Strategic Mechanisms, Functional Modeling and Experimental Design in Neurolaw

Oliver R. Goodenough
Professor of Law, Vermont Law School
Faculty Fellow, Berkman Center for Internet and Society, Harvard University

August 3, 2009

Introduction

This paper has four goals. The first is to provide an overview of the emerging sub-discipline of "Neurolaw." The advances of cognitive neuroscience are challenging many disciplines. Incorporating the insights of neuroscience into legal analysis and policy-setting is a rapidly expanding enterprise. It is not just a one-way street, however, with law borrowing second-hand conclusions and ideas from a sister science. Rather, Neurolaw is developing its own methods and its own tradition of empiricism.

The second goal is to focus on game theory and mechanism design. These provide useful analytic starting points for applying cognitive neuroscience in a social context. This leads to the third goal. I will suggest that the formal structures of these mechanisms of sociality can be expected to be represented in the structures of the cognitive processes which implement them. These cognitive processes are central to, but also draw on, the brain/culture interactions that produce human social behavior. I do not mean to suggest that there is a simple and direct homology between such mechanisms and any particular physical structures within the brain. It is widely recognized, however, that the brain is a computational device, and the brain processes which carry out a particular type of computation will necessarily reflect the requirements of the computation being made.

Finally, the paper will suggest an application of this mechanism-based approach to a particular instance: understanding the nature of human moral commitment. This suggestion is still very much a work in progress, however. The program of empirical

study that can help to confirm or confound this approach can be envisioned, but it is not yet carried out.

The Emergence and Targets of "Neurolaw"

Many disciplines have a model of human thought and action at their heart. Some, such as psychology, are explicit sciences of the mind. Others, such as economics and law, have, at their core, models of our decision making processes that determine how they approach the world. In the past few decades, cognitive neuroscience has brought profound new understandings about the workings of the brain. It has also invented nearly magical methods for investigating the physical basis of mental processes. These advances have, in turn, begun to revolutionize discipline after discipline. Sometimes the new knowledge overturns old certainties; in other cases it confirms and clarifies our existing understandings. What is clear, however, is that cognitive neuroscience is an intellectual force that cannot be ignored.

The role of neuroscience in the study and the application of law is gathering momentum. Law is deeply interested in how people think and behave. One reason for the impressive impact of economic ideas on law in the last half-century has been the perceived success of economics in explaining human motivation and decision making (e.g. Goodenough 2006). But economics itself is now turning increasingly to neuroscience for methods and explanations (e.g. Schultz 2008, Zak 2004), and the study of law is following suit. The trickle of "neurolaw" scholarship that started about a decade ago has become a steady stream, gaining increasing acceptance within the academy and support from such forward looking institutions as the MacArthur Foundation (Law and Neuroscience Program 2009).

Much of the Neurolaw scholarship has involved a traditional legal academic taking the conceptual advances of cognitive science and applying them to law and legally-related questions. Law is an inveterate intellectual scavenger, and it has a long history of looking to the best in contemporary science both for inspiration and, where

possible, as a source of applications (Elliott 1985). But legal scholars are not simply consumers at the Neurolaw table: a growing number of lawyers and even judges are directly involved in the design and execution of experiments and other primary research combining law and neuroscience (e.g. Buckholtz et al. 2008, Prehn et al. 2008). As this trend develops, we are beginning to see explicit thinking about the methods which such research might employ. This essay will aim to address one of these methodological questions: the role which mechanism design principles can play in helping to understand the functional structures likely to underlie solutions to the problems of sociality.

The Targets of Neurolaw Investigation

Before getting to the specific possibilities of mechanism design, it will be helpful to take a broad view of the current subject-matter targets of neuro-legal investigation. These can be helpfully sorted into a few broad categories. For many, law means a courtroom, and Neurolaw means brain science techniques with direct courtroom application. This category includes attempts to develop objective methods for assessing such traditionally subjective questions as pain (Kolber 2007, Miller 2009), truth-telling (Garland & Glimcher 2006, Merikangas 2008, Langleben 2008, Sinnott-Armstrong 2009) and memory (Phelps & Sharot 2008, British Psychological Society 2008). Notwithstanding the interest sparked by these efforts at "mind reading," the development of courtroom-ready applications has been slow, although greater success is always a possibility as the science improves. Furthermore, important questions of neural privacy are only beginning to be addressed (e.g. Kolber 2007).

Getting neuroscientific results into the courtroom in a specific case raises a number of issues for the law of evidence. The science itself must meet tests of reliability and acceptance, and its application to the legally relevant question must also be established (e.g. Garland & Glimcher 2006). Much of the work in neuroscience involves cumulating the results of multi-subject studies, which don't always translate to reliable conclusions about a particular subject. Undue influence on judges and juries is also a potential problem. Studies have shown that people are too easily persuaded by arguments

– even faulty ones – that include brain-scanning images (McCabe & Castel 2008). Judges, as the gate-keepers of admissibility, will need some convincing before brain-scans and other neuroscientific data come into the courtroom in any volume.

