competitive; 81(2) renders conduct that falls within 81(1) automatically void; and 81(3) sets forth requirements for exemptions to Article 81. Articles 81(1) and 81(3) are somewhat complex and for purposes of this Note, they will be broken down into their general material requirements.

The “agreements” prohibited by 81(1) are broadly interpreted so as to encompass any joint intention of undertakings regarding specific conduct in a market. In addition to agreements between undertakings, anti-competitive decisions made by “associations of undertakings” are also prohibited by 81(1).

Agreements fall within 81(1) only if they prevent, restrict, or distort competition. Agreements can prevent, restrict, or distort competition by either their object, or effect. Furthermore, 81(1) only applies if...
the ultimate effect of the restriction on competition is “appreciable.” Additionally, this appreciable restriction must affect a defined common market. Finally, it must not be forgotten that Article 81 only applies to agreements that affect trade between member states.

Agreements within Article 81(1) are automatically voided by Article 81(2), unless they qualify for an exemption under 81(3). Article 81(3) allows agreements to stand if their pro-competitive benefits outweigh their restrictive effects. An agreement must satisfy four conditions in


132. If an agreement is not restrictive of competition by object, its effect may restrict competition if it “affect[s] actual or potential competition to such an extent that on the relevant market negative effects on prices, output, innovation or the variety or quality of goods and services can be expected with a reasonable degree of probability.” See id.

133. Although the requirement of an appreciable effect on competition is not included in the text of 81(1), it is formalized in the Commission’s De Minimis Notice of 2001. Accordingly, an agreement between actual or potential competitors will not be deemed to appreciably restrict competition if the aggregate market share held by the parties to the agreement does not exceed 10%. Such agreements will not be caught by 81(1). See European Commission, Commission Notice on Agreements of Minor Importance Which Do not Appreciably Restrict Competition Under Article 81(1) of the Treaty Establishing the European Community (De Minimis), 2001 O.J. (C 368) 13.

134. Defining relevant markets requires a two-part analysis—defining the relevant product market, and defining the relevant geographic market. Defining the relevant product market involves assessing “all those products and/or services which are regarded as interchangeable or substitutable by the consumer, by reason of the products’ characteristics, their prices and their intended use.” Defining the relevant geographic market involves assessing “the area in which the undertakings concerned are involved in the supply and demand of products or services, in which the conditions of competition are sufficiently homogeneous and which can be distinguished from neighbouring areas because the conditions of competition are appreciably different in those areas.” See European Commission, Commission Notice on the Definition of the Relevant Market for the Purposes of Community Competition Law, 1997 O.J. (C 372) ¶¶ 7–8.

135. Establishing an effect on trade between member states involves three elements—the concept of “trade,” establishing “effect,” and establishing that the effect is appreciable (note, this third analysis is independent of the finding that an agreement works an “appreciable” restriction of competition). “Trade” is defined broadly to include “all cross-border economic activity.” European Commission, Guidelines on the Effect on Trade Concept Contained in Articles 81 and 82 of the Treaty, 2004 O.J. (C 101) 81, 83. Establishing an “effect” requires that “it must be possible to foresee with a sufficient degree of probability on the basis of a set of objective factors of law or fact that the agreement or practice may have an influence, direct or indirect, actual or potential, on the pattern of trade between Member States.” Id. Appreciability is a fact-sensitive inquiry that considers “the nature of the agreement and practice, the nature of the products covered and the market position of the undertakings concerned.” Id. at 85.

order to qualify for an exception under Article 81(3). The first condition is that the agreement must create efficiency gains by improving production or distribution of goods or services, or by promoting technical or economic progress. The second condition requires that a fair share of the benefits created by the agreement be passed on to consumers. The third condition under 81(3) is that the restrictions must be indispensable to achieving the efficiencies created. Lastly, under the fourth condition of 81(3), the agreement in question cannot afford the parties the opportunity to eliminate “competition in respect of a substantial part of the products concerned.”

