INSTITUTIONS, EMOTIONS, AND LAW: A GOLDILOCKS PROBLEM FOR MECHANISM DESIGN[‡]

Oliver R. Goodenough[*†]

INTRODUCTION

This Essay lays out a framework for thinking about institutions, emotions, and law. It combines two threads, one drawn from work that links our emotions with internal, psychological commitments, and the other drawn from an understanding of institutions as mechanisms that reframe the strategic landscape in a world of potential cooperation and conflict between social actors. The combination of these two ideas can provide us with a positive theory of moral sentiments and with a clearer way to understand the role of emotion in law. This approach also has implications for the old is/ought distinction—on the one hand undercutting it as a necessary distinction in the world of external "reality," and on the other providing an explanation for its subjective experience as a cognitive imperative.

The internal institutions that can be created through our moral sentiments are hugely important to human social life, but they are also limited. Our intuitive tool kit for structuring cooperative institutions, and for reacting punitively to transgressions against these institutions, provides a good, but still relatively limited, set of reframing solutions. Institutions, however, can originate in a number of different ways and can be located in a number of different media—psychology, culture, even the physical world. Law can be seen as an institution built across many of these layers, a composite that can create a more nuanced and capable set of structures and responses, leading to a better set of outcomes than would be available in an emotion-only world. The problem for the composite is getting the mix right: both too much emotion and too little can prevent us from fully capturing the expanded solution space that a rule of law can open to its adherents.

Law can be too cool; it can also be too hot. Like Goldilocks, the designers of legal institutions are left to search for a mix that will be just right.

---

‡ At the author's request, and in accordance with the interdisciplinary nature of this Essay, the references are presented in scientific format with a bibliography at the end of the Essay. In these references, the author has not cast as wide a net as could be possible, but rather makes references to works, including an overrepresentation of his own, which can provide a starting point for further research.

* Professor of Law, Vermont Law School; Faculty Fellow, Berkman Center for Internet and Society, Harvard Law School; J.D. 1978, University of Pennsylvania; B.A. 1975, Harvard University.

† The ideas set out in this Essay owe debts to many sources, including, in particular, Carl Bergstrom, Susan Bandes, Robert Frank, Monika Cheney, Randolph Nesse, and Paul Zak. I am grateful for the financial support of the Gruter Institute for Law and Behavioral Research, the John Templeton Foundation, and the Vermont Law School that helped make this work possible.

I. WHAT DO INSTITUTIONS DO?

The potential rewards of cooperative sociality are huge. Since the time of Adam Smith, economic thinking has recognized that factors such as specialization, scale, and trade are crucial to increased welfare, broadly conceived, and that such factors require an ordered social structure within which relatively dependable cooperative activity can take place (Smith 1776). Game theory, however, tells us that strategic interactions between actors with potentially divergent goals do not always lead to a plus-sum, cooperative outcome. Some structures are likely to produce such positive interactions; others are not (Goodenough 2008, generally Bowles 2004).

One way of looking at institutions is to view them as mechanisms that can redenominate the strategic structure so as to make different, and on the whole preferable, outcomes more likely than would be the case in their absence (Goodenough 2008, Goodenough & Cheney 2008). A property institution, for example, when reasonably respected, removes many of the barriers that would otherwise exist to straightforward barter trade. It also permits investment, the accumulation of capital, conservation, and a number of other beneficial outcomes (Goodenough & Decker 2008).

We usually think of institutions as the product of conscious design. Indeed, the game-theory subdiscipline of mechanism design, which can be viewed as a kind of workshop for institutions, generally supposes the existence of an active, intentional designer (Parkes 2001). While such may often be the case, evolutionary processes can also lead to the emergence of institutions and to their instantiation in a variety of forms and locations (Goodenough 2008). For instance, a protection against theft may be embodied in a socio-cultural artifact, such as a law; it may be embodied in a mechanism that prevents theft while allowing a trade, such as an armored soda machine in a college dorm or other rough neighborhood; it may also be embodied in a set of internalized psychological mechanisms of expectation and inhibition. It is in such an internal context that emotions play a key role in creating and maintaining productive human institutions.

There are a number of ways in which you can restructure the moves and payoffs of a potentially plus-sum strategic interaction so as to make mutually beneficial play the likely outcome (Benkler 2009, Goodenough 2008, Nowak 2006). Among those recognized early on in biology were reciprocity (particularly in an open-ended, repeat-game context), signaling, and commitment.