Then there are questions of criminal law, some relating to specific cases, and others to matters of legal policy more generally, such as responsibility (Greene & Cohen 2004, Sapolsky 2004, Goodenough 2004, Roskies 2006, Morse 2006a, 2006b and 2008), addiction (e.g. Bonnie 2005, Morse 2006c, Erickson 2007), and mental health (e.g. Sapolsky, 2004, Pustilnik 2006). The work on responsibility, a question that arises in the guilt or innocence phase of a U.S. criminal trial, remains controversial, at least in its application to adults. The traditional folk-psychological approaches seem to fit with our desires for punishment (Morse 2008), and the law may better reflect the psychology of punishers as apposed to those being punished (Goodenough 2004).

There is much broader acceptance for suggestions that we need a more nuanced approach to the consequences we inflict on those adjudged criminal. Both courts and legislatures have taken some initiative at this level. The recent development of such tailored responses as drug courts (Nolan 2003, but see O'Hear 2009) and mental health courts (Erickson et al. 2006, Council of State Governments Justice Center 2008) show a desire to get away from incarceration – and long incarceration at that – as the principal weapon in the anti-crime arsenal. These developments had already been underway before Neurolaw became incorporated into the mix, but neuroscience has helped to increase their effectiveness and accelerate their spread.

An additional category, and the one that will be a particular focus in this paper, involves normative reasoning. How do we generate and apply normative rules – questions of ought – whether derived from abstract ethics, moral sentiments, or legal formulations? These include rules and judgments that we apply to ourselves and those we apply to others, along with those made in formal, explicit contexts, such as by a judge or jury in a court of law, and those made viscerally and in the heat of the moment. Although these processes are tied together by the concept of some form of behavioral

obligation, there is a long history of disputes in philosophy, jurisprudence, and psychology about the function, merits and interrelations of these different normative systems and domains (e.g. Prinz, 2007, Greene 2008). It is in just such an area of tangled explanations and discordant models about the nature of thought that cognitive neuroscience has the potential to make a major contribution, not only to our academic understanding, but to the shaping of our institutions of law as well

The Methods and Models of Neuroscience with Application to Neurolaw

Neuroscience itself is a discipline with a number of subcomponents and sub-disciplines. At its heart, neuroscience combines three key elements of interest to neurolaw: (i) technical and methodological advances, (ii) richer cognitive models, and (iii) an understanding of the strategic dimensions of thought and action, particularly in the human social context (Goodenough 2006).

When people consider neuroscience, the technology often gets the most attention. Scanning technologies in particular, with their nearly magical ability to probe and represent the anatomy and functioning of a living brain, cause awe and excitement – sometimes beyond their real ability to deliver (Roskies 2007, Sinnott-Armstrong et al. 2008). In some aspects of Neurolaw, such as the presentation of evidence of pathology in a criminal defense or prosecution or in the investigation of subjective states like truth-telling, these technological advances are front and center. Indeed, for some, the application of technology directly to finding answers to the questions raised in a particular case is the main occupation of Neurolaw.

It is worth remembering, however, that the term "neuroscience" is often combined with the term "cognitive" (Gazzaniga et al. 2002). It is not just a physical science of the brain, but also a physically-grounded science of thought. Cognitive research reveals a complex picture of mental activity. We have moved beyond the unitary model of thought that had dominated much of philosophy and psychology since Descartes (Damasio 1994). Rather, our thought processes are multiple and complex, often the product of different

physical systems activating different parts and combinations of the brain itself. As we study cognition, we can ask "what is the function that the brain activity is aimed at performing?" Different systems will work in ways that are shaped, at least in significant part, but the necessities of the problems they are being asked to address.

By working out the functioning of the physical systems involved in our thinking, we can create better models of how that thought works, and these models in turn help us in our next round of investigation into the physical systems and their function. In this context, the "magical" technology has an important role to play as a research tool in helping to find how our cognition works – but here as a means to an important end and not just the end in itself.

The third important strand is the strategic dimensions in human thought and action, particularly when such thought and action occurs in a human social context (Sanfey 2007, Fehr & Camerer 2007, Lee 2008, Krueger et al. 2008). If the functions of the brain do indeed reflect the computational contours of the problems they are seeking to solve, then an important step in studying both the functions and the problems will be to lay out, as clearly as possible, the structure of the challenges and of the pathways to solutions for those challenges. Lee (2008) summarizes the approach:

> "Many neurobiological studies have exploited game theory to probe the neural basis of decision making and suggested that these features of social decision making might be reflected in the functions of brain areas involved in reward evaluation and reinforcement learning."

For such inter-personal domains as law, morality, and normative thinking more generally, a key task is to understand the strategic dilemmas of sociality – i.e. the discordant and concordant interests of agents where mutually necessary resources are limited or our goals otherwise overlap for good and ill. Game theory is a powerful analytical system for providing insight on these questions, made even more powerful when we add the concepts of mechanisms and mechanism design to the mix. The remainder of this essay will suggest how these approaches can be particularly helpful as

we consider an empirical program of research on many of the questions of Neurolaw, briefly examining, in the process, how we might put this approach to work on a particular subject: moral commitment.