B. Competition Law and Sports

In the years after the Bosman decision, as the EC sharpened its focus on European football and sports in general, it recognized that the application of competition law must take into account certain characteristics that are unique to sports. The EC is not concerned with regulations that are essential to a sport. If the regulation in question is inherent to a particular sport, its organization, or the organization of competitions within that sport, so that without such a regulation the sport would not be able to exist, the regulation will not fall within Article 81(1) regardless of its restrictive effect on competition. Thus, if the hard cap or the luxury tax is determined to be essential to European football, Article 81(1) will

---

137. Id.
138. Only objective benefits to the market are measured in applying the first condition. Specifically, the Commission will take into account the nature of the claimed efficiencies; the causal link between the agreement and the efficiencies; the likelihood and magnitude of the claimed efficiencies; how and when each of the claimed efficiencies would be achieved; and any costs incurred in achieving the efficiencies. See id. at 104–05.
139. “Consumers” under 81(3) are the customers of the parties to the agreement in question. See id. at 109.
140. This condition functions as a least restrictive means test, whereby there cannot be any other economically practical and less restrictive means of creating the same efficiencies. However, it is important to note that the Commission holds that it will not question the business judgment of the parties by analyzing hypothetical or theoretical alternatives. Rather, the Commission will only intervene under this condition where it is reasonably clear that other realistic and attainable alternatives are available. See id. at 107–08.
141. Determining whether competition is eliminated depends on the degree of competition prior to the agreement, and on the degree to which competition is reduced by the agreement. See id. at 113. “Competition” for purposes of this final condition of Article 81(3) is referring to competitors in the industry, who are not a party to the agreement in question.
not apply. The EC applied this test in dismissing a complaint against UEFA in 2002. ENIC, a company that owned controlling interest in five European football clubs, petitioned the EC under Article 81 to strike down a UEFA rule that prohibited one company from owning more than one club in UEFA competitions. The EC found that though the rule had a restrictive effect, such effects were “inherent in the pursuit of the very existence of credible pan European football competitions.” Since the rule was necessary and proportionate to the need to maintain the public’s perception that UEFA competitions were genuine, and without such a perception long-term competition would prove impossible, the rule fell outside Article 81(1).

The EC has also recognized that competition in sports is different from competition in other industries. In sports, it is not within a competitor’s economic interest to put other competitors out of business, rather competition is necessary for the very existence and success of sports. This means that some interdependence among competitors is inherent in sports.

Furthermore, as the ECJ concluded in Bosman, the EC has recognized the need to preserve uncertainty as to results and to maintain a degree of equality among competitors as legitimate aims that are essential to sports.

V. ANALYSIS

The following analysis considers the legality of a UEFA-implemented hard salary cap, and a UEFA-implemented luxury tax in light of Article 81. It is, to some extent, unnatural to consider both policies since UEFA

144. The Commission found that without such a rule, two clubs with one owner could be forced to play each other in UEFA competitions. As a result, fans could lose confidence in the honesty of the competition, and ultimately lose interest which would devalue the sport. Id. ¶ 32.
145. Id. ¶ 38. Though the EC has held that this inquiry is unique to sports, the underlying logic does extend to virtually all sectors. If an industry cannot exist without certain restrictions on competition, those restrictions must be allowed to stand. See Wouters, 2002 E.C.R. at ¶¶ 73–110, where the ECJ held that a prohibition by the Dutch Bar on partnerships between bar members and other professions (including accountants) did not fall within Article 81(1) because it was reasonable to conclude that such a regulation was necessary for the proper practice of law within the Netherlands.
146. By contrast, in almost all other industries, it is within an undertaking’s interest to eliminate weaker competition and take over their market share. See Monti, supra note 142.
147. See id.
is likely to adopt one or the other, but not both.\footnote{149} The EC would then review the legality of whichever policy UEFA chose to implement.\footnote{150}

Insofar as the analysis of the hard cap and the luxury tax overlap, they are analyzed together in order to avoid redundancy. For the most part, however, they are considered separately. The first part of the analysis applies Article 81(1), and considers whether the hard cap or the luxury tax can be characterized as essential to European football and thus fall outside of Article 81(1). The second part applies Article 81(3), and the third part offers additional considerations.