Commitment involves some kind of guaranty of a promised performance; classic examples from human life include such “sunk cost” approaches as bank architecture, letters of credit, pawn shops, and

engagement rings. Commitment can also be achieved by taking steps to make reconsidering the promised strategy impossible, even if there were good reasons to do so as the game progresses. Bringing an army across a river to make an attack and then burning the bridges behind it is a proverbial example from human warfare that takes retreat off the board as an option (Adams 2001).

There is a kind of prospective irrationality in a commitment move, an irrationality that is at the heart of its power. The inability to reconsider the committed step, either because of the step's inherent irrevocability or because of some penalty or loss attached to defection, is what makes a commitment useful. If a rational retraction, in light of changed circumstances, is an easy possibility, then the step loses its power to convince.

A strategic actor, wishing to construct a mechanism of cooperation that will give assurance of reliability to other players, will often seek to make commitments that can be both reliable and recognized as such. The kinds of external commitments discussed above can meet these requirements. The critical question for this discussion is whether such a commitment can be made *internally*, as some kind of neurologically grounded bridge-burning that both constrains behavior and is recognizable to other observers. A number of authors have argued that it can (see Nesse 2001). Indeed, some, including Frank (2008) and Hirshleifer (1984), have pointed to that cognitive state we recognize as emotion as being able to anchor exactly this form of internal, subjective commitment.

II. WHAT DO EMOTIONS DO?

In a limited essay such as this, it would be folly to attempt to adequately explore what emotions *are*. For instance, there is the idea of emotion as a state of physical and mental arousal that produces a particular set of cognitive and behavioral consequences. There is also the idea of emotion as the perception of that arousal, both in ourselves and in others, on the screen of our consciousness (see generally Dolan 2002, Goodenough & Prehn 2004, Damasio 1994). For the purposes of this argument, I will avoid such nuance, and simply view emotions in the everyday folk-psychology sense as a bundle of all of the above.

So let us ask instead, what do emotions *do*? That is the critical question here. Of course, emotions do many things, from promoting memory and attention to giving a powerful goad to action (Goodenough & Prehn 2004). In the context of institutions, their critical property is the ability to act as a commitment device. Emotions allow humans to construct credible and recognizable internal institutions, rooted in our internal mental states (Frank

1998, 2001, 2006, 2007 & 2008; Goodenough 2008). The physical facts of neurological arousal have the power to commit us to courses of action in ways that are hard to retract and hard to fake. This capacity greatly expands the range of possible plus-sum interactions in which we can engage with each other and the contexts in which we can establish them. Hirshleifer (1984), for instance, points to the ability of emotion to anchor threats and promises, both critical elements in building and maintaining reliable institutions.

How do emotions accomplish this? At this juncture I can only offer a probable explanation. A good starting point is the widely observed presence of an emotion-like arousal in animals, often linked to threat displays (Adams 2001). Like the burnt bridges of human conflict, a potential aggressor can often be deterred by a credible threat given by the defender that the fight will be a serious one, no matter the cost. This strategic truth, coupled with the zoological evidence, strongly suggests the existence of an evolutionary pathway for psychological and behavioral commitment. The story must include some kind of hard-to-fake and hard-to-retract biological factor, perhaps building emotion's role as an effective signal on the underlying physiology of the fight-or-flight preparation provided by a jolt of adrenaline and other neurochemicals on the recognition of a threat. Indeed, Ward Goodenough (1997) has suggested a link between animal territoriality—where a credible commitment to defense is a key element—and human moral outrage. And once the signal gets going, then the normal arms races of reliability, recognizability, and susceptibility to cheap faking get underway.

What is special about humans in this context is the recruitment of that set of strong reactions we call emotion into a more complex and generalized toolkit of commitment strategies. This complexity can, in turn, support a variety of deeply held values and moral sentiments, ranging from trustworthiness and promise-keeping to a taste and a respect for both fairness and punishment (Goodenough 2008).

This approach can shed light on other concerns of moral theory, most notably the separation of emotion and reason as normative frameworks (Frank 2008). It is exactly because a commitment may be costly that it is effective in inducing the other party or parties to come into the game; and it is exactly when the commitment is going to be costly that we are most tempted by "rationality" to chuck the whole thing over the side and head for the exits, hopefully absconding with as much of the potential mutual benefit us as we can carry.