What Do Mechanisms Do?

Many readers will be familiar with the core ideas of game theory, and its efforts to formalize the opportunities and challenges of social interaction (e.g. Gintis 2000). In cases where cooperation around a task would expand the possible returns, it is not always clear that actors can stay on the path to such an outcome. Because of the pay-offs from specialization or the gains of scale and trade, cooperation can often be a road to overall gains for its participants – a "non-zero" opportunity. In such circumstances, however, there may also opportunities for short-term gain by one of the players from an immediate defection at the expense of the others. The classic formulation of such a problem is "the prisoners' dilemma" game, but the basic problem occurs in many other circumstances as well. Fields as diverse as economics and evolutionary biology have focused on these concerns when seeking to model cooperation (Gintis 2000).

Recent developments in game theory suggest new ways of thinking about these problems, developments some put under the label of "mechanism design". The simple move here is to understand that strategic actors are not necessarily prisoners of the prisoners' dilemma, or of any other game structure that they can foresee will not lead to positive interactions. Rather, competent strategic actors will not only understand the nature of the moves in the game on offer, but will also understand the nature of the game, and whether it is to be embraced or avoided (Goodenough 2008). Such actors will make choices on whether to join the play, rightly treating the prisoners' dilemma as quicksand on the strategic landscape, to be avoided when possible (except, perhaps, when the other player is clearly enough of a sucker to make the defection move actually profitable).

Nor will a truly competent actor only choose the better opportunities from those that present themselves. The actor can also work to *design* the strategic landscape,

actively opening up solution pathways for interactions where mutually plus-sum outcomes are the likely result. These structures can be referred to as "mechanisms," and the process of their creation "mechanism design" (Parkes 2001). The ingredients in such mechanisms will include many of the favorites already delineated in game theory, economics, and evolutionary theory, including such perennials as reciprocity, repeat play, punishment, and commitment (Goodenough 2008). The idea of mechanism design helps to give coherence to their deployment as people meet the challenges of long-term cooperation. A respect for property, for instance, is a mechanism that helps to solve a number of such dilemmas, while creating a few new ones along the way (Goodenough & Decker 2009).

These mechanisms can be usefully equated with the idea of institutions, as such term is sometimes used to denote behavior-shaping sets of rules, practices, and interrelations (Goodenough & Cheney 2008). At their best, institutions can create reliable interaction spaces, where such plus-sum activities as trade, work on a common project, and investment can occur. It is not too much of a stretch to think of normative rules, whether embodied in moral sentiments, respected custom, or formal law, as examples of such institutions, and to look for their structures to be grounded in the principles of mechanism design. And it is not too much of a further stretch to conclude that the structural properties of the mechanisms will be present in the institutions of morality and law, and will be represented, at least in part, in the functional outlines of the mental processes which support them. The recurring neuroscience question of "what is the brain doing" has a potential answer in the normative domain: it is creating and working within mechanisms that will help solve the recurring dilemmas of a highly social and reasonably competent animal: *Homo sapiens*.

Where Mechanisms are Located?

The use of mechanisms in thinking about the function of the brain is complicated by the fact that the structures of mechanisms can be located in a number of different media, ranging from external physical artifacts such as locks, architecture or software, to

cultural institutions such as law, and on to the workings and functional structures of the brain itself. Indeed, if the mechanism requires some form of intentionality and choice by the actor, involving the actor's thoughts and behavior, the brain will, by necessity, be at least a part of the loop.

The ubiquitous soda-dispensing machine is a good example of the multiplicity of places in which a successful mechanism locates the tracks and boundaries that turns a potential pit of defection into a widely used avenue for plus-sum commerce (Goodenough 2008). The core transaction is the offer to trade pleasant and refreshing sugar water – the soda – for a token of stored value – money. The principal opportunities for defection are, on the soda company's side, the provision of nothing, or of a poor or adulterated product, and, on the drinker's side, of taking the product and failing to make the required monetary exchange.

The drinker's defection is prevented, in the first instance, by a clever baffle and dispensing mechanism that prevents simply reaching up into the machine to take the soda. To this is added a sophisticated money recognition mechanism and armor that fends off all but the most violent attacks on these devices. Morally supported concepts of property, and inhibitions against theft, are also useful in keeping the drinker honest. In the deeper background are culturally and legally supported rules about the production of money and about theft or the destruction of the machine itself. These legal institutions are themselves, even more deeply still, supported by fear of punishment and by commitments to law-abidingness made in human brains.

The soda company's defection is made unlikely by its necessity to be a trusted repeat supplier of pleasant and refreshing sugar water. It has developed a strong signal in its brand and trademarks – a signal backed up by a massive investment in otherwise unproductive advertising and by the law of trademark that makes it difficult for others to appropriate the signal and to parasitize its worth. Thus the most salient aspect of most soda machines is the illuminated sign for "Coca Cola" or some other highly recognizable sign of origin and dependability. Again, the law sits in the background, both as the

guarantor of the legitimacy of the signal and as a final preventative against a strategy of dangerous adulteration by the soda company.