\textit{A. Article 81(1)}

Both a UEFA hard cap and a luxury tax would satisfy many of the requirements of Article 81(1). UEFA is an “association of associations of undertakings.”\footnote{151} Therefore a hard cap or luxury tax imposed by UEFA would be a decision by an association of associations of undertakings within the meaning of Article 81(1).

---

\footnote{149} Although it is likely that UEFA itself would conduct an analysis similar to ours as it must assess the legality of all policies that it considers adopting.\footnote{150} On May 1, 2004, the procedure for implementing European competition law was revised. New procedures give competition authorities of member states, national courts of Community member states, and the EC the power to implement Articles 81 and 82. See Council Regulation (EC) No. 1/2003 of 16 December 2002 on the Implementation of the Rules on Competition Laid Down Under Articles 81 and 82 of the Treaty, 2003 O.J. (L1) 7–9. Prior to these changes UEFA would have been able to notify the EC that it adopted a luxury tax or a hard cap, and request that the EC review the legality of its action thereby securing temporary immunity from fines even if the EC determined that the policy in question violated competition law. Now, however, notifications are no longer possible, and UEFA must assess the legality of its actions at the risk of being fined for violations of competition law. See \textit{Van Bael & Bellis}, supra note 104, at 1024–25. Notwithstanding these rule changes, it is likely that the EC and not national courts or national competition authorities would review UEFA actions because they often affect many member states. See \textit{id.} at 1026–27.\footnote{151} \textit{See ENIC/UEFA, supra note 143, ¶ 25.} Professional clubs are undertakings insofar as they engage in economic activity. National associations which group clubs together are associations of undertakings. Thus, UEFA, which groups together national associations within Europe, is an association of associations of undertakings.
Both a hard cap and a luxury tax are forms of horizontal agreements that typically fall within 81(1) because they restrict competition in either object or effect. While UEFA could argue that the narrow object of a hard cap or a luxury tax is to improve competitive balance within European football, it would have to concede that the effect of such regulations restricts competition in the market for players. In the case of a hard cap, a club that is at or near the spending limit would not be able to compete with other clubs for the services of any player whose salary would bring the club above the limit. In the case of a luxury tax, a club that is at or near the tax threshold, though free to compete for the services of players, will be taxed for taking on salary in excess of the threshold, and thus competition is at least distorted, if not inhibited.

Clearly the market for players is one of the relevant markets affected by a hard cap or luxury tax. This market represents the “upstream” market where clubs compete to purchase players’ services which are necessary for the finished product: a football match. However, the market for players is not the only relevant market under 81(1) analysis. The “downstream” selling market where football is sold to spectators, media, and other consumers, would also be affected. This market would benefit from an improved product—more competitively balanced football with less predictable results—as a result of a salary cap or luxury tax. The effect of a UEFA hard cap or luxury tax on both the upstream market for players and the downstream selling market is sure to be appreciable since clubs organized under UEFA represent 100% of the professional European football market.

Finally, a hard cap or luxury tax would certainly affect trade between member states. One example of such an effect can be seen in situations where clubs from one member state transfer players to clubs in other member states. Under a hard cap, a club from one member state will not be able to transfer a player to a club in another member state if the salary

---

152. “Horizontal” agreements are agreements entered into by companies “operating at the same level(s) in the market.” See European Commission, Guidelines on the Applicability of Article 81 of the EC Treaty to Horizontal Cooperation Agreements, 2001 O.J. (C 3) 2. A hard cap or luxury tax would be an agreement among clubs, with all clubs operating at the same level within the market for European football—all clubs are buyers in the sense that they have to contract players, a stadium, etc., and sellers in that they participate in matches and competitions, to which they can sell tickets, broadcast rights, etc.