In the anticipation of such temptation, how is the reliability of an emotion-based commitment to be maintained? To be credible, subjective

values and internal morality need to be armored against the blandishments of rationality, much as a Coke machine needs to be armored against a thirsty undergraduate. In this light, the often-noted opacity of moral sentiments to the processes of reason (Hume 1740, Kelsen 1992) can be seen as a kind of mental firewall helping to preserve the effectiveness of the passions as guarantors of behavior. This firewall manifests itself in explicit discussions of normative philosophy through assertions of the "naturalistic fallacy." The is/ought distinction is not so much a fact of the external world as it is an extremely useful separation in our modes of thought, a separation that opens up a set of strategic options that would be unavailable to a purely "rational" creature.

As this brief discussion suggests, approaching moral cognition from the starting point of mechanism and institutional design provides a useful new framework for investigating old questions and for developing a program to investigate the instantiation of such mechanisms in the physical basis of thought.

III. LIMITS OF OUR EMOTION-BASED INSTITUTIONS

But how effective are these mechanisms? How complete? The human intuitive toolkit is remarkable compared with most other species, but it is not effective enough or comprehensive enough to do the whole job of mechanism building. Our emotionally rooted moral systems, in their application both to ourselves and to other parties, are susceptible to gaps and failures. Perhaps somewhat surprisingly, humans often run *too* hot. In the realm of punishment, for instance, a recent experiment by Dreber, Rand et al. (2008) suggests that in a two-player context, the use of direct punishment against defection is less likely to prompt a return to cooperation than it is to provoke a retaliatory punishment. Such a spiral of escalating punishment is all too familiar, whether in experiments in the lab, in literary contexts such as the Capulets and Montagues of *Romeo and Juliet*, or in real-life conflicts in hot spots around the world.

Some of this failure is to be expected. While we may not know very much for certain about the conditions of the so-called environment of evolutionary adaptiveness (EEA), we can be reasonably assured that the complexity and scope of the human social milieu has expanded from the world in which many of our social emotions evolved (Jones 2001, Jones & Goldsmith 2004). As with many areas of human existence, we can learn better ways of doing things, and the medium of culture allows us to preserve and transmit our improvements. Humans have no natural equipment for flying, but we can engineer ourselves into the air just the same.

In the realm of institutions and mechanism design, we have supplemented our moral sense with both formal and informal cultural structures. We have norms; we have customs; we have religions; we have laws.

IV. WHAT DOES LAW DO?

We can usefully consider law as part of this continuum of structuring mechanisms. The development of a legal institutional framework in many societies around the world is one of humanity's great collective achievements, and also, even where most advanced, a flawed work in progress.

Law works through a mixture of institutional mechanisms. It has a composite structure, embedded in a socio-cultural fabric (Goodenough & Prehn 2004). Law places at least portions of its institutions in a different locus from personal morality, and it can craft solutions to social dilemmas and open up productive interactive possibilities that would be unavailable to our moral intuitions alone (Zak & Knack 2001). For instance, the law of intellectual property (IP) is clearly formulated. Its official doctrine has been well articulated, in both national laws and international treaties. We can argue over some regional and cultural variations, such as the split between Anglo-American and Continental European approaches to the moral rights of authors, but the variations are, for the most part, perfectly well expressed. Many believe it to be a positive element in society, at least when used in moderation. (Defining the proper boundaries of IP is not a policy debate I want to engage in here—rather, see, e.g., Goodneough & Decker 2009 and Fisher 2004.)

IP is also an application of property-flavored institutions that seems not to be a "natural" part of our moral sentiments (Goodenough & Decker 2009). We may respect, even revere, creativity; we may keep knowledge private as a secret; but we are also happy to appropriate ideas, expression, and other products of the intellect, and put them to our own use, with little if any qualm . There appears to be little in our intuitive, emotion-grounded sense of morality that would help us to capture the gains of an IP system. We are affectively cold on the issue. Without the intervention of a legal rule for IP, at least some areas of our technical and creative achievement would have progressed more slowly and with less accomplishment.

