

Beyond the Bad Man and the Knave: Law and the interdependence of motivational vectors.
Yochai Benkler

In his seminal work of legal pragmatism, *The Path of the Law*, (1897), Oliver Wendell Holmes, Jr. wrote:

“If you want to know the law and nothing else, you must look at it as a bad man, who cares only for the material consequences which such knowledge enables him to predict, not as a good one, who finds his reasons for conduct, whether inside the law or outside of it, in the vaguer sanctions of conscience....

Holmes' recommendation was that, as a positive matter, law as a social fact *is* constituted by the predictions of the bad man about the likely application of public force in cases that might involve his behavior. Coupled with David Hume's normative recommendation that “in contriving a system of government... every man ought to be supposed to be a knave and to have no other end, in all his actions, than private interest” (1742), Holmes and Hume provide the foundations of the law and economics movement of the past four decades. Law as a practical fact acts through its effects on material interests of individuals; the method of legal design ought therefore be to predict the effects of law on the material interests of individuals assumed, for simplicity, to be self-interested. Nothing so well captures this spirit than Douglas Baird's statement of the core simplification that informed economic analysis of law: “people maximize, markets clear.” (Baird 1997 p. 1132). In order for markets to clear whatever it is that people maximize, people must maximize something that markets can price. Hume, like Adam Smith in *The Moral Sentiments* (1759), had nonetheless emphasized the importance of moral sentiments, or emotionally- and socially-enforced intrinsic constraints on individual action, and their importance in complementing markets and self-interest in allowing society to function well. Indeed, Hume specifically argued that the supposition of knavery, while true in the design of political institutions, is “false in fact.” But both Hume and Smith were able to sustain the recognition of the importance of moral sentiments in interpersonal relations, with the independence of interventions aimed to harness self-interest to the public good, by maintaining an assumption of strict separability between the effects of interventions on material interests or preferences, and their effects on the emotional drivers, social motivations, and moral commitments that provide decisional guidelines for many of us, for much of our behavior (Bowles and Hwang 2008). Assuming that it was possible to intervene to affect the material interests without affecting the direction of any of the other motivational vectors, the law could be designed for Holmes's Bad Man and Hume's Knave without affecting the good man. The same assumptions continue in contemporary thinking about institutional design and incentives. Thus, Kenneth Arrow was able to both identify the necessity of trust and trustworthiness to correct markets given persistent imperfection of information (1971) while at the same time rejecting Richard Titmuss's (1970) claim that a blood market crowded out blood donations to the detriment of the quality of the blood supply, arguing that donors and sellers respond to different incentives systems and are unaffected by each other (Arrow 1972). Extensive experimental and observational evidence, however, has established that interventions with well-defined effects on material interests are not separable from, or additive to, their effects on other motivational vectors. (Surveys: Bowles 2008; Frey and Jegen 2001). Explicit guidelines for action, coupled with either material rewards or punishment, can result in either increase or decrease in the magnitude and directional effect of other sources of motivation, which in turn may act either to negate or amplify the intended effects (Bowles and Hwang 2008). That is, forcing law can either crowd-out, or crowd-in other, motivations. There is some work in law on dynamic preference shaping by law (Bar-Gill and Fershtman 2004; 2005). There has been

substantial work within the law and economics movement focused on social norms, considering the effects of law on social norms and on the use of law to shape behavior through its influence on social norms (Cooter 1998a, 1998b; Lessig 1995; Kahan 1997; McAdams 1998). That work largely took the effects of law through norms and through material incentives as additive. Norm-driven behavior and incentives-driven behavior were imperfectly substitutable, but they were treated as separable and additive.¹ There is some work on reciprocity that sees law as potentially crowding-out reciprocity dynamics. (Kahan 2003; Jolls 2002; Scott 2003). There is no synthesis of the literature on the interdependence and misalignment of motivational forces and its application to legal design. This paper outlines such a synthesis, and uses it to explain core features of legal practice—the pervasiveness of normative argument, the centrality of proceduralism, and the pervasiveness of vagueness and mushy standards—and offers an example of how such an approach could be brought to bear on specific legal debates, such as the matter of efficient breach, or the advisability of gap-filling rules in contract. In the process, the paper offers a structured approach to analyzing interventions intended to shape incentives in the presence of interdependent motivations beyond what is currently available in the existing economic literature on crowding out.

Four motivational vectors and framing

To begin to analyze the ways in which law is designed, and can be improved, to deal with non-separable or imperfectly separable motivational dynamics, we need (a) a characterization of the set of motivational forces; and (b) a characterization of the mechanisms or dynamics of interaction between motivational forces in response to interventions aimed at one of these forces. Substantial experimental and observational literature has established that people behave towards each other in ways that are more consistent with substantial levels of altruism and reciprocity, with solidarity and empathy, responding to considerations of fairness, moral commitment, and social conformism, in addition to their responses to reward and punishment. (Camerer, 2003; Henrich et al., 2005; Gintis et al., 2005). These can be usefully clumped into four major motivational elements:

- *Material interests*: These includes elements of material welfare, are most readily convertible to a universal medium, money, and are partly and imperfectly convertible into satisfaction of social, emotional, and moral motivations, as in the case of status, which in contemporary society is partially susceptible to purchase as an external measure of success or autonomy, or the ability to use money to pursue the instantiation of moral commitments through donations.
- *Emotional responses*: In the context of the literature on non-separability of motivations, the theory that speaks most directly to crowding out emphasizes a background of emotional needs for autonomy, competence, and relatedness (Deci and Ryan 1999; Deci and Ryan 2000). This theory anchored Bruno Frey's importation of crowding out into economics. (Frey 1997; Frey and Jegen 2001). One need not, however, necessarily accept Deci and Ryan's particular theory of needs to accept that there is a dimension of a person's processing of the world around them, be it their brain makeup or their personality, that will move them to act in various ways and

1 E.g., Cooter 1998a says: “Assume that enacting a law without enforcing it induces at least 31% of the citizens to change, so the system in Figure 4 moves to the stable interior equilibrium. Although improvement is dramatic, 25% of the actors continue doing wrong. Further reductions in wrongdoing would require state coercion. Supplementing informal sanctions with state coercion shifts down the curve representing expected payoffs to wrongdoers in Figure 4, thus reducing the equilibrium number of wrongdoers.” This is perfect example of the way in which he assumes non-substitutability, rather than non-separability. The introduction of coercion here only affects the payoffs to the recalcitrant wrongdoers. It has no negative effects on the motivations of the “rightdoers”.

respond differently to materially similar situations. The role of emotions and affective responses in mediating human behavior is central to psychology (Davidson, Scherer, and Goldsmith 2003). In addition to Deci and Ryan, one mechanism proposed for support of prosociality is the empathy-altruism hypothesis, suggesting that empathy plays a core role in sustaining prosocial behaviors. (Batson 1991). This approach has received substantial support from work in neuroscience in the past 5 years. (Zak and Barraza 2009). The emphasis is on emotional or “gut” responses, as opposed to more reasoned, consciously or sub-consciously, processes, has also long been present, if ignored, in economics. (Smith 1759; Frank, 1988).

- *Social motivations*: By social motivations I mean to capture three distinct motivational elements that may be uncorrelated with, and on occasion oppositional to, the pursuit of material interests in the short or intermediate term. First, and most readily susceptible of assimilation to material interests, is social capital. (Granovetter 1974; Ben-Porath 1977; Coleman 1989). The idea here is that some social contacts function like capital in making one more economically productive, but they are not fungible with material resources. Because they are not fungible, however, the introduction of money into a social connection can reduce or eliminate its value as social capital. (Benkler 2004). Second, social network analysis suggests that our daily behavior, including what we eat, is affected by what our close neighbors in social networks do. (Fowler and Kristakis, 2008; Hanaki et al 2007). This suggests that we act in part by benchmarking our own behavior against that of others we deem “similar” to us, and imitating the behavior we deem to be “normal” or successful. We have, it appears, a preference for conformism; at least many of us do, and sometimes, in teenagers most obviously, this can take the form of sub-group scripted non-conformism. Third, we have a specific sub-set of emotional needs that have to do with relatedness to others. Whether this takes the form of emotionally seeking others (Deci & Ryan) or of a strong response of in-group bias to a solidarity group from which we take our identity or from which we expect high degrees of reciprocity (Tajfel and Turner 1979; Yamagishi et al, 1999; Efferson et al 2008), we have a motivational drive to belong and connect with others that may be distinct from what they can practically offer us in the future as social capital or from conformism.
- *Moral commitments*: We are capable of moral commitments to guide our behavior. While the term “moral sentiments” as a critical glue for material interaction goes back to Adam Smith, in contemporary economics Amartya Sen's *Rational Fools* (1977) is likely a good place to locate the emphasis on the distinct force applied by moral commitment. There is substantial current debate over the extent to which moral commitment is best explained in terms of mixed emotional and sub-conscious and semi-conscious cognitive processing, or in terms of a theory based more purely moral emotions. (Hauser 2006; Greene 2007; McGuire et al 2009; Green 2009). Participants from all sides of that debate, however, identify and assume a distinct force, instantiated in our brain and psyche, that incorporates moral judgments and influences our actions upon them. Some of what goes into moral commitment is an aspect or dimension of conformism. But enough of what goes under conformism is clearly not understood in moral terms, like fashion, and enough of what we understand to be “moral” drives action that leads to social isolation when it leads to resist a socially-shared behavior that one deems immoral—be it racism or homophobia in some groups, or refusal to smoke pot or drink in others—that it is worth locating in its own independent category. What makes these moral commitments “commitments” is that they are not continuous functions. They include ranges in which they take on an infinite value preventing or demanding an action. Disgust responses to taboo actions, like incest or eating taboo foods, are the obvious example. But they are also not purely that—they can also simply moderate action. Moral commitments probably include distinct

lines for: what is the fair thing to do, identified with substantial work on fairness and inequality aversion as a motivator for action (Fehr and Gächter 2000); what is the right thing to do (that is, what is behavior that would be appropriate for anyone acting under these conditions to do) (Hauser 2006; Greene 2007); and what is the good thing to do, which has more to do with identity and virtue than necessarily with more Kantian notions of right or justice and comparative conceptions of fairness. (Batson, 1998; Kahneman and Knetsch 1992). These subsets of moral commitment can explain and prescribe different aspects of how law acts and can be made to act to frame actions in terms that are susceptible to assimilation into these different modalities of moral commitment.