153. For a description of the process of defining relevant markets, and use of the terms “upstream” and “downstream” markets, see Van Bael & Bellis, supra note 104, at 143–44.

154. Indeed, UEFA has been accused of operating as a monopoly in professional European football. See, e.g., ENIC/UEFA, supra note 143, ¶ 15.
of the player being transferred would raise the transferee club’s payroll above the spending limit. In the case of a luxury tax, while a transfer could go through even if the transferee club’s payroll would exceed the tax threshold, the transferee club would be taxed, and possibly refrain from such a transfer as a result.

Thus far, both a UEFA hard cap and a UEFA luxury tax would fall within Article 81(1). However, it is still necessary to consider whether such restrictions can be characterized as inherent to European football. The decisive question is this: can European football, particularly national competitions, continue to exist in the long term without a hard cap or a luxury tax? If the answer is yes, then 81(1) will apply and we must continue our analysis by applying 81(3). However, if the answer is no, then 81(1) will not apply, the restrictions will be consistent with competition law, and our inquiry will end.

The basic argument that some restriction on player salary is necessary to the long-term existence of European football can be used to defend either a hard cap or a luxury tax. UEFA can claim that without such restrictions, the lopsided dominance of large-market clubs in recent years will continue. A continued imbalance in competition on the field means that results will remain predictable, and spectators will lose interest in the sport. Ultimately, European football will lose all marketability as spectators, advertisers, and broadcasters all turn away. Essentially, UEFA would be arguing that without restrictions on the upstream market, where clubs purchase players’ services, the downstream market, where football matches are sold, would cease to exist. Since the upstream market is dependent on the downstream market, restrictions on competition in the upstream market (a hard cap or a luxury tax) are necessary. This argument is strong in light of the fact that it comports with the EC’s reasoning in dismissing ENIC’s claim against UEFA, and promotes

155. See supra text accompanying notes 83–95.
156. The EC held in relevant part:

Without the UEFA rule, the proper functioning of the market where the clubs develop their economic activities would be under threat, since the public’s perception that the underlying sporting competition is fair and honest is an essential precondition to keep its interest and marketability. If UEFA competitions were not credible and consumers did not have the perception that the games played represent honest sporting competition between the participants, the competitions would be devalued with the inevitable consequence over time of lower consumer confidence, interest and marketability. Without a solid sporting foundation, clubs would be less capable of extracting value from ancillary activities and investment in clubs would lose value.
aims that both the EC and the ECJ have recognized as legitimate.\textsuperscript{157} However, this argument is too broad, and the EC is sure to push UEFA to justify its choice of one policy over the other. Taking a closer look at a hard cap and a luxury tax separately, each policy has different strengths and weaknesses that the EC would certainly consider, and which it may find dispositive.

Once UEFA establishes that there is a lack of competitive balance in European football and that the consequences are potentially devastating to the sport, it is difficult to argue that a hard salary cap would not substantially benefit the sport. By separating player budget from revenue, a hard cap would effectively redistribute top talent among clubs so that a competitive balance on the field is restored. However, a hard cap is perhaps the most restrictive spending limit employed in professional sports, placing an unconditional ceiling on spending. The EC, which stresses that restraints on competition be proportional to their desired effects,\textsuperscript{158} would likely frown on the harshness of a hard cap. As a result, less restrictive policies would be preferred alternatives. UEFA, in defending a hard cap, would be forced to argue that less restrictive policies would not provide enough of a competitive balance to render results unpredictable. However, the hard cap is such an extreme measure, one that would fundamentally alter the balance of power in European football, that the EC would likely have a hard time accepting the argument that without such drastic change the sport could not continue to exist. This is particularly true in light of the fact that less restrictive practices have effectively balanced competition in other sports.\textsuperscript{159} Ultimately, because of its restrictive

\textsuperscript{157}See supra note 146 and accompanying text.