We see how productive mechanisms/institutions can be set up in a complex mix of systems. The remainder of this essay will examine ways in which mechanisms of law and morality are likely to be anchored in the brain. It will take as its focus for this discussion a question from my own current work: the role of emotion as a guarantor of internal moral commitments.

Using Mechanism Design as a Starting Point for Neurolaw Research

The core hypothesis developed in the paper can now be simply summarized: The solution structures for the challenges of productive sociality will have usually have at least partial instantiation in the functionality of the brain systems working to implement those solutions. If a particular strategic mechanism is at issue, the structural elements necessary for that mechanism to work will be present in the brain/culture/external artifact world continuum, in linked ways that cumulatively make the mechanism possible and functional. By analogy, if we observe things flying, it is reasonable to look for the wing, i.e. for the structure that provides the lift that makes flying work. If we observe actors perceiving the world through light-based images, it is reasonable to look for the structures of focus and sensitivity we call an eye. And if we observe things having certain kinds of positive-outcome interactions with others of their kind, it is reasonable to look for the mechanism that helps to produce such behavior and that creates the pathway to that outcome. *Cherché le mécanisme*.

This hypothesis provides a possible roadmap for investigating brain function in the normative domain. Evolutionary science has famously identified a number of different points and paths for the evolution of the eye (Dennett 1995). While each instance has its own characteristics, each also shares some common strategies. And the search for physical structures in which those strategies are embodied would be a good starting point for investigating the functional anatomy of a previously un-described

animal that has the capacity for seeing. In the same way, I believe that understanding the strategies of sociality is a good starting point for forming hypotheses about the functionality of the brain as it engages in thought and produces action in a normative context. Those hypotheses, in turn, will be tested by the very empirical program which they help to inform, in the classic process of scientific advancement.

Steps along this path have already been taken by researchers into what some call "social neuroeconomics" (Fehr & Camerer 2007, Sanfey 2007). Some of these efforts have explored the brain circuitry of utility and reward (e.g. Kable & Glimcher 2007) while others have focused on how the players keep track of the options and actions of others in the game (e.g. McCabe et al. 2001), and yet others do both (e.g. Seo & Lee 2008). This research often uses animal subjects. For instance, in an experiment which put monkeys into a repeat-play, competitive game, Barraclough et al. (2004) the researchers were able to show how "neurons in the dorsolateral prefrontal cortex (DLPFC) encoded the animal's past decisions and payoffs, as well as the conjunction between the two, providing signals necessary to update the estimates of expected reward." Krueger et al. (2008) provide an excellent review of much of this work, and point to the need for further research,

Often, this research has focused on the first level question of play within a game – but not on the higher level questions of how to chose or design games, or to maintain compliance with rules, both by the actor and by other players. Among those who have examined the mechanisms that maintain cooperation, some of the most productive work has been on trust. As a leader of this research, McCabe and his collaborators have focused on imaging work while looking at the so called "trust game" (e.g. McCabe et al. 2001, Krueger et al. 2008). Zak and others have convincingly established the importance of oxytocin in the neurochemistry of trust (Zak 2004, 2007, 2008, Kosfeld et al. 2005, Zak et al. 2005). Neurochemistry has sometimes taken a back-seat to structure in recent neuroscientific studies, in part because the power of imaging made structure so accessible for study. If one thing is clear about neurochemistry, however, it is that it works in

partnership with structure, and few explanations will be complete if they take only one aspect of the puzzle into account.

The use of game theory in neuroeconomic studies is well established. As Neurolow becomes an empirical science, attention to the nature of the strategic problem to be solved in the brain/law interaction should become a customary factor in the design and analysis of controlled experiments and of "wild" behaviors in the greater laboratory of life. And such approaches will often benefit by moving beyond consideration of behavior *within* a game to focus equally on the questions of the design and maintenance of the mechanisms that *define* the game.

The Role of Emotion in Morality and Law

There is a long history of argument on the role – and importance of emotions in morality and law. There are arguments over what emotions *are* and there are arguments over what emotions *do*.

As to what the emotions *are*, part of the difficulty is that our folk psychology of emotions seems to encompass both (i) a set of underlying neurological and physiological states of arousal and (ii) the awareness of such states that we experience about ourselves through our subjective abilities at self-monitoring and that we experience about others through observation, empathy, and reasoned awareness (Goodenough & Prehn 2004). In some ways, this latter observational aspect is like identifying an explosion that takes place behind some kind of opaque barrier by analyzing the patterns of smoke that emerge after the fact and waft up into our field of vision. We know something has happened, but do most of our speculating about exactly what it was based on secondary phenomena. The cognitive opacity to the emotions has been a particular stumbling block for legal theorists such as Kelsen (1992). All that said, for the purposes of this essay we can take emotion as a package, in the generally accepted way, and get on with the argument

As to what the emotions *do*, they of course do many things, from promoting memory and attention to giving a powerful goad to action (see, e.g., Rolls 1999, Phelps 2002, Dolan 2002). In the context of the *moral* emotions, and the role of such emotions in law, we can gain some insight by following the method suggested in this essay. We might ask, in order: i) "what is the problem that emotion-base moral judgment is aimed at solving," ii) "what is the structure of the mechanisms that could be at work solving that problem" and iii) "how might that structure be manifest in the mental processes and brain functions at work in performing the calculations that support that solution." In looking for answers, the work of Hume (1739/40), Hirshleifer (1984), Frank (1988, 2001, 2006, 2007 & 2008), Greene (2008) and Prinz (2007), gives useful guidance.