Each vector is, as is obvious from the brief characterization, itself a fairly complex amalgam of different elements. I cluster them together in order to ease the conceptualization of potential misalignments, or “crowding out” effect. But the value each will take will itself be the sum of its internal dynamics. Understanding the limitations of clustering these diverse motivational forces into a mere four, we can nonetheless think of these motivational vectors as discrete arguments within a utility function driving a given action. We could denote them as M, E, S, and R (for moRal or for Right). Each of these motivational dimensions can be thought of as a directed force operating on a person's behavioral decisions (and by saying “decisions” I don't mean to imply fully rational and articulated, but simply to describe the actual action settled upon in the face of non-perfectly-forced practical choice.) Each has a magnitude and direction, in the sense of pulling toward one or another action from a set of feasible actions. For simplicity, this could be modeled simply as having a magnitude and a valence (positive, negative) relative to a discrete question about each possible action (yes/no, act/do not act). The critical insight of non-separability of motivations is that interventions that may have a well-understood magnitude and direction/valence within one of the vectors, in particular material interests, may have effects within the dynamics of one or more of the other forces, which may work in the same direction or a different direction, at a force that may be equal, greater, or smaller. Ignoring these effects will lead an intervention to undershoot or overshoot its desired behavioral effect, depending on the magnitude and direction of the effect on other vectors. (Bowles and Hwang 2008; Benabou and Tirole 2006; Bowles 2008). Motivational vectors are (a) discrete: each one has its own internal dynamics of which way it will point; (b) interdependent: any single design intervention can affect more than one vector, and the sum of motivational forces will depend on the total effect; and (c) susceptible to misalignment.

All of these motivational forces are implemented in a cognitive framework through which we perceive our situation and our choices. Framing can affect our evaluation of the material interests involved, as in the case of loss aversion or irreversible indifference curves. (Kahneman and Tversky 1979, Knetsch 1989). But framing also affects our understanding of the emotional, social, or moral situation and the range of possible actions. (Goffman 1974). Often, there may be competing explanations of which motivational pathway will be affected by a given frame intervention, such as the possible interpretations of the Wall Street vs Community Game experiments. (Lieberman, Samuels, and Ross 2004). There, a materially identical finite prisoners dilemma game was played, with one group told it was playing the Wall Street Game, and the other the Community Game. Levels of cooperation began and remained high in the latter, and defection began and remained high in the former. This behavior could be explained in material terms—the frame led agents to self-fulfilling predictions about how others would respond to cooperation. But it could also be explained in terms of understanding what the socially conformist thing to do is; or of framing the decision as one where self-interest is the appropriate moral framework. (Hoffman et al., 1994; Cardenas et al., 2000; Gneezy and

Rustichini, 2000; Tversky and Kahneman, 1981). In other words, framing could have its effect early in the decisional process, placing the decision on a track where certain considerations, like cost, simply would not enter the calculus, or farther along in the decisional process, where it is appropriate to make a cost-benefit analysis, but the frame weights moral commitment more or less heavily. In modeling terms, the idea of “appropriate moral framework” should translate into a framing parameter that allows for interaction between the four motivations, such that under certain circumstances the weight of M is magnified, and in others, E, S, or R might be. More research is needed on which pathways are affected by which kinds of framing intervention. Law, in any event, includes elements, as we will see, that are best explained as performing framing effects that moderate potential tensions between material consequences and moral, social, and emotional motivational drivers.

Misalignment Mechanisms

Several mechanisms and pathways have been proposed for how interventions aimed at nudging material motivations can affect the magnitude and direction of other motivational forces (Bowles and Hwang 2008). These include:

- *Normative Framing.* Explicit incentives can frame a decision as one where self-interest is the appropriate moral framework, as opposed to one where other-regarding moral commitments are appropriate. (Hoffman et al., 1994; Cardenas et al., 2000; Gneezy and Rustichini, 2000; Tversky and Kahneman, 1981)
- *Interparty negative signaling.* Use of explicit incentives signals the beliefs of one party to an interaction (the incentive-provider) about the other party—about their competence (Benabou and Tirole 2003) or trustworthiness (Falk and Kosfeld 2006; Fehr and List 2004).
- *Confounded social signaling.* Explicit rewards or punishments in social settings can confound the signals of social behavior—where pro-social behavior can be interpreted as self-interested, and thus when agents observe each other behaving “cooperatively” they do not learn that cooperativeness is good, but that punishment avoidance or reward seeking is good. (Benabou and Tirole 2006; Seabright 2004; Hanaki et al 2008)
- *Confounded individual signaling/identity formation.* Explicit rewards or punishment can confound the individual's ability to use given acts to form an identity, a self-conception, or to self-signal that he or she is a certain kind of person he or she wants to be. (Benabou and Tirole 2006)
- *Contrarian self-assertion.* Use of explicit incentives can be experienced by the person offered as an effort to control them, and a rejection of their own internal motivation, particularly for tasks that are understood as not universally money-motivated. (Deci and Ryan 1999; Frey 1997; Falk and Kosfeld 2006)
- *Antisocial Negative Reciprocity.* Explicit negative payoffs can lead to negative reciprocity, which may take the form of non-cooperation (Fehr and Rockenbach) or of anti-social punishment, either directly reciprocal or even anticipatory, in expectation of future punishment (Gächter, et al 2005, Hermann et al 2008; Dreber et al 2008).
- *Endogenous shifting of levels of prosociality.* In a dynamic analysis, the presence of explicit rewards and punishments can decrease the desirability of or payoffs to developing or possessing preferences or stable behavioral-psychological characteristics or identity that are oriented toward non-material motivations. (Bowles 1998; Bohent et al 2001; Falkinger et al 2000; Gächter and Falk 2002; Gächter et al 2007; Bar-Gill and Fershtman 2005.)

To recapitulate. We can usefully describe our myriad motivations as four motivational vectors: material interests, emotional responses, social motivations, and moral commitments. For given interactions, the relative weights of each of these vectors are given by a framing function, which weights both the relative magnitude of the four vectors, and influences the relative magnitude of elements within each vector, such as whether a given act is disgusting or merely uncomfortable on the emotional level, or whether a given act is morally troubling or taboo, and hence takes a continuous moderate value or a discontinuously large value within R . Several pathways are currently suggested and modeled or experimentally tested, by which the direction or magnitude of one or more of the vectors other than material interests can be affected by interventions intended to affect the magnitude and direction of the material interests vector—by offering rewards or threatening punishment for actions that comply with, or diverge from, respectively, the desired behavior.

Introducing motivational interdependence and misalignment into design and law

A recent experiment concerned with the effects of money on blood donation offers a simple example of how one would introduce insights about motivational interdependence and misalignment into institutional design. Mellström and Johannesson (2008) introduced three treatments to a group of potential volunteers who were facing physical examinations to clear them for offering blood donations in Sweden. While this is not a perfect setup for studying crowding out in a general population, because this is a preselected relatively donative group, it will serve for purposes of illustrating how design interventions informed by an understanding of motivational interdependence can operate. Sweden has a background of a purely voluntary blood donation system. Volunteers must pass a physical exam first. The experiment consisted of three treatments administered before the exam was taken, and the output was the number of participants who are willing to go through with the exam and become eligible to be donors. The control group was given no treatment. Of the control group, 43% in fact volunteered to complete the exam. One treatment group was offered 50 SEK to volunteer to go through with the exam and donate blood. In that treatment group, 33% volunteered, a difference that was not significant. The effect was, however, significant in women, who dropped from 52% to 33%. In both cases the predicted effect, either at a significant level or as a trend, appeared as predicted by Titmuss. The authors hypothesized, based on Benabou and Tirole (2006), that women were more susceptible to social signaling effects. The third treatment therefore introduced a possibility of maintaining a clean social signal of donative intentions. Potential volunteers were given the option to donate their 50 SEK to the Swedish Children's Cancer Foundation. This treatment eliminated the negative effect in women, but the overall level of giving was brought back to the original, 44%, not higher. In sum, the modified intervention did no better and no worse than pure reliance on volunteering. But the experiment is useful in the sense that it suggests that: (a) motivational vectors are discrete targets; (b) there can be subpopulation effects in different directions, in this case, gender-based; (c) the effects are susceptible to theoretical prediction; (d) they are susceptible to experimental piloting; (e) their effects are complex, and require both experimentation and *ex post* observation to have significant confidence in the effects of policy interventions.