\textsuperscript{158}Before dismissing ENIC’s complaint against UEFA, the EC held that “the rule does not seem to go beyond what is necessary to ensure its legitimate aim. . . .” ENIC/UEFA, supra note 143, ¶ 41. Similarly, in Bosman, Advocate General Lenz stated in his opinion that only restrictions which are “indispensable” to achieving a legitimate aim fall outside of 81(1). See Bosman, 1995 E.C.R. at I-5033.

\textsuperscript{159}MLB’s luxury tax is an obvious alternative. MLB first instituted the luxury tax in the 2002–2006 CBA which recently expired. The tax has coincided with unprecedented parity among MLB teams both economically, and in terms of on-the-field results. In the five years from 2002–2006 that the luxury tax has been in effect, five different teams have won the World Series. See Barry M. Bloom, MLB, Union Announce New Labor Deal, MLB.COM, Nov. 10, 2006, http://mlb.mlb.com/news/article.jsp?ymd=20061024&content_id=1722211&vkey=ps2006news&fext=.jsp&c_id=mlb. Furthermore, on the heels of signing a new five-year CBA which extends the luxury tax, players’ salaries in

ENIC/UEFA, supra note 143, ¶ 32. Recall that in this decision the EC was considering a UEFA rule prohibiting one entity from owning controlling interest in more than one club participating in the same UEFA competition. Public perception that the clubs participating in a competition are balanced so that the outcome is not entirely predictable is likewise a precondition to continued interest in, and marketability of European football.
effect on competition, the EC is likely to reject the argument that a hard cap is necessary to European football. As a result, the hard cap is likely to fall within Article 81(1), and will be rendered void unless it qualifies for an exemption under 81(3).

Turning to the luxury tax, its strengths and weaknesses are the inverse of those of the hard cap. Since a luxury tax only penalizes clubs for spending over the tax threshold, and imposes no absolute limits on a club’s ability to spend, the EC will not find that it greatly restricts competition. Still, the EC might question the effectiveness of a luxury tax. A luxury tax will allow big market clubs the choice to pay the tax in order to keep top players. If a luxury tax cannot create enough of a competitive balance to fix European football’s problem, UEFA cannot claim that it is essential to the economic viability of the sport. However, it is unlikely that the EC would press this issue since defeating the luxury tax for its lack of effectiveness would only encourage UEFA to adopt a more restrictive practice such as the hard cap.\(^\text{160}\) The EC is more likely to defer to UEFA’s judgment that a luxury tax will balance competition. Assuming that the EC accepts UEFA’s basic argument that player salaries need to be restricted, there is a strong possibility that the EC would find that a luxury tax is essential to credible competition in national European football leagues. Thus, a luxury tax would fall outside of Article 81(1). How-

---

\(^{160}\) Other alternatives such as revenue sharing and a player entry draft have already been dismissed. \(\text{See supra note 104.}\)
ever, for purposes of this Note, we will also apply Article 81(3) to the luxury tax to determine if it might qualify for an exemption in the event the EC would apply Article 81(1).

B. Article 81(3)

Recall that exemptions under Article 81(3) are efficiency-based in that restrictions are allowed to stand if the benefits created outweigh the restrictions on competition. This inquiry not only considers what benefits are created for consumers, but also whether the restrictions in question are the least restrictive means of creating such benefits. Ultimately, 81(3) analysis will mirror the 81(1) analysis, and as a result, the respective strengths and weaknesses of a hard cap and a luxury tax considered in 81(1) analysis will remain significant in 81(3) analysis. Our inquiry here is guided by Article 81(3)’s four conditions. The fourth condition—that the restriction in question cannot eliminate competition in respect to the products concerned—can be dismissed outright since it is clear that neither a hard cap nor a luxury tax imposed by UEFA would prevent clubs that are not organized under UEFA from competing in the market for players’ services. In considering 81(3)’s first three conditions, the hard cap and the luxury tax will be taken up separately.