Hume identified moral emotions – or sentiments as he generally termed them – are a critical element in moral thinking (Hume 1739/40). He recognized the separation of this kind of thought from the processes which we sum up under the labels "reason," or "rationality," and went so far as to formulate what is often called Hume's law. As restated by Prinz (2007), the assertion is that "there is no way to derive an ought from an is," at least as a matter of language-based logic. Prinz (2007) and Greene (2008) are essentially modernizers of Hume's insights – expressly so in Prinz' case. They too recognize emotion's role in moral judgment, and give accounts that integrate this role into brain function.

Hirshleifer and Frank have taken the argument to the next stage, proposing a mechanism based explanation of the kind adopted here. One step an actor can take to restructure the strategic landscape in order give an assurance against defection from a promised action is to make a *commitment.* In this context, a commitment is a hard to fake and hard to retract change of state or circumstance that makes defection costly or simply prohibitively difficult to accomplish. To be effective in convincing others that the game form has been altered by a commitment, an honest signal of the commitment is a very useful element as well.

As the often used saying attests, an army can commit itself to ferocious fighting and no surrender by coming across a river to make an attack and then burning the bridges behind it. Animals can commit themselves in a number of ways, including to a strategy of fierce defense of territory or a mate (Adams 2001), and it appears that this commitment can be made *internally*, perhaps through some kind of neurochemical equivalent of behavioral bridge burning. The raised hackles, bared teeth, and vicious snarl of a pugnacious dog are all signals of an internal commitment to make the attacker pay dearly for an intrusion.

This kind of arousal is a highly plausible starting point for recruitment in humans into a system of commitment that can support moral judgment. Ward Goodenough (1997) noted a link between animal territoriality – where a credible commitment to defense is a key element – and human moral outrage. Hirshleifer (1984) suggested that emotions can act as guarantors of threats and promises, with the internal cost of guilt. Frank (passim) extended the argument to give the emotions a key role in supporting many of our morally based actions. This applies both to moral positions taken about our behavior, as Hirshleifer suggests, and to our reactions to the transgressions of others, where a commitment to punishment or other sanction is an important element in keeping other players "honest" in a positive-outcome game structure (e.g. Fehr & Fischbacher 2004, Fehr & Gintis 2007, Greene 2008).

This promising line of thinking can be taken a further step – and, using the approach suggested here, we can anticipate how the functions of such an internal commitment mechanism may be organized in the brain/behavior complex. As noted above, the story must include some kind of hard-to-fake and hard-to-retract biological factor, constituting an effective signal as well as an internal commitment. If the aggressive display of a threatened animal is the source, this could build on the underlying physiology of the fight-or-flight preparation provided by a jolt of adrenalin and other neurochemicals on the recognition of danger (Goodenough 2009).

This supposition has a further intriguing possibility: perhaps there is a *functional* reason that our moral sentiments are opaque to the kind of thinking we call reason or rationality. It is exactly because a commitment may be costly that it is effective; and it is exactly when the commitment is going to be costly that we are most tempted by "rationality" – which seems to be particularly good at at "rational actor" model, short-term, cost benefit analysis – to chuck the whole thing over the side and head for the exits. It is plausible to imagine that the subjectively perceived opacity of our moral emotions to reason may be the product of a cognitive "firewall," insulating the utility and power of an ability to make reliable, intuitive commitments against erosion by rationality. This suggests a new approach the Hume's Law itself. The perception of separation in our thinking that Hume raises to principle of logic may in fact be an artifact of the cognitive architecture of subjective commitment (Goodenough 2009).

The link of brain function to the manifestations of thought and behavior is, of course, central to much of the work in neuroscience – and a number of researchers into the basis of moral cognition have suggested mechanisms, of sorts, as underlying normative thought. For instance, Moll et al. (2008) list six combinatory elements that underlie the moral emotions: attachment, aggressiveness, social rank, outcome assessment, agency, and norm violation. Mikhail (2008) cites a computationally complex "moral grammar." This approach sets out a process for moving from perception to conclusion that is potentially complimentary to the more game theoretic approach suggested here. These approaches could often benefit, however, from the incorporation of modeling drawing on the mechanism design principles suggested here.