These observations are entirely sympathetic to the American pragmatist way of doing law, whether it be law and economics or, broadly speaking, law and society. Unlike much of current law and economics, except the behavioral branch, it requires different behavioral assumptions, and a more complex set of potential effects to be considered. Unlike the bulk of behavioral law and economics, it focuses on the non-material components of the utility function, rather than on systematic biases in evaluating information, or assessing outcomes, or evaluating the material consequences of action.

While the insights into interdependent motivational vectors are susceptible to theoretical prediction and experimentation, and thus to scientific analysis, the state of the underlying social and behavioral science has not reached a stage at which theoretical prediction alone will give us determinate answers to the effects of law on behavior, and thus on outcomes. Given the complexity of human motivation, it is not clear that we will ever be able to recover the simplicity and predictive clarity of simple rationality and the assumption of independence of material incentive effects from other motivational vectors. But what we might lose in clarity of prediction we will gain in scientific validity and quality of institutional design over time.

In the literature on law and interdependence of motivations, four major mechanisms have been offered, which cover only some of the pathways and possible effects. First, following Deci and Ryan's model of contrarian self-assertion, Frey suggested that highly controlling and bureaucratic structures and non-participatory constitutional structures would result in less publicly spirited behavior. He showed that tax compliance, for example, was higher in looser, more trusting tax systems, than in more tightly-monitored, punishment-based systems (Frey 1997), probably due to "tax morale" effects (Graetz et al 1986). Second, law can affect the value of developing and exhibiting a reliable pro-social motivational profile. It can undermine the need for that profile where it perfectly replaces internal social motivations with extrinsic control and monitoring (Bohnet, Frey, and Hack 2001, for the hypothetical high-enforcement state). It can undermine the social value of prosocial behavior by confusing the social or self-signaling effects (Bar-Gill and Fershtman 2005). Law can also degrade the value of developing a prosocial motivational profile where law exists at a moderately effective level, but is corrupt and imperfectly applied, so that its application itself can be gamed through antisocial moves to undermine the material and social payoffs to prosocial behavior (Bohnet Frey, and Hack 2001). Third, law can increase the value of a pro-social motivational profile where it reduces the payoffs to antisocial behavior, so that marginal players who could have adopted either an antisocial or pro-social profile develop a prosocial profile, further strengthening the social norms-based enforcement against antisocial actors, and further reducing their expected payoffs of defection. (Cooter 1998a). This latter is consistent with the findings that punishment can improve cooperation (Bowles and Gintis 2002; Fehr and Gächter 2002; Fehr and Rockenbach 2002; Falk et al, 2005), where legal enforcement takes the role played by strong reciprocators in those lawless experiments. To the extent that this is in fact the role of law; and to the extent that antisocial costly harming responsive to negative reciprocity is a species of interdependent motivations that can undermine the total value of cooperation (Dreber et al. 2008), it is possible that one of law's main effects is to provide punishment services in support of social cooperation while reducing the risk of cycles of antisocial negative reciprocity. If this were the sole or dominant effect of law, then Holmes's characterization of law as the predicted calculations of the bad man about his material payoffs from action deemed undesirable by law would be adequate. The critical difference in the effects of law on cooperation along this dimension from Holmes's view is that law is not irrelevant to the good man. Instead, it crowds *in* prosocial behavior over time. In that sense this approach treats motivations as endogenous and affected by law; but the mechanism is a traditional one: people adopt "good" dispositions by living in a society in which "crime doesn't pay." But each of the proposed three mechanisms for altering prosociality over time—contrarian self-assertion against a too-bureaucratic and controlling state, alteration of the payoffs to internalized trustworthiness and cooperativeness, or systematic reduction of the payoffs from defection—leads to dynamic changes at the population level in the prevalence of agents with prosocial motivational profiles relative to those with more material and self-interested motivations. In these approaches, levels of prosociality in a society are endogenous to its legal system. We do not, however, have evidence of such interdependence. The sole relevant piece of evidence, a multi-country study that found correlation

between levels of rule of law in a country and levels of antisocial punishment in experimental games, (Hermann, et al 2008), cannot distinguish the causal direction of the effect, and does not clearly correlate this antisocial costly harming with non-cooperation in the first instance of the games.

Normative Framing I: Material interests, social framing, and moral commitment

Holmes's admonition, with which the paper began, was a positive one—if you want to understand what law is, as a social fact, you have to look to its material consequences under given circumstances. That is all that matters, because only “the bad man” is ruled by law, as opposed to a broader range of considerations. This basic claim, that the effect of law on behavior runs through its effects on material costs and benefits of given actions to its individual subjects, has been the central working assumption of law and economics (Becker 1968; Posner 1973). The three primary alternative hypotheses that sound in normative framing or the role of legitimacy in the shaping of behavior (as distinguished from philosophical accounts of when law ought to bind morally, e.g., Raz 1977) have been:² (a) that law is part of the cultural practices through which we develop and adopt moral commitments, which, in turn, guide our actions (Cover 1983); (b) that law shapes the social meaning of an act, affects its social signaling attributes, and shapes the behavior of people who care about regulative responses of others in society like gossip or shaming (Lessig 1995; Kahan 1997); (c) that particular legal forms and aspects of legal practice can either “legitimate” official actions and cause internalized compliance (Tyler, 1990, Tyler, 2003) or fail to legitimate the acts and engender resistance (Frey 1997), a social-scientific claim with strong resonance with the critical idea of law as ideology or hegemonic practice (Kennedy 1982; Jost and Banaji 1994). Another potential approach, anchored in a more Aristotelian view of perfectionism, would also add the idea that law may affect practice coupled with normative framing. Updated with contemporary social psychology, it might suggest that habitual practice or even forced routine, through self-justification (Jost and Banaji 1994) and coherence seeking (Simon and Holyoak 2002, Simon et al 2004), can be internalized as a normative reason for action.

Assuming that people do indeed act in some measure in pursuit of what they consider to be moral commitments (Sen 1977), a legal practice that directly expresses the moral content of its commands can align its institutional commands with the perceived direct applications of moral principles and commitments. If it succeeds in doing so, then internal commitments to moral action will lead to congruence and strengthening of compliance. If it fails, then conflict between perceived legal commands and moral commitments will lead to weaker compliance. A legal practice that abjures moral mapping of its commands, as Holmes recommends in *The Path of the Law*, or that self-consciously treats the moral modalities within legal argument as pater (Cohen 1935), may at least forgo the opportunity to strengthen prosocial behavior, and at worst place material-incentives-based compliance with its commands in conflict with moral commitments. Take the relatively anodyne question of “efficient breach” in contracts. Contracts are enforced as a legal obligation, where the threat of a damage award is intended to act on the material incentives of potentially-breaching parties to prevent them from breaching. To the extent that agents inhabit a society that also holds the moral belief that promises should be kept, whether or not legally binding, that moral commitment adds a second motivational vector pointed in the same direction as the material incentives implied by the potential of contractual enforcement. An innovation of law and economics, the theory of efficient breach is that the refusal to enforce penalty clauses in contracts cases is evidence that common law looks favorably on

2 Cooter's views emphasized the role of law as a coordination focal point, which is independent of its normative claims, and as a source of coercion against defectors, which decreases their prevalence and increases payoffs to norm-followers. This too is a mechanism that passes through material interests, rather than normative commitment.

welfare-enhancing breaches—where fulfilling the obligations by the breaching party would cause it more harm than it would benefit its counterparty. (Goetz and Scott 1977). Such a legal framework assured that parties would not execute a contract whose execution would, by definition, decrease welfare. Philosophers of law have long argued that this is a fundamental misunderstanding of contract, which at root is a moral relationship—promise—whether as an expression of the individual autonomy of promisors (Fried 1981) or as an expression of community and human relations (Markovitz 2004). That complaint seeks to replace the conception of the normative task of law—from maximizing welfare to embodying a certain moral commitment to autonomy, or community. It does not respond to the core behavioral assumption: that if legal damages reflect the expectations of the non-breaching party, then the breaching party shall have internalized the costs of its action and will behave efficiently.