In the case of a hard cap, it seems that the first two conditions of an 81(3) exemption will be easily satisfied. Read together, the first two conditions require that a hard cap creates a benefit that is passed on to the customers of the parties to the hard cap agreement. It has been established that a hard cap would limit the amount any club could spend on players in the upstream market. Such limits will improve competitive balance amongst clubs, creating less predictable results, which means a better product is being sold to spectators, advertisers, and media in the downstream market. Thus it seems a hard cap would create an improved

---

161. The inquiry under 81(3) is slightly narrower in focus than our inquiry under 81(1). Under 81(1), we asked whether a hard cap or luxury tax was necessary for the existence of European football. Under 81(3), we first ask how a hard cap or luxury tax will improve European football, and then ask whether they are necessary to create such efficiencies.

162. This condition is actually moot for purposes of our inquiry since it is principally aimed at protecting competitors that are not a party to the agreement in question from being squeezed out of the market. However, in the case of a UEFA hard cap or luxury tax, all professional European football clubs would be a party to the agreement (all professional European football clubs are organized under UEFA) and, therefore, for purposes of this fourth condition, there is no competition to be eliminated. Furthermore, a hard cap and a luxury tax seek to enhance competition between clubs, and do nothing to prevent new clubs from coming into existence.
product\textsuperscript{163} that is being passed on to the consumers, thereby satisfying the first two conditions. However, it is not clear that a hard cap would satisfy the third condition under 81(3). The third condition requires that a hard cap be the least restrictive means of creating the benefit of improved competitive balance and less predictable results. The luxury tax is clearly a less restrictive measure, and one that the EC might consider.\textsuperscript{164} While UEFA will not have the burden of proving that a hard cap is necessary to the existence of European football, as it did under 81(1), defending a hard cap under a least restrictive means test is not an easy task. Essentially, UEFA will have to convince the EC that a hard cap instills competitive balance to an extent that less restrictive policies cannot match.\textsuperscript{165} Additionally, UEFA will have to argue that the benefit of the extra parity that only a hard cap can yield outweighs, and thus justifies, the hard cap’s greater restrictive effect on off-the-field competition. The EC would likely reject such a defense of the hard cap. The EC and ECJ have recognized that a degree of equality among clubs is a legitimate aim for sports. This does not mean that the almost perfect equality created by a hard cap is necessary or even desired. Parity must be imposed so that results are no longer predictable, and nothing more. Thus, the EC would not value the extreme balance created by a hard cap, and would likely find that it is not the least restrictive means of instilling competitive balance in European football. As such, a hard cap would not qualify for an exemption under Article 81(3) and thus would be void under Article 81(2).

In analyzing the luxury tax, the first two conditions under 81(3)—that a benefit be created and passed on to customers—represent a bigger obstacle than the third condition. The luxury tax does not eliminate the financial advantage that large-market clubs have in the upstream market, the way the hard cap does. However, it does penalize clubs for making use of that advantage and thus they may be less likely to do so. While the luxury tax will never foster equality to the extent the hard cap does, it would still likely improve competitive balance and, in turn, add to the unpredictability of results by allowing small-market clubs a better chance to contract top players. These effects amount to an improved product on

\textsuperscript{163} Improved goods and services are generally considered a legitimate form of efficiency gain under 81(3). See European Commission, Guidelines on the Application of Article 81(3) of the Treaty, \textit{supra} note 131, at 107.

\textsuperscript{164} See \textit{supra} note 159, ¶ 2.