Law can also help moderate overly hot reactions. Established legal process can short circuit the kind of escalating tit-for-tat punishment contests that seem all too likely to break out when self-help is the only option. Moving the locus of punishment for a defection from the person with the grievance to a more generalized societal response appears to

increase the inclination of the person punished to "take his medicine," damping down the inclination to retaliate and making punishment-based enforcement a much more useful technique. Part of the legal composite's job is to manage our moral emotions. Even the law's delay (within reason) can play a part, as a kind of circuit breaker against emotion-driven responses (Goodenough & Prehn 2004). We can reframe the old proverb on marriage to a form applicable to retribution: "Punish in haste, repent at leisure." The law as we know it in the United States does a particularly good job of insuring against the problems of an excessively impulsive response.

V. THE LIMITS OF REASON-BASED LAW

This ability to move rules into areas that are free of emotion-based moral content (or where they even run counter to our intuitions) is not just an advantage for the law—it is also one of its greatest challenges. The IP story is not finished yet. Until recently, copying and other appropriative uses of the products of the intellect were often difficult and depended on technology requiring relatively high-cost investments. In such a context, passionless policing by external monitoring was relatively successful in promoting compliance. When unauthorized copying and use became technologically and financially trivial (e.g., the iPod and the photocopy machine), suddenly self-policing became a necessary element for a successful IP regime. Without an anchor in emotion-based commitments to the respect of property, such self-policing is all too likely to be missing. Attempts have been made to build the "don't take me" institution into the physical fabric of the world, through anticopying software on DVDs and CDs, but these have proved only partially successful (Fisher 2004). If some more compelling emotional basis for compliance cannot be found, aspects of IP law may simply need to be rethought (Goodenough & Decker 2009).

One tactic, well exploited in the most effective systems of justice, is to make *adherence* to law the matter of subjective commitment, detached from the specific content of any particular rule. Such deep-rooted and emotionally committed respect can allow law to incorporate at least some passionless, rational processes and still have emotional valance for action and inhibition. But such respect can also allow the law to incorporate brutal, unfair, and even outright wicked rules. Our intuitive checks and balances have a place in the law as well.

VI. LAW AND EMOTION: A GOLDILOCKS PROBLEM

So what is the proper relation of emotion and reason in building the institutions that underpin a successful social order? The two need to have a dynamic partnership. Although some have viewed law and emotion as either mutually irrelevant or perhaps even antagonistic, this is to undervalue the contribution and challenge that emotion brings to the law (see the strands of discussion in Bandes 2001). A competent system of law will both incorporate and moderate emotion in some places, piggybacking onto the intuitive system of value-based commitments to "good" behavior and norm enforcement. As the Latin poet Horace famously stated: "Quid leges sine moribus vanae proficiunt?" This can be translated as "What good are laws when there are no morals?" (Horace Book 3, Ode 24). In other contexts, the legal system will redirect or even suppress emotion. It is not a one-size-fits-all relationship.

As with many dynamic systems in this complicated world, the law-and-emotion challenge is getting the balance reasonably correct for the particular context. So our job, as designers of the composite of mechanisms that make up the law, suggests a combination of the film *Groundhog Day* with the Goldilocks fable. We wake up every day to the same task: looking for the right mixture of reason and emotion to create legal institutions that aren't too hot, and aren't too cold, but are as close as we can get to just right.

BIBLIOGRAPHY

Eldridge S. Adams, *Threat Displays in Animal Communication: Handicaps, Reputations, and Commitments*, *in* EVOLUTION AND THE CAPACITY FOR COMMITMENT 99 (Randolph M. Nesse ed., 2001).

Yochai Benkler, *Law, Policy, and Cooperation*, *in* GOVERNMENT AND MARKETS: TOWARD A NEW THEORY OF REGULATION (Edward J. Balleisen and David Moss eds., forthcoming 2009).

SAMUEL BOWLES, MICROECONOMICS: BEHAVIOR, INSTITUTIONS, AND EVOLUTION (2004).

ANTONIO DAMASIO, DESCARTES' ERROR: EMOTION, REASON, AND THE HUMAN BRAIN (1994).

R.J. Dolan, *Emotion, Cognition, and Behavior*, 298 SCIENCE 1191 (2002).

Anna Dreber et al., *Winners Don't Punish*, 452 NATURE 348 (2008).