Consider the state of motivations for contract compliance effects under a rule that enforces penalties, assuming we accept the empirical validity of the misalignment mechanisms described above. Recall that the utility function includes M, S, E, and R. M will refer to the size of the expected payment upon breach, pushing material interests towards compliance. S stands for the social signaling effects of breach, symbolized by “Welchers are no good as social contacts.” E stands for the emotional response and sense of self related to performance, the sense of self-determination one has when one does what one has decided to do, etc. The R factor is the moral commitment of “promises should be kept.” It is likely the kind of commitment that for most people will be in the continuous part of the function, rather than in the same category as incest or eating a taboo food, in the discontinuous portion of the commitment function which is simply excluded as an option. In the baseline case, M has a magnitude related to expectation damages; S has a positive valence and a given magnitude related to signaling oneself as a good social contact; E, may be neutral, if the contract is undertaken under conditions of arms length bargaining but with a sense of choice, negative if the contract was undertaken under conditions experienced as “having no choice,” or positive in the case of a contract that is experienced as self-chosen and in pursuit of one's own goals; and R is positive and of a given magnitude related to the strength of “promises must be kept.” Now, imagine a legal system that enforces penalties, and a contract that includes an enforceable penalty. The effects against a background of an enforceable contract with no penalty would be as follows: The M factor increases by the magnitude of the penalty, and with a positive valence, towards compliance with observable requirements; The S factor would decline in magnitude, because the social signaling effect of compliance would be confused with the effects of the penalty clause. The E factor would change to a negative valence, consistent with both contrarian self-assertion and interparty negative signaling, as the requirement of a penalty would be experienced as an expression of lack of trust and/or a forcing mechanism. The R factor should not change, and continue to support “promises must be kept” to the same extent as without the penalty. Whether the penalty would, or would not, increase compliance, would depend on the relative magnitudes of the increase in M, the decrease in S, and the valence reversal in E. Now, imagine that law changes, and refuses to enforce penalties; for now setting aside the reasoning. Assuming that contracting parties would now cease to use unenforceable penalties, the effects would as follows. M would equal the expectation damages, a smaller magnitude than where the penalty exists. S would increase in the direction of compliance, as the signal of reliability becomes clearer. E would return to the neutral position. And R would remain positive and of the original magnitude. Compliance with observable aspects of the contract would decrease, with the decline in M. Compliance with unobservable aspects would increase because of the increase in S and the neutrality of E, now that the negative emotional response to penalty is removed. Whether this is material welfare³ improving or

3 It is difficult to speak of “welfare enhancing” in this case once we recognize that utility includes all four motivational factors, and they are all affected differently for different people, sometimes in different affected subpopulations, and

undermining depends on your empirical expectations about how important observable relative to unobservable aspects of performance are, and whether you are in a condition of over-performance of an otherwise inefficient contract, or underperformance. Because the increase in performance is along the unobservable performance aspects, it is likely to improve efficiency of performance within the range of efficient contracts, although not if this simply means higher investment yet in a contract that should be abandoned as inefficient.

Now imagine that the justification given for the refusal to enforce penalties is “efficient breach.” There is no effect on M and E. But S is now framed so that it is appropriate for contracting parties, as social actors, to seek their own welfare. The court adopting this theory effectively frames the prosocial thing to do as calculating at the margin as to whether compliance is efficient or not. The value of S declines, although its valence may not change. R, on the other hand, will decrease in magnitude or become neutral, because the moral framing of compliance and non-compliance is now explicitly decoupled from “promises must be kept.” The law frames the contracting relationship as one where “promises must be kept” is external to it, appropriate for social promises, which the law does not enforce, but not for commercial contracts. Compliance with observable requirements will decrease relative to penalty, and equal the level of M in the unframed no-penalty condition. Compliance with unobservable elements of contract fulfillment will decline relative to the normatively and socially unframed elimination of penalties, as the values of internalized self-monitored compliance, S and R, decline, while E remains unchanged. Overall performance will decline relative to the unframed no-penalty condition. Whether this is efficient or inefficient will depend on whether you think that observable performance aspects dominate, or whether the contracts governed by this framed rule usually are subject to overcompliance from high levels of unobservable performance, or subject to undercompliance with unobservable performance elements.

If one thought that the full costs of a breach may be difficult to assess; and “breach” may be hard to detect and takes the form of imperfect or half-hearted performance, then internal and internally-enforcing motivations are paramount. In such a case you might want to eliminate penalties, to prevent a negative valence in E, but to frame it in ways that support positive and large values for S and R. For example, you would say that the reason courts refuse to enforce penalties is that they are unfair, that they would result in “unjust, extravagant, or unconscionable *quantum* of damages n case of a breach.” (Goetz and Scott 1977, 555, citing traditional justification for refusal to enforce penalties). This is in fact how common law has justified this rule. Now, M is as it would be without penalties, and E is unchanged. But S and R are no longer re-framed as lower value. Indeed, contract remedies are located right back inside the domain of moral commitment and social obligation—of fairness and conscience. In these cases, an internal normative commitment to fulfill promises would lead to gap-filling fulfillment above and beyond what is institutionally enforceable. This, as it turns out, was the combined institutional solution in common law. Not simply the material interests elements, as the proponents of efficient breach proposed, but precisely the wrapping of the effect on material interest in a normative frame that also aligned the internal motivations. The normative framing is not pattered, as Felix Cohen would have it. The analysis here is not a moral one, as Fried or Markovitz would emphasize. Rather, the analysis is a positive, behavioral one, where the normative framing is part of the institutional design for diversely-motivated subjects under conditions of interdependent motivational effects on discrete motivational vectors.

might be endogenous to aspects of the interaction. Here, I only focus on the traditional concern of law and economics, efficiency, which in our case is limited to material welfare, as it is affected given the assumptions about interdependent effects on the four motivational vectors.

Normative Framing: Legitimacy and contrarian self-assertion

One function of law is to deter antisocial behavior. The quandary arises when we consider the Frey/Deci and Ryan point of contrarian self-assertion, and the social-psychology findings of compliance and its absence as related to the perceived legitimacy of the asserted control. The basic point is that law functions, when it functions as prohibition coupled with a threat of force, on the material incentives of its subjects. This was Becker's great innovation in reference to criminal law (Becker 1968). When it does so, it risks functioning like the parent saying to a teenager "If you drive recklessly, you're grounded!!" This is the classic crowding-out effect: the teenager would assert their self-mastery by screeching their wheels, at best only after they had turned the corner. The continued widespread use of p2p filesharing systems, after years of litigation by the recording industry against individual users, and after more-or-less everyone knows that using p2p networks is illegal, seems to be at least in part a result of users finding ways of seeing continued use of p2p filesharing as a way of "sticking it to the man" (Madden and Rainie 2005; Lessig 2008). The potential mediating force of law here is to convert this external regulation into what Deci and Ryan (2000) called "integrated regulation." That is, a set of commands with an external source, which are perceived by the individual as accepted and consistent with self-determination. Tom Tyler's work on legitimacy is pertinent here. In this work, legitimacy is a psychological affective response to an external command, characterized by internalization of that command. We would properly describe an external command as being experienced as "legitimate," if the individual internally takes the fact that the command was issued by the given authority, in the given form, with the given content, as a good independent reason for compliance (independent of any material or other social consequences of noncompliance). The effect here is on E, not on R. One way of interpreting works of philosophers on the conditions of legitimacy (Raz 1977) is describing the conditions under which those who hold R of a certain kind should also include compliance in their R vector. Beyond the scope here, but worthwhile noting for future reference, the kind of methodologically-individualist, empirical investigation of legitimacy as an affective response opens avenues to investigating how one might integrate social theories of legitimacy at a population level, which would have their effects not only in E, but in S and R as well. (Habermas 1984, 1987, 1996).

One can imagine legitimacy being generated by several different mechanisms. In the context of criminal enforcement, Tyler has shown that individuals' perceptions of the procedural fairness and motive trust of the enforcement process are the most powerful predictors of future compliance, particularly under conditions where monitoring is difficult and compliance has to be motivated by the internalization of the behavioral norm. (Tyler 2003). Frey, on the other hand, focuses on participation, and provides some evidence that legal systems that are more open to participation of citizens in law-making enjoy higher levels of voluntary compliance (Frey 1997). This is consistent with experimental observation that, where participants were able to choose their own rules of action and punishment for noncompliance, levels of cooperation in public goods experiments were high and punishment rarely necessary (Ostrom, Walker and Gardner 1994). While Tyler's argument about the importance of perceived intentions to agent's perceived fairness of the interaction, and hence willingness to accept it, is consistent with experimental findings, it is less consistent with experimental findings that participants care about the substantive fairness of the distributive outcomes independently of caring about the fairness of intentions of the other participants (Fehr and Schmidt 2001). How bad the unfairness of substantive distribution of outcomes from a legal process needs to be before subjects on

the losing end of the distribution begin to either question its procedural fairness (as a proxy), or directly reject the system itself, is hard to tell. There is certainly evidence that people on the losing end of a social distribution have a substantial capacity for system-justification (Jost and Banaji 1994; Jost, Banaji, and Nosek 2004). In general, it appears that one way in which the legal system reduces the potential negative effects of its nature as command backed by force is through systemic *procedural* fairness. In biblical text, Moses is enjoined to create an independent, impartial judiciary: “appoint judges and officials for each of your tribes in every town..., and they shall judge the people fairly. Do not pervert justice or show partiality. Do not accept a bribe, for a bribe blinds the eyes of the wise and twists the words of the righteous.” (Deuteronomy 16:18-20). In medieval Europe, procedural justice was publicly performed in its most extreme form through the ordeal, “precisely for dealing with situations in which certain knowledge was impossible but uncertainty was intolerable.” (Bartlett 1986, 33). In today's courthouses, blindfolded *Justicia*, her scales in hand, plays a central role in signifying that procedural impartiality. A central aspect of judicial practice is the performance of necessity—the organization of legal materials in a given case to show the inevitability of the result, its necessity irrespective of what the judge wants personally. (Kennedy 1997). Law performs impartiality and procedural fairness, almost to the point of comic exaggeration, sometimes with costumes, wigs, and excessive performance of deference toward officials, likely in part to sustain its legitimacy in the eyes of litigants, half of whom will generally go home having lost, but whose continued internalized compliance needs to be sustained. To the extent that the treatment of law and its fairness or impartiality are subject to different understanding in different classes in society, so that certain elites understand the human elements in legal argument and decision, while others see it as remote, impartial, and legitimate, to that extent law can be used by the former to cause the latter to accept, as an internal, psychological matter, substantive inequalities that they might have resented and resisted in the absence of law's mediation. This is the idea of law as a hegemony. (Kennedy 1982). Whether or not one accepts the claims of insidious social control through law, the core practice of combining highly symbolic procedural performances with continuous references to a shared morality and normativity is certainly a common practice in law. For people at least partly motivated by moral commitment and norm compliance and partly psychologically responsive to procedural performance in experiencing a decision as legitimate, these practices seem to combine to play a role in assuring voluntary compliance even where enforcement is weak or impossible. That is, hyper-emphasis on procedure may contribute to aligning E, and possibly S, with whatever the command of the court is, that is to say, with the M incentives imposed by the court. To the extent that a court achieves this through normative suasion and is seen as legitimate, it will also add magnitude and align the valence of R. The perhaps surprising point here from the perspective of law and economics is that the overall procedural and rhetorical modes of the adjudication process that result in a command and potential penalty operate directly on individual-level motivations, at the behavioral level, for a given action, independently of the effect on material interests. If done poorly, they can lead to negative valences along these other motivational vectors, without any change in the content of the command or its material implications for the action and actor, or the can feed back on each other causing excessive motivation in the direction intended.