\textsuperscript{165} While UEFA could alternatively argue that all policies that are less restrictive than a hard cap will fail to instill any competitive balance, evidence, at least in the case of MLB’s luxury tax, does not entirely support such an argument. \textit{See} Neil deMause, \textit{supra} note 159.
the field. As such, the EC will recognize that the luxury tax does create a benefit, and therefore satisfies the first condition. Since this improved product is sold directly to customers in the downstream market, the luxury tax would also satisfy the second condition.

The third condition of Article 81(3) requires that the luxury tax be the least restrictive means of improving competitive balance. Though the tax might deter some large-market clubs from spending in excess of the threshold, clubs can still determine for themselves how much to spend on players. Few balancing mechanisms are as deferential to natural market forces, and as a result, the luxury tax is likely to pass a least restrictive means test. Thus the luxury tax would qualify for an exemption under Article 81(3) in the event that the EC applies Article 81(1).

**C. Final Considerations**

On balance, the luxury tax seems to have a better chance of surviving Article 81 scrutiny; whether it be through a determination that 81(1) does not apply, or through an 81(3) exemption. The EC’s potential argument that the hard cap is not the least restrictive means of creating competitive balance and less predictable results appears to be a strong one. On the other hand, the argument that the luxury tax would not be effective at imposing competitive balance and adding to the unpredictability of results is weak in light of its effectiveness at balancing the market for players in MLB. Furthermore, in light of the EC’s distaste for restrictive practices, the broad and diverse range of markets represented by UEFA, and the powerful interests of both the players’ union FIFPro and influential large-market clubs (such as the G-14) that wish to sustain their popularity, the luxury tax seems to be a more moderate policy that better accommodates the varying agendas of the loose association that is European football.

**VI. CONCLUSION**

In analyzing European football’s problem of a lack of competitive balance, and searching for a solution, a few significant conclusions emerge. First, with few exceptions, the EC and the ECJ are going to subject European football, and all sports, to the same standards as other industries in Europe. This is evident in the *Bosman* decision and throughout the EC amendments to the transfer system. Compliance with European law can sometimes conflict with the interests of European football. In such conflicts, the law often wins out, and to some extent, this phenome-

---

166. See *supra* note 159.
non has contributed to European football’s current problem of poor competitive balance.

Second, irrespective of its causes, it is clear that European football is lacking competitive balance. Over the last few seasons in national competitions across Europe, large-market clubs have dominated consistently, and by such a substantial margin, that supporters of small-market clubs have no hope that their clubs will realistically compete for their respective league championship. This problem has affected ticket sales, and has caused certain people within the sport to call for action to remedy the situation.

This Note suggests that a luxury tax is a viable solution that comports with European competition law better than a hard salary cap. That said, competition law is not the only obstacle to instituting a luxury tax in European football. Other practical considerations not explored here remain relevant. Chief among such considerations are the interests of large-market club owners, and the players themselves, who might prefer a luxury tax to a more restrictive measure such as a hard cap, but who still would likely oppose any policy that is contrary to their interests.168

Finally, whether a luxury tax, a hard cap, or some other option not explored here, it is clear that European football must take steps to improve competitive balance, and increase the unpredictability of results. Arguably, professional football in Europe is played at a higher skill level than any other competition in the sport. It is unquestionable that Europe is the biggest market for the biggest game in the world. Thus, for the sake of the sport, something must be done to make European football matches matter again.

Thomas M. Schiera**

---

168. It is likely that some large-market owners would see a luxury tax as inhibiting their ability to pay hefty sums for the top talent that makes their clubs successful and in turn popular. In the case of players, and the players union, FIFPro, it is easy to see how the luxury tax’s restrictive effect on player salaries would be contrary to their interest.

** B.A. Boston College (2002); J.D. Brooklyn Law School (expected 2007). I would like to thank Professors Jean J. Davis, for her resourcefulness, and Maryellen Fullerton, for her input in the writing process. This Note would not have been possible without the love, support, and understanding of my wife, Cecilia, and my Mother and Father—I love you all.