WILLIAM W. FISHER III, PROMISES TO KEEP: TECHNOLOGY, LAW, AND THE FUTURE OF ENTERTAINMENT (2004).

ROBERT H. FRANK, PASSIONS WITHIN REASON: THE STRATEGIC ROLE OF THE EMOTIONS (1988).

Robert H. Frank, *On the Evolution of Moral Sentiments*, *in* FOUNDATIONS OF EVOLUTIONARY PSYCHOLOGY 371 (Charles Crawford & Dennis Krebs eds., 2008).

Robert H. Frank, *The Status of Moral Emotions in Consequentialist Moral Reasoning*, *in* MORAL MARKETS: THE CRITICAL ROLE OF VALUES IN THE ECONOMY 42 (Paul J. Zak ed., 2008).

Robert H. Frank, *Cooperation Through Moral Commitment*, *in* EMPATHY AND FAIRNESS 197 (Greg Bock & Jamie Goode eds., 2006).

Robert H. Frank, *Cooperation Through Emotional Commitment*, *in* EVOLUTION AND THE CAPACITY FOR COMMITMENT 57 (Randolph M. Nesse ed., 2001).

Oliver R. Goodenough, *Values, Mechanism Design, and Fairness*, *in* MORAL MARKETS: THE CRITICAL ROLE OF VALUES IN THE ECONOMY 228 (Paul J. Zak ed., 2008).

Oliver R. Goodenough, *Law and the Biology of Commitment*, *in* EVOLUTION AND THE CAPACITY FOR COMMITMENT 262 (Randolph M. Nesse ed., 2001).

Oliver R. Goodenough & Monika G. Cheney, *Preface: Is Free Enterprise Values in Action?*, *in* MORAL MARKETS: THE CRITICAL ROLE OF VALUES IN THE ECONOMY, at xxiii (Paul J. Zak ed., 2008).

Oliver R. Goodenough & Gregory Decker, *Why Do Good People Steal Intellectual Property?*, *in* LAW, MIND AND BRAIN 345 (Michael Freeman & Oliver R. Goodenough eds., 2009).

Oliver R. Goodenough & Kristin Prehn, *A Neuroscientific Approach to Normative Judgment in Law and Justice*, 359 PHIL. TRANSACTIONS OF THE

ROYAL SOC'Y OF LONDON, SERIES B, 1709 (2004), *reprinted in* LAW AND THE BRAIN 77 (Semir Zeki & Oliver R. Goodenough eds., 2006).
Ward H. Goodenough, *Moral Outrage: Territoriality in Human Guise*, ZYGON, March 1997, at 5.

Jack Hirschleifer, *On the Emotions as Guarantors of Threats and Promises* (UCLA Dep't of Econ., Working Paper No. 337, Aug. 1984), *available at* http://econpapers.repec.org/paper/clauclawp/337.htm.

DAVID HUME, A TREATISE OF HUMAN NATURE (L.A. Selby-Bigge ed., Oxford Univ. Press 1978) (1740), *available at* http://www.gutenberg.org/etext/4705.

Owen D. Jones, *Time-Shifted Rationality and the Law of Law's Leverage: Behavioral Economics Meets Behavioral Biology*, 95 NW. U. L. REV. 1141 (2001).

Owen D. Jones & Timothy H. Goldsmith, *Law and Behavioral Biology*, 105 COLUM. L. REV. 405 (2005).

HANS KELSEN, PURE THEORY OF LAW (Max Knight trans. & ed., 1967).

Randolph M. Nesse, *The Future of Commitment*, *in* EVOLUTION AND THE CAPACITY FOR COMMITMENT 310 (Randolph M. Nesse ed., 2001).

Martin A. Nowak, *Five Rules for the Evolution of Cooperation*, 314 SCIENCE 1560 (2006).

David C. Parkes, Iterative Combinatorial Auctions: Achieving Economic and Computational Efficiency (May 2001) (unpublished Ph.D. dissertation, University of Pennsylvania), *available at* http://www.eecs.harvard.edu/~parkes/diss.html.

THE PASSIONS OF LAW (Susan A. Bandes ed., 1999).

(R.H. Campbell et al. eds., Oxford Univ. Press 1976) (1776), *available at* http://www.econlib.org/library/Smith/smWN.html.

Paul J. Zak & Stephen Knack, *Trust and Growth*, 111 ECON. J. 295 (2001).