Signaling.

Two kinds of signaling are primarily at stake: interparty signaling and social signaling. Interparty signaling occurs when one party to an interaction acts in ways—in this case, by deploying material incentives, such as offering payment or threatening punishment—that signals to the other party a given negative belief about that party, which undermines the other party's cooperative motivations. This may be a signal that the offeror thinks that the offeree is incompetent (Benabou and Tirole 2003)

or untrustworthy (Fehr and Rockenbach 2002). In both cases, the meaning of the signal as perceived by the offeree makes that offeree withdraw from the cooperative exchange: refuse to perform as required by the offeror. Law can mediate this effect in two ways. First, law can “normalize” the monetization of a given interaction, thereby removing the implication that offering payment is an interparty signal of disrespect, suspicion, or lack of confidence in the competence of the offeree. If offering money is simply the way things are required legally, for example, then the offering of money ceases to be an interpersonal signal (although that does not change the fact that it changes the normative and social valence of the activity). Second, law can depersonalize the act. If law sets a certain default relationship, in the absence of choice by the parties to the contrary and without a performance that can render its existence as a default salient in the minds of the parties, then that relationship obtaining as between the parties is not an affirmative signal from one to the other. It becomes a background constraint on their behavior, if it is a negative consequence, or a non-willed aspect of the interaction rather than an expression of the view of one regarding the other. Compare trade secret law to non-compete contracts. In the case of non-competes, courts enforce them, where they do, if they are entered into *ex ante*. Where they provide protection, they do so only in the context of relationship already framed as non-trusting. The classic analyses comparing Silicon Valley to Route 128, relating the success of the former to refusal by California courts to enforce non-compete agreements (Gilson 1999) fit the findings about motivation crowding out in the presence of efforts to control the employee's behavior (Falk and Kosfeld 2006; Fehr and List 2004). Trade secrets, by comparison, apply to information the employer deems secret and takes steps to keep secret, without requiring specific agreement from employees to that effect, and hence without sending the negative interparty signal. (Uniform Trade Secrets Act 1985, secs. 1(2), 1(4)). Nonetheless, ill-advised practices can cause parties to lose this desirable feature of trade secrets law by requiring a contract up front nonetheless.

The World Intellectual Property Organization's tutorial for small and medium size businesses offers a nice example of such an error. It calls for employers to require employees to sign confidentiality agreements, under the heading “Employees are the biggest threat,” the division advises that “An employee automatically owes confidentiality to the employer. However, in an environment where employee mobility is high, “psychological contracts” are no longer reliable; it implies that giving formal legal contracts a greater importance becomes important. Such contracts or clauses in contracts enhance legal protection of trade secrets and provide an enterprise security in case of litigation.” (WIPO SME division 2009). Such a token recognition of the new psychological contract literature (Stone 2001) coupled with advice that directly contradicts what we know both from that field work and from the experimental and theoretical work on crowding out is as crisp an example of policy that ignores the non-separability of motivations as we will likely find. A background rule of law that makes trade secrets law available as a background constraint on all employees retains the ability of employers to use the regime without forcing them to confront employees with distrust from the first day. Voluntary departures from that affordance in pursuit of litigation advantages against the few rotten apples come at the cost of reduced reciprocal trustworthiness of the workforce as a whole.

A similar idea seems to animate the following quotation from *Shepard v. Purvine* (Or. 1952):

These people were close friends and neighbors, and they were not dealing at arm's length. One's word was considered as good as his bond. Under the circumstances, for plaintiffs to have insisted upon a deed would have been embarrassing; in effect, it would have been expressing a doubt as to their friend's integrity. We do not believe that the evidence warrants a conclusion

that plaintiffs were negligent in not insisting upon a formal transfer of the rights accorded. An oral license promptly acted upon in the manner plaintiffs acted is just as valid, binding, and irrevocable as a deeded right of way.

The court appears to recognize that one cannot simply add the invocation of a formal legal right to a friendly, neighborly relationship and expect that relationship to remain unaffected. By holding the oral agreement between neighbors to be as valid as a deeded easement, the court adapts law to make room for social cooperation. By refusing to make a formal invocation of the legal system a precondition to the availability of law, the court locates its authority as a background, uncertain power to be invoked only after the breakdown of social cooperation. To condition the availability of the power of the state on up-front formalization of the relationship, the court seems to assume, would not merely add a layer of security to an otherwise stable cooperative relationship. It would undermine the cooperative relationship. The idea that presence of law as a background norm, applicable to everyone, but therefore to no one in particular as an interpersonal communication, could avoid negative interparty signaling is consistent with the experimental findings that, while threat of punishment in two-player games can sometimes undermine cooperation, it improves cooperation in a multi-player public goods game, where the threat of punishment can be seen as pro-social sacrifice to the group, and not directed at any given player individually unless and until they have in fact defected (Fehr and Rockenbach 2002).

Social signaling is a different type of concern. This combines the insights from Benabou and Tirole (2006) with those of Hanaki et al (2007). The former identify social signaling as the act of signaling one's type to others in society, through one's behavior. It assumes that it is possible to gain status; make functional social connections; or engage in sustaining social relations by performing tasks that are culturally coded as social. The introduction of a material incentive clouds the quality of the signal. I might now be doing it for the money, for example, instead of for the reasons that earlier were important for me to exhibit to others in my social milieu. The latter looks more toward social network-based mutual observation and imitation. It assumes that cooperative behavior, to the extent it improves the payoffs to those who engage in it, signals to others, on the periphery of cooperative subnetworks, that they too might want to get in on the cooperation behavior. Here, again, monetary rewards can degrade signal quality when observers cannot tell what motivated a successful practice, and therefore obscure the fact that it was cooperativeness, or sociality, that was the trait to be copied, as opposed to assiduous pursuit of rewards. To the extent that law is perceived as imposing an effective constraint on behavior, it, like monetary rewards, would cloud the social signaling effect of a prosocial act. Note that there is a tension between this effect of law and the possibility of diffusing an interpersonal social signal by requiring and normalizing monetization. While requiring monetization in a certain context can eliminate the negative interparty signal, it will also have the effect of degrading the social signal of the behavior. It would be causing a positive shift in E at the expense of neutralizing the effect of S.

Punishment and risk of antisocial costly harming.

Punishment and its threat have been shown experimentally to be effective at disciplining that part of the population whose behavior is most consistent with *homo economicus*. (Bowles and Gintis 2002; Fehr and Gächter 2002; Fehr and Rockenbach 2002; Falk et al 2005). And yet punishment is not without cost. In particular, punishment can lead to its own cycles of negative reciprocity when parties can retaliate against their punishers, whether the initial round of punishment was deserved or not. (Hermann et al, 2008; Dreber et al, 2008, but Gächter et al 2008). The degree to which what is often called “antisocial punishment,” but should be called *costly harming* given that it is not tied to a

norm infraction, and thus is not properly considered “punishment,” occurs experimentally is cross-culturally diverse, and seems to be correlated with measures of rule of law. (Hermann et al 2008). Ostracism and public punishment can play an important role in maintaining social cooperation (Bowles and Gintis 2002). But where punishment is not formalized and is left to individual parties, its efficacy does not negate its capacity occasionally to reel disastrously out of control, as in the legendary Hatfields and McCoys or vendetta culture more generally. Antisocial negative reciprocity, or costly harming, is a species of crowding out, in the sense that imposition of an expected price (punishment) for antisocial behavior can lead to spiteful harming behavior, at least as long as the spiteful parties can decrease their counterparty's payoff by more than the punishment decreases their own (Falk et al 2005). Law is, if nothing else, a cultural adaptation that formalizes and institutionalizes punishment so as to lower the probability that it will engender these bouts of destructive negative reciprocity.