The Next Steps: Empirical Investigation

Do mental processes involved in the internal commitments which help support human morality and adherence to law have a recognizable commitment-based structure? The obvious next step in this work-in-process is a program of empirical study that can help to confirm or confound this approach. The outlines of such a program can be envisioned, involving the classic mix of behavioral data, functional imaging,

neurochemical sampling, and other physiological measurements. The tasks should involve both commitments about one's own behavior as well as commitments to a particular response to the behavior of others. The detailed design of such a program, along with its funding and execution, are the next steps in this "work in progress."

While developing such a program is an immediate goal for my own research, I also hope that this essay may help to stimulate others to work on the approach as well.

Conclusions

The investigations of Neurolaw have a number of different targets, ranging from courtroom applications to policy considerations. One of these targets is the neural processes involved in normative judgments, both about our own actions and about the actions of others. Game theory, illuminated by the insights of mechanism design, provides a set of expectations about the shape of solutions to strategic problems. These expectations can be a source for modeling the structures that we can look for in the functionality of the brain as it addresses these problems. Taking these structures into account can provide a roadmap for empirical studies about normative thinking, and the empirical studies, in turn, will provide a test for the roadmap and for the underlying method.

References

Adams, Eldridge S. 2001. Threat displays in animal communication: handicaps, reliability, and commitments, in Randolph M. Nesse, ed. *Evolution and the Capacity for Commitment*, New York: Russell Sage Press.

Barraclough, D.J., Conroy, M.L., Lee, D. 2004. Prefrontal cortex and decision making in a mixed-strategy game. *Nature Neuroscience*, 7: 404-410.

Bonnie, Richard J., 2005. Judicially Mandated Treatment with Naltrexone by Opiate-Addicted Criminal Offenders. *Virginia Journal of Social Policy and the Law* 13: 64-88.

Bowles, Samuel 2004. *Microeconomics: Behavior, Institutions, and Evolution*, Princeton, N.J.: Princeton University Press.

Buckholtz, Joshua W., Asplund, Christopher L., Dux, Paul E., Zald, David H., Gore, John C., Jones, Owen D. and Marois, Rene, 2008. The Neural Correlates of Third-Party Punishment, Neuron 60: 930-940.

British Psychological Society, 2008. *Guidelines on Memory and the Law: Recommendations from the Scientific Study of Human Memory*, Liecester: British Psychological Society, available at http://www.forensic-centre.com/files/Memory%20and%20the%20Law.pdf

Council of State Governments Justice Center, 2008. *Mental Health Courts: A Primer for Policymakers and Practitioners*, Washington, D.C.: Bureau of Justice Assistance, available at http://consensusproject.org/mhcp/mhc-primer.pdf

Damasio, Antonio 1994. *Descartes' Error*, New York: Putnam.

Dennett, Daniel C. 1995. *Darwin's Dangerous Idea: Evolution and the Meanings of Life*, New York: Simon & Schuster.

Dolan, R. J., 2002. Emotion, Cognition, and Behavior, *Science*, 298: 1191 – 1194.

Elliott, E. Donald, 1985. The Evolutionary Tradition in Jurisprudence, *Columbia L. Rev.* 85: 38- ___.

Erickson, Carlton K. 2007. *The Science of Addiction: From Neurobiology to Treatment*, New York: W.W. Norton & Co.

Erickson, Steven K., Campbell, Amy, & Lamberti, Steven J., 2006. Variations in Mental Health Courts: Challenges, Opportunities, and a Call for Caution. *Community Mental Health Journal*, 42: 335-344.

Fehr, Ernst, & Camerer, Colin F., 2007. Social neuroeconomics: the neural circuitry of social preferences. *Trends in Cognitive Sciences*, 11: 419-427.

Fehr, Ernst and Fischbacher, Urst, 2004. Third-party punishment and social norms. *Evol. Hum. Behav.*, 25: 63–87.

Fehr, Ernst and Gintis, Herbert, 2007. Human Motivation and Social Cooperation: Experimental and Analytical Foundations. *Annu. Rev. Sociol*.33:43-64.

Fisher, William W., 2004. *Promises to Keep: Technology, Law and the Future of Entertainment*, Stanford, California: Stanford University Press.

Frank, Robert H. 1988. *Passions Within Reason*, New York: W.W. Norton.

--- 2001. Cooperation through Emotional Commitment, in Randolph M. Nesse, ed. *Evolution and the Capacity for Commitment*, New York: Russell Sage Press.

--- 2006. Cooperation Through Moral Commitment, in Chris Frith et al., eds., *Empathy and Fairness (Novartis Symposium 278)*, Chichester: Wiley.

--- 2007. On the Evolution of Moral Sentiments, in Charles Crawford, ed., *Foundations of Evolutionary Psychology*, Mahwah, NJ: Erlbaum.

--- 2008. The Status of Moral Emotions in Consequentialist Moral Reasoning, in Paul Zak, ed., *Moral Markets: The Critical Role of Values in the Economy*, Princeton, NJ: Princeton University Press.

Garland, Brent, & Glimcher, Paul W. 2006. Cognitive neuroscience and the law. *Current Opinion in Neurobiology*, 16:130-134.