First, as discussed, law can engender a psychological response of legitimacy and integrated regulation, which in turn can lead the external punishment to be accepted internally as fair and as a guide to behavior. This is the most important effect, as it allows parties to accept punishment and conform behavior without seeking revenge and counter-punishment, and allows law to prevent negative reciprocity of types that are permissible without breaking the state's asserted monopoly over the use of violence. Second, law is a formal and depersonalized form of assigning and applying punishment. Its execution is based on role adoption, rather than on interpersonal anger or emotional reciprocation. The participants often are stripped of their names and become roles: the judge; counsel; defendant; plaintiff. In many cultures, the public performances of the participants are stylized and visually striking: the American simple black robes worn only by judges are among the least elaborate; wigs and robes, ceremony associated with courts, are all part of the public performance of justice, rather than personal punishment. They diffuse the possibility of antisocial costly harming of the punishers, because none of the actors appear to be there in person; they are there as role players. And they offer these roles as a buffer between the parties to the dispute so as to avoid the possibility of inter-party punishment. They do so, however, at a cost. The lost humanization in court often misses opportunities for more cooperative problem solving, a fact that has underwritten reforms in mediation and informal dispute resolution that specifically divert parties away from the formal, justice-based system into informal, more cooperative dynamics.

Core aspects of legal practice and the non-separability of motivations

A full catalog of the ways in which law can be explained or criticized from the perspective of the non-separability of motivations is well beyond the scope of this essay. Here I will only note three major characteristics of legal practice that can better be explained by non-separability than by a focus on the material incentives effects of law. The first two were already discussed: *formality*, or the centrality of process to law; and *normativity*: the continuous invocation of ethical norms in terms of right, fair, and good, and social/conventional norms, like reasonableness. The third is discussed in greater detail: *indeterminacy*: the pervasive use of of vagueness, gaps, conflicts, and ambiguities in law as a normal part of legal practice.

Formality and process. Lawyers are obsessed with process. Courts are sites of ceremonious performance that cannot be explained primarily by reference to precision or accuracy. Judges do not wear robes because it helps them get the facts straight, or their holdings be more predictable so that people can better plan their actions. Many of the rules of evidence are highly unlikely to produce

greater accuracy: Asking an expert to testify in person on the stand under dramatic cross examination is performance, not science. Accuracy would be better achieved on paper. Excluding evidence that is plausibly credible but prejudicial is about producing a certain performance of fairness, not about bringing all relevant information to bear on a decision. Preventing amendment of claims or proceedings after a certain point, finality, none of these are in the interest of *ex ante* planning based on the most accurate possible application of law to the context at hand for the decision maker. They are, however, consistent with the psychological finding that fair process is associated with higher degrees of experienced legitimacy, understood not as a moral or systemic requirement, but as a psychological experience of law that allows its subjects to internalize its commands as integrated regulation, rather than resisting them as extrinsic commands. Process depersonalizes the application of the command to a person in context, rendering it inappropriate to antisocial retaliation against the enforcing party. This is the effect discussed above in context of legitimacy and antisocial costly harming.

Normativity. Law consistently uses the language of normative commitment. It is rare to find a court explicitly abjuring the claim that its decision recognizes what is right or fair. Chief Justice Marshall's refusal to justify the claims of European settlers that property through conquest was right, opting instead to speak of the institutional constraint of courts of the conqueror to uphold the conqueror's law, was a rare exception. (*Johnson v. M'Intosh* 1803). In the usual case, judicial opinion writing will speak of right; of fairness and justice; or good faith; and of common actual human behavior, "normal behavior," as a source of normative input in to how one ought to act, by setting standards of action such as "the reasonable person," or importing commercial practice among merchants into commercial law. In English and American law, this emphasis on normativity was historically highlighted by the existence of courts of equity, whose remit was precisely to decide cases "in conscience," formally in the teeth of what common law required as a matter of formal process. In continental law, this would fall under the category of general clauses of the German Civil Code, such as the duty of "good faith," or the French doctrine of "un bon père de famille" (like a good father) (Zimmermann and Whittaker 2000). As the discussion of normative framing suggested earlier, one of law's main functions is to prescribe to people what a person who cares about doing what is right, doing what is fair, and doing what is normal for people in that position to do. The result is that through their narrative and analytic emphasis on normative claims, legal decisions speak not only, or perhaps not even primarily, to the litigants, but to society at large. Given the diversity of moral beliefs across human populations, it is likely that the capacity for moral commitment is a basic human capacity, but that it can take on many particular forms. (Hauser 2006). To the extent that judicial opinion-writing succeeds in aligning the perceptions of a large number of individuals in society of what is the right, fair, and normal thing to do with what law demands, then the demand is internalized, whether cognitively or emotionally, and reinforced along the moral commitment vector. In other words, R is aligned with M. To ignore this norm-setting, social meaning effect of law is to ignore quite possibly its main effect (Lessig 1995; Kahan 1997). There are large, historical debates over the importance of law in setting societal norms. Even on one of the most apparently law-driven norm-setting exercise, racial equality and desegregation, there is serious argument about whether court decisions like *Brown vs. Board of Education* drove changes in norms regarding racial segregation, or whether changing practices drove the law. (Rosenberg 1991; Klarman 2004). That debate, however, leaves untouched the question of whether the politically-driven statutory changes may have caused the shift in normative commitment, even if the Supreme Court's decisions did not. There remains substantial room for social scientific investigation of the question of the degree to which the primary effect of law is its structuring of material incentives, and how much of law's effect is its characterization of what a person bearing normal moral commitments, and seeking to signal normal social conformism, should do.

A major sub-set of that research agenda is investigation of the dynamic effect of normativity in legal opinions. That is, whether, assuming a change in legal opinions can shape moral commitments at all, it does so in a way that achieves sustained shifts in a population over time in the population subject to the law, or whether it must do so, if at all, through repeated normative symbolic intervention. That is to say, whether moral commitments are elastic, plastic, or neither. In the context of implicit bias, Banaji differentiates between elastic and plastic responses to debiasing interventions. Elastic responses are those that change behavioral measures of implicit bias in the short term after the intervention, biases that return after a period has passed. Plastic responses are those that change behavioral measures over time. Responses of the first sort require constant reinforcement. They suffer from a certain kind of weakness, therefore, but also allow for much greater steering over time as broadly shared moral convictions shift. Plasticity in human moral commitment would suggest that moral commitment was responsive to moral steering, but more slowly and with longer-term results. Taking the rapid revolution in moral perceptions of appropriate relations—in terms of race in the 1960s, gender in the 1970s, and perhaps most rapidly and starkly, sexuality in the 1990s and the 2000s, it appears that moral commitments are not fixed, but are susceptible to updating within a human lifetime. At a minimum, they are elastic. If human beings have a general capacity for moral commitment, which in turn takes on the surrounding moral culture, and if law is a cultural form through which we discuss and construct a given moral culture, then the contents of moral commitment in a given society can change to fit the prescriptions of its legal culture, and legal culture in turn can, and likely does, shift to fit changes in prevailing moral culture with direct bearing on legal prescriptions. Nowhere is this more crisply evident than in comparing the U.S. Supreme Court's core opinion on the rights of homosexual couples in the 1986: “To hold that the act of homosexual sodomy is somehow protected as a fundamental right would be to cast aside millennia of moral teaching.” (*Bowers v. Hardwick* 1986) to its core opinion on the same subject 17 years later: “At the heart of liberty is the right to define one’s own concept of existence, of meaning, of the universe, and of the mystery of human life. ... Persons in a homosexual relationship may seek autonomy for these purposes, just as heterosexual persons do.” (*Lawrence v. Texas* 2003). It is possible that the mechanism is relatively Kantian—that individuals take the normative narratives created by courts as the input into their moral commitments, whether implemented through cognitive or emotional mechanisms. A more plastic shift in norms may take a more Aristotelian form. Legal requirements can lead to practice. Practice, even when rooted in an external command, at least when that command is “legitimate” in the sense of internalized, can turn over time into virtue—a set of beliefs, desires, and practices that comes to characterize the individual exhibiting it (Benkler and Nissenbaum 2006). Through the cognitive practice of individuals to seek coherence between their actions and decisions and their beliefs or preferences (Festinger 1957; Simon and Holyoak 2002; Simon et al, 2004), and through the needs of individuals to justify their practices, particularly when these are incongruent with prior practices or perceptions of what is required of them (Jost and Banaji 1994; Jost et al 2004), people come to adopt the belief that their current practice is virtuous. They then continue the practice to show themselves that they are in fact the kind of person who adopts this virtuous position. One way of modeling this is as self-signaling—where individuals engage in a practice to signal to themselves that they are the kind of person that they would like to be. (Benabou and Tirole 2006). Law can instigate this shift by connecting required actions with plausible justifications to accelerate changes in practice, internalized moral commitments, and the adoption of the new practice as virtuous and ethically binding. While plausible as a reading of evidence from available social and behavioral sciences, both the Kantian and Aristotelian hypotheses about the role of law and practice on internalized commitment and virtue require substantial empirical investigation.

Indeterminacy. Law is full of gaps, conflicts, and ambiguities. (Kennedy 1976; Kennedy 1997; Llewelyn 1960; Cohen 1935). As soon as an area becomes so clear and predictable as to assure the knave that he can safely take advantage of a sucker, judges muddy up the waters to create uncertainty. (Rose 1988). For every textual rule of construction, we have a counter rule. (Llewelyn 1960). For every set of clear rules, we have a vague guiding standard. Indeterminacy plays a role in improving the dynamic value of prosocial motivational profiles and identities. Following Bohnet, Frey, and Hack (2001), we can say that if law were perfectly enforceable, there would be no need for trustworthiness or internal moral compass. Individuals would be indifferent when interacting with selfish or moral actors: all would be equally constrained by the rules. Perfect enforcement is, however, impossible. The only choice we really have is between middling enforcement and low enforcement. Clear rules that fit the world imperfectly (by definition, since the world is more complex and dynamic than rule sets of sufficient generality to allow planning across diverse and changing environments can be) create a middling level of legal enforcement. There are some settings where they will be perfectly enforceable. In others, there will be no ex-post adjustment to fit settings outside of the rule set. Shylock will have his pound of flesh. In this environment, trustworthiness increases in value, and people can seek out trustworthy parties, or else they must limit their transactions to those that fall very discretely within the bounds of the clear rules. (Arrow 1971). If, as Charles Dickens put it in Mr. Bumble's words, "the law is a ass—a idiot," if, in other words, the law does not provide mechanisms to import general understandings of proper behavior and human relations, then it offers no support for that trustworthiness, and provides opportunities for the knaves to prosper alongside the pro-social actors. Standards applied *ex post*, relying on general understandings of judges of what is, and is not, trustworthy behavior, create an uncertainty as to the payoffs of strategic behavior, while increasing the probability that cooperative behavior will pay off, and if it is met by defection, be protected. The quote from *Shepard v. Purvine* earlier in the paper identifies the effect. If the court in a state insists on a written instrument between neighbors, and will invalidate a neighborly agreement as to boundaries, it allows parties to plan more efficiently. But it also forces them out of a trusting, cooperative relationship, and into a formalized, arms-length one. By providing a background rule, applicable as between all parties, that it will enforce informal neighborly land use arrangements, the court allows neighbors to treat each other informally, through social exchanges, in the majority of cases, without ever having to resort to formal claims of right as against each other. It then offers a backstop against defectors should they emerge.

Robert Scott (2003; 2004) argues that when courts use contextual interpretation, good-faith duties, fair dealings, reliance, and other gap-filling rules in contracts, they crowd-out social enforcement and reciprocity, and therefore undermine social cooperative results in the majority of transactions: those not litigated. Classical contract law, which he argued enforced only the text of the contract, left more bargains unenforceable, and therefore the scope of reciprocity-based enforcement broader. Modern contract law crowds out this system of reciprocity-based self-enforcement. In the Bohnet et al framework, this would suggest that classical contract law was a low-enforcement setting, where developing a type—in this case, strong reciprocators with a reputation for retaliating against contract breachers—could gain higher payoffs, in which case over time more agents would adopt that profile (by process of imitation and self-formation or by a process of selection over time among businesses). The problems with this analysis are: (a) that it assumes only one type of prosocial motivation and enforcement mechanism—strong reciprocity—as opposed to considering moral commitment or normatively-driven behavior, which gives content to, and is reinforced by, many of the gap-filling rules; and (b) that the vagueness and uncertainty leave sufficient room for developing a

benefit from prosocial identity or type, while at the same time being somewhat tilted toward aiding prosocial actors against antisocial defectors.

Take Scott's hypothetical, *the case of the falling phosphate prices* (Scott 2004). Two companies enter an agreement: *A* will buy from *B* 250,000 tons of phosphate for five years at a specific price; the contract includes a merger clause. The market price of phosphate tanks. *A* reduces its orders from *B* to 50,000 tons at the contract price. *B* sues for damages equal to the difference between the market price of 200,000 tons and the contract price. *A* claims that course of dealings was to adjust quantities as market conditions changed, and lack of commercial reasonableness to the contract given the price drop. Scott implicitly accepts reformation as the “cooperative” outcome, but argues that courts should reject *A*'s defenses because to enforce cooperation would be equivalent to introducing third-party punishment into a cooperative relationship, which would crowd out cooperation. Scott's hypothetical suggests something fundamental about the choice between rules and standards. To understand why, we must first reconsider the “cooperative” outcome in this hypothetical.

To simplify, assume that the parties thought that phosphate prices were stable, and had entered the long term contract purely to reduce transactions costs. (If the contract had intended to assign risk explicitly, then the cooperative outcome is to leave it where agreed, rather than reform.) In that case, neither imposing the full contract nor accepting the buyer's defense is the cooperative outcome. Assuming that the experimentally observed mode of 50% distribution of gains or losses is a Schelling coordination point in market-based societies, the cooperative outcome would be that the buyer buy 125,000 tons of phosphates at the contract price, sharing the mutually-unanticipated loss equally with the seller. Enforcing the contract as written “punishes” the buyer, who has to bear an “unfair” burden (divergent from modal “fair” outcome). Accepting the buyer's defenses “punishes” the seller, who is left to bear 80% of the loss. A court faced with a binary choice, breach/no-breach, backed by expectation damages, has no non-punishing outcome.

The parties' expectations about litigation will, however, shape their cooperation about how to deal with unanticipated loss. It determines (a) whether they are in a dictator game or an ultimatum game, and (b) who is proposer/dictator and who responder. If parties are uncertain as to outcome, or if litigation is under the American rule and costly relative to contract value, they are in an Ultimatum Game. The responder has a serious option of imposing significant costs on the proposer by “rejecting the offer”—rejecting the proposed reformed contract and either breaching (if he is the buyer responding to a seller's requirement of a full or more-than-half order) or suing (if he is the seller deciding to reject buyer's offer of reformation to a lower quantity). If the parties are certain of the outcome, and litigation is either inexpensive relative to contract value or if the English rule applies to costs, then they are in a Dictator Game: whoever is expected by both parties to win an unfair share of the gains (or immunity to losses) at an acceptable cost is the dictator, who can impose whatever lesser loss on the counterparty that he chooses. The first major effect of the choice between a clear rule and a fuzzy standard is therefore the type of game the parties are in. In this case, if we aim to improve cooperation (without deciding for the parties what the cooperative outcome is), a fuzzy standard with litigation costs under the American rule would improve cooperation. DG is cooperative only if the dictator is altruistic. Cooperative outcomes in UG are driven by altruism, strong reciprocity, and are somewhat more robust to framing effects. Systematically, proposers share a higher portion of the pie with responders in UG than in DG, and few participants in market societies offer respondents nothing in UG, which is not true of DG (Camerer 2003).

Moreover, the judicial gap-filling rules are not a random or uncorrelated uncertainty introduced

into the system. They are explicitly framed in moral terms and oriented towards norms of practice in a given community of practice. They: (a) frame the relationship between trading partners as one that requires good faith, fair dealing, etc.; and (b) are tilted in likely outcome towards the outcome that most comports with what standard practices or understandings of what those morally-laden categories imply in the given social context. These, in turn, are implemented by a system that is procedurally and in terms of cultural performance and signification oriented toward achieving legitimacy and internalization of their expected outcomes. The meaning of the application of the law as background law in this case is to signal that the state backs a certain kind of cooperative business dealings; and that sharp dealing and taking advantage of mishaps will not be supported. The payoffs to behavior that is attuned to social norms and moral behavior are high; the risk to non-morally driven behavior higher than the risk to normatively-compliant behavior, and the risk associated with trusting that one's counterparties are indeed going to act in ways that are socially normal and capable of moral explication is lower than it might otherwise be. All this suggests that the relative long-term payoffs to adopting and sustaining an identity of an actor who is attuned to social norms and cares about moral commitments will likely guide one to behaviors that will result in congruent cooperative behavior from others, and will gain the support of the state in the teeth of defection, while strategic behavior operates in an uncertain environment, where the uncertainty is systematically biased against behavior that is understood within the social context as strategic and antisocial or contrary to prevailing norms. (Lack of congruence between the cultural assumptions of judges and the cultural assumptions of the parties are an important basis of critique of courts when they apply the judges' own, usually elite and majority group, cultural understandings to a setting, instead of being able to reflect those of the parties, where they are not. This goes, however, to whether the particular legal system is well-aligned with the actual distribution of normal and normatively-understood behaviors in its society, not to whether a legal system that uses vague rules guided by normative frameworks that are well-aligned with social understandings of the relevant behavior will elicit greater internalized compliance than one that adopts a sharp, rules-based approach.) Finding trustworthy partners, in the meantime, is still beneficial, because the cost of enforcing one's rights is high and victory never certain. Unlike the argument in Bohnet et al (2001), the institutional environment most conducive to the dynamic emergence of more cooperative types is not one where law enforcement is low, but rather an environment where law enforcement is vague but predictably tilted toward normatively driven behavior consistent with understandings of norms in the relevant setting.

Conclusion

Shaping the material consequences of action for self-interested rational subjects, Holmes's "bad man," has been the focus of institutional design for decades. It relied on the assumption that interventions that align the material consequences of action for individuals with socially desirable outcomes unambiguously improved the attainment of those desirable outcomes, because whatever other motivational drivers might affect individual behavior would remain unaffected, while the material intervention would direct action towards the desirable end. The assumption of separability of motivational vectors is not, however, supported empirically. Substantial work in both psychology and economics supports the proposition that a given intervention may have positive effects on motivation whose sum effect is action, by at least some agents in a population, that is the opposite of the desired effect. The mechanism is not the result of a market failure, information shortfall, or failed rationality. It is the result of a motivational profile that has different discrete arguments, several or all of which may be affected by a given intervention, even if that intervention is aimed solely at material interests.