Gazzaniga, M. S., Ivry, R. B. & Mangum, G. R. 2002. *Cognitive Neuroscience: the Biology of the Mind*. New York: W.W. Norton.

Gintis, Herbert 2000. *Game Theory Evolving*, Princeton, Princeton University Press.

Goodenough, Oliver R., 2006. Can Cognitive Neuroscience Make Psychology a Foundational Discipline for the Study of Law? in Belinda Brooks-Gordon and Michael Freeman, eds., *Law and Psychology, Current Legal Issues Vol. 9*, Oxford: Oxford University Press.

--- 2004. Responsibility and punishment: whose mind? A response. Phil. Trans. R. Soc. Lond. B, 359: 1805–1809.

--- 2008. Values, Mechanism Design, and Fairness, in Paul Zak, ed., *Moral Markets: The Critical Role of Values in the Economy*, Princeton, NJ: Princeton University Press.

--- 2009. Institutions, Emotions and Law: A Goldilocks Problem for Mechanism Design. *Vermont Law Review*, 33: 395-404

Goodenough, Oliver R. and Cheney, Monika G. 2008. Preface: Is Free Enterprise Values in Action? in Paul Zak, ed., *Moral Markets: The Critical Role of Values in the Economy*, Princeton, NJ: Princeton University Press.

Goodenough & Decker, 2009. Why Do Good People Steal Intellectual Property? In Michael Freeman & Oliver R. Goodenough, eds., *Law, Mind and Brain*, London: Ashgate.

Goodenough, Oliver R. and Prehn, Kristin, 2004. A neuroscientific approach to normative judgment in law and justice. *Philos Trans R Soc Lond B Biol Sci.* 359(1451): 1709–1726.

Goodenough, Ward H. 1997. Moral Outrage: Territoriality in Human Guise. *Zygon*, 32(1): 5-27.

Greene, Joshua D., 2008. The Secret Joke of Kant's Soul, in Walter Sinnott-Armstrong, ed., *Moral Psychology Vol. 3, The Neuroscience of Morality: Emotion, Brain Disorders, and Development*, Cambridge, MA, MIT Press, at 36-79.

Greene, J, and Cohen, J. 2004. For the law, neuroscience changes nothing and everything. *Phil. Trans. R. Soc. Lond.* **B**, 359: 1775-1785.

Hirshleifer, Jack 1984. On the Emotions as Guarantors of Threats and Promises. *UCLA Economics Working Papers*, Number 337, available at http://econpapers.repec.org/paper/clauclawp/337.htm.

Hume, David, 1739/40. *A Treatise of Human Nature*, variously reprinted, available at http://www.gutenberg.org/etext/4705.

Kable, Joseph W. & Glimcher, Paul W., 2007. The neural correlates of subjective value during intertemporal choice. *Nature Neuroscience*. 10: 1625-1633.

Kelsen, Hans, 1992. *Introduction to the Problems of Legal Theory*, a translation of the first edition (1934) of *Reine Rechtslehre*, or Pure theory of the law; translated by Bonnie Litschewski Paulson and Stanley L. Paulson, with an introduction by Stanley L. Paulson, Oxford: Oxford University Press.

Kolber, Adam J., 2007. Pain Detection and the Privacy of Subjective Experience. *American Journal of Law & Medicine*, 33: 433-456.

Kosfeld, Michael, Heinrichs, Markus, Zak, Paul J., Fischbacher, Urs, Fehr, Ernst, 2005. Oxytocin increases trust in humans. *Nature*, 435: 673-676.

Krueger, Frank, Grafman, Jordan, & McCabe, Kevin, 2008. Neural correlates of economic game playing. *Phil. Trans. R. Soc. B*, 363: 3859-3874.

Langleben, Daniel D., 2008. Detection of deception with fMRI: Are we there yet? *Legal and Criminological Psychology* ,13: 1-9.

Law and Neuroscience Program, 2009. See generally the web resources available at www.lawandneurosciencepprogram.org.

Lee, D. 2008. Game theory and the neural basis of social decision making. *Nature Neuroscience*, 11: 404-409.

McCabe, David P. & Castel, Alan D. 2008. Seeing is believing: The effect of brain images on judgments of scientific reasoning. *Cognition*,107: 343-52.

McCabe, Kevin, Houser, Daniel, Ryan, Lee, Smith, Vernon, & Trouard, Theodore, 2001. A functional imaging study of cooperation in two person reciprocal exchange. Proceedings of the National Academy of Science. 98: 11832-11835.

Merikangas, James R. 2008. Commentary: Functional MRI Lie Detection. *J Am Acad Psychiatry Law*. 36: 499-501.

Mikhail, John, 2008. Moral Cognition and Computational Theory. In Walter Sinnott-Armstrong, ed., *Moral Psychology Vol. 3, The Neuroscience of Morality: Emotion, Brain Disorders, and Development*, Cambridge, MA, MIT Press, at 82-91.

Miller, Greg, 2009. Brain Scans of Pain Raise Questions for the Law. *Science*, 323: 195.

Moll, Jorge, de Oliveira-Souza, Ricardo, Zahn, Roland, & Grafman, Jordan, 2008. *The Cognitive Neuroscience of Moral Emotions*, in Walter Sinnott-Armstrong, ed., *Moral Psychology Vol. 3, The Neuroscience of Morality: Emotion, Brain Disorders, and Development*, Cambridge, MA, MIT Press, at 1-17.

Morse, Stephen J., 2006a. Moral and Legal Responsibility and the New Neuroscience, in *Neuroethics in the 21st Century: Defining the Issues in Theory, Practice and Policy* (Illes, J. ed,) pp. 33-50. Oxford: Oxford University Press.

--- 2006b. Brain Overclaim Syndrome and Criminal Responsibility: A Diagnostic Note. *Ohio St. J. Crim. Law*, 3: 397- 412.

--- 2006c. Addiction, Genetics and Criminal Responsibility, *Law and Contemporary Problems*, 69: 165-___

--- 2008. Determinism and the Death of Folk Psychology: Two Challenges to Responsibility from Neuroscience. Minn. J. L. Sci & Tech. 9: 1-36.

Nesse, Randolph N. 2001. The Evolution of Subjective Commitment, in Randolph M. Nesse, ed. *Evolution and the Capacity for Commitment*, New York: Russell Sage Press.

Nolan, James L. Jr., 2003, *Reinventing Justice: The American Drug Court Movement*, Princeton: Princeton University Press

O'Hear, Michael M. 2009. Rethinking Drug Courts: Restorative Justice as a Response to Racial Injustice. Forthcoming, *Stanford Law and Policy Review*, available at http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1365027

Parkes, David C. 2001. Iterative combinatorial auctions: Achieving economic and computational efficiency. Unpublished dissertation. Available at http://www.eecs.harvard.edu/~parkes/diss.html.

Phelps, Elizabeth A. 2002. The cognitive neuroscience of emotion. In *Cognitive Neuroscience: The Biology of Mind, 2nd ed.* (eds. M.S. Gazzaniga, R.B. Ivry, & G.R. Mangun), pp. 537-536. NewYork: Norton.

Phelps, Elizabeth A. & Sharot, Tali, 2008. How (and Why) Emotion Enhances the Subjective Sense of Recollection. *Current Directions in Psychological Science*, 17: 147-152.

Prehn, K., Wartenburger, I., Mériau, K., Scheibe, C., Goodenough, O. R., Villringer A., van der Meer, E., and Heekeren, H. R., 2008. Individual differences in moral judgment competence influence neural correlates of socio-normative judgments., *Social Cognitive and Affective Neuroscience* 3(1): 33-46.

Prinz, Jesse J. 2007. *The Emotional Construction of Morals*, Oxford: Oxford University Press.

Pustilnik, Amanda C., 2006. Prisons of the Mind: Social Value and Economic Inefficiency in the Criminal Justice Response to Mental Illness. *Journal of Criminal Law and Criminology*, 95: 217-___.

Rolls, E.T. 1999. *The Brain and Emotion*. Oxford: Oxford University Press.

Roskies, Adina L., 2006. Neuroscientific challenges to free will and responsibility. *TRENDS in Cognitive Sciences*, **9**: 419-423.

--- 2007. Are Neuroimages Like Photographs of the Brain? *Philosophy of Science*, 74: 860–872.

Sanfey, Alan G., 2007. Social Decision-Making: Insights from Game Theory and Neuroscience. Science 318: 598-602.

Sapolsky, R. M. 2004. The frontal cortex and the criminal justice system. *Phil. Trans. R. Soc. B*, 359: 1787–1796

Schultz, Wolfram, 2008. Introduction. Neuroeconomics: the promise and the profit. *Phil. Trans. R. Soc. B* 363: 3767-3769.

Seo, Hyojong & Lee, Daeyeol, 2008. Cortical mechanisms for reinforcement learning in competitive games. *Phil. Trans Royal Soc. B*, 363:3845-3857.

Sinnott-Armstrong, Walter, 2009. Neural Lie Detection in Courts, in *Using Imaging to Identify Deceit: Scientific and Ethical Questions*, Cambridge, MA: American Academy of Arts and Sciences.

Sinnott-Armstrong, Walter, Roskies, Adina, Brown, Teneille. Murphy, Emily, 2008. Brain Images as Legal Evidence. *Episteme* 5: 359-373.

Zak Paul J., 2004. Neuroeconomics. *Phil. Trans. R. Soc.* **B**, 359:1737-48.

--- 2007. The Neuroeconomics of Trust. In Roger Frantz, ed., *Renaissance in Behavioral Economics*, Abingdon: Routledge.

--- 2008. The Neurobiology of Trust. *Scientific American*, June, 2008, 88-95.

Zak, Paul J., Kurzband, Robert, and Matznere, William T. 2005. Oxytocin is associated with human trustworthiness. *Hormones and Behavior*, 48: 522-527
.