The effects of an intervention are not unambiguously aligned, and their sum may result in reduction of a desired activity, rather than its increase, even if they consist of adding a material reward for such action, and vice versa. A useful way of clustering the motivational drivers is to conceive of material interests, moral commitments, social motivations, and emotional responses as four discrete arguments in a utility function, with a magnitude and a valence. Several mechanisms have been proposed for causing some of these arguments to change magnitude and/or valence in response to interventions that unambiguously increase the magnitude of the material interests vector and direct it toward the desired action. While incomplete, these mechanisms are supported by empirical evidence. These include normative framing, interparty negative signaling, confounded social signaling, self-signaling or identity formation, contrarian self assertion, antisocial negative reciprocity, and longer term, endogenous shifts in levels of prosociality as a result of shifting emphasis away from sociality-supporting systems. I have offered here an account of how practices central to law—its emphasis on procedure; its continued use of normative rhetoric and references both to substantive moral commitments and conformism as sources of law; and its widespread use of vague standards oriented toward normative commitments and “normal” practices can be explained in terms of aligning the material interests with the other motivational vectors. I suggested that these effects are susceptible to theoretical prediction, experimental verification, and observational assessment. Substantially more empirical work on separating out the relative weights of these effects; the universality of their application, and the dynamic effect is needed. But institutional design must adopt the practice of incorporating the discreteness, interdependence and potential misalignment of motivational vectors into its basic approach. We now know, theoretically and empirically, that if we design law for the bad man, we cannot assume that “the good one” will continue to “find[] his reasons for conduct, whether inside the law or outside of it, in the vaguer sanctions of conscience....”.

- Arrow, Kenneth J. 1971. Political and Economic Evaluation of Social Effects and Externalities. *Frontiers of Quantitative Economics*. (Michael D. Intriligator ed., 1971) . 3, 22.
- Arrow, Kenneth J. 1972. Gifts and Exchanges. *Philosophy and Public Affairs* 1:343 pp.349-350.
- Baird, Douglass. 1997. The Future of Law and Economics: Looking Forward. *Chicago Law Review* 64:1129-1165.
- Bar-Gill, Oren, and Chaim Fershtman. 2004. Law and Preferences. *Journal of Law, Economics, and Organization*. 20:331.
- Bar-Gill, Oren, and Chaim Fershtman. 2005. Public Policy with Endogenous Preferences. *Journal of Public Economic Theory* 7:841.
- Bartlett, Robert. 1986. *Trial by Fire and Water: The Medieval Judicial Ordeal*. Oxford University Press, Clarendon.
- Batson, C.D. 1991. *The Altruism Question: Toward a Social-Psychological Answer*. Hillsdale, NJ. Erlbaum.
- Batson, Dan. 1998. Altruism and Prosocial Behavior. In *Handbook of Social Psychology*. Vol. 2, ed. Daniel T. Gilbert, Susane Fiske, and Gradner Lindzey, 282-316. New York. McGraw-Hill.
- Becker, Gary S. 1968. Crime and Punishment: An Economic Approach. *Journal of Political Economy* 76:169-217.
- Benabou, Roland, and Jean Tirole. 2003. Intrinsic and Extrinsic Motivations. *Review of Economic Studies*. 70:489-520.
- Benabou, Roland, and Jean Tirole. Incentives and ProSocial Behavior. 2006. *American Economic Review* 96:1652-1678.
- Bohnet, Iris, Bruno S. Frey, and Steffen Huck. 2001. More Order with Less Law: On Contract Enforcement, Trust, and Crowding Out. *American Political Science Review*. 95:131-144.
- Bowles, Samuel. 1998. Endogenous preferences: the cultural consequences of markets and other economic institutions. *Journal of Economic Literature*. 36(1):75-111.
- Bowles, Samuel and SA. Hwang. 2008. Social preferences and public economics: Mechanism design when social preferences depend on incentives . *Journal of Public Economics* 92:1811-1820. p. 1812.
- Bowles, Samuel. 2008. Policies Designed for Self-Interested Citizens May Undermine “Moral Sentiments”: Evidence from Economic Experiments. *Science*. 320: 1605-1609.
- Bowles, Samuel and Herbert Gintis. 2002. Homo Reciprocans: Altruistic Punishment of Free Riders. *Nature* 415:125
- Camerer, Colin, 2003. *Behavioral Game Theory: Experimental Studies of Strategic Interaction*. Princeton University Press, Princeton.
- Cardenas, Juan Camillo, John K. Stranlund, Clive E. Willis. 2000. Local environmental control and institutional crowding out. *World Development*. 28(10)1719-1733.
- Cohen, Felix. 1935. Transcendental Nonsense and the Functional Approach. *Columbia Law Review*. 35:809-849.
- Cooter, Robert 1998a. Expressive Law and Economics. *Journal of Legal Studies*. 27:585-608.
- Cooter, Robert. Punitive Damages, Social Norms, and Economic Analysis. 1998b *Law and Contemporary Problems*. 60:73-91.
- Cover, Robert. 1983. Supreme Court Forward: Nomos and Narrative. *Harvard Law Review*. 97:4-68.
- Davidson, Richard. J., Klaus Scherer, & H. Hill Goldsmith, (Eds.). 2003. *Handbook of Affective Sciences*. Oxford University Press. Oxford, UK.
- Deci, Edward L., and Richard M. Ryan. 2000. The “What” and “Why” of Goal Pursuits: Human Needs and Self-Determination of Behavior. *Psychological Inquiry*. 11:227-268.
- Deci, Edward L., Richard Koestner, and Richard M. Ryan. A Meta-analytic review of experiments examining the effects of extrinsic rewards on intrinsic motivation. *Psychological Bulletin*. 125(6):

627-668

- Efferson, Charles, Rafael Lalive and Ernst Fehr. The Coevolution of cultural groups and ingroup favoritism. *Science*. 321:1844-1849.
- Falk, Armin, Ernst Fehr, and Urs Fischbacher. 2005. Driving Forces Behind Informal Sanctions. *Econometrica* 73:2017.
- Falk, Armin and Michael Kosfeld, 2006. The Hidden Costs of Control. *American Economic Review*. 96(5) 1611-1630.
- Falkinger, J., Ernst Fehr, Simon Gächter, and R. Winter-Ember. 2000. A simple mechanism for the efficient provision of public goods. *American Economic Review*. 90(1) 247-264.
- Fehr, Ernst and Simon Gächter. 2002. Altruistic Punishment in Humans, 415 *Nature* 137.
- Fehr, Ernst and Bettina Rockenbach. 2002. Detrimental effects of sanctions on human altruism. *Nature* 422:137
- Fehr, Ernst and John A. List. 2004. The Hidden Costs and Returns of Incentives—Trust and Trustworthiness Among CEOs. *Journal of the European Economic Association*. 2(5)743-771.
- Fehr, Ernst K. Schmidt, Theories of Fairness and Reciprocity, Evidence and Economic Applications, in *Advances in Economics and Econometrics*. In 8th World Congress, *Econometric Society* (M Dwatripont et al, eds 2001).
- Festinger, Leon. 1957. *A Theory of Cognitive Dissonance*. Row, Peterson.
- Frank, Robert. 1988. *Passions Within Reason: The Strategic Role of Emotions*. Norton. New York.
- Frey, Bruno. A constitution for knaves crowds out civic virtues. *The Economic Journal*. 443:1043-1053.
- Frey, Bruno, and Reto Jegen. 2001. Motivation Crowding Theory: A Survey of Empirical Evidence. *J. Economic Surveys* 15(5):589.
- Fried, Charles. 1981. *Contract as Promise*. Harvard University Press. Cambridge. Mass.
- Gächter, Simon, Elke Renner and Martin Sefton. 2008. The Long-Run Benefits of Punishment. *Science*. 322:1510.
- Gächter, Simon Benedikt Herrman, and Christian Thoni. 2005. Cross Cultural Differences in Norm Enforcement. *Behavioral and Brain Sciences* 28:822
- Gächter, Simon and Armin Falk. 2002. Reputation of reciprocity? Consequences of the labour relation. *Scandinavian Journal of Economics*. 104(1) 1-26.
- Gächter, Simon, E. Kessler, M. Königstein. 2007. Performance Incentives and the dynamics of voluntary cooperation. University of Nottingham, School of Economics working paper 2640.
- Gilson, Ronald J. 1999. The Legal Infrastructure of High Technology Industrial Districts: Silicon Valley, Route 128, and Covenants Not to Compete, *New York University Law Review*. 74:575.
- Gintis, Herbert, Bowles, S., Boyd, R., Fehr, E. (Eds.), 2005. *Moral sentiments and material interests: the foundations of cooperation in economic life*. MIT Press, Cambridge.
- Gneezy, Uri and A. Rustichini. 2000. A Fine is A Price. *Journal of Legal Studies*. 29(1) 1-17.
- Goetz, Charles and Robert Scott. 1977. Liquidated Damages, Penalties, and the Just Compensation Principle: A Theory of Efficient Breach, *Columbia Law Review*. 77:554
- Goffman, Erving. 1974. *Frame Analysis. An Essay on the Organization of Experience*. Harper and Row. New York.
- Graetz, Michael, J.F. Reinganum, L.L. Wilde. 1986. The Tax Compliance Game: Toward an Interactive Theory of Law Enforcement. *Journal of Law, Economics and Organization*. 39:1-32.
- Greene, Joshua. D. 2007. The secret joke of Kant's soul. In W. Sinnott-Armstrong (Ed.). *Moral psychology. The neuroscience of morality emotion disease and development*. (Vol. 3). Cambridge, MA: MIT Press.
- Greene, Joshua D. 2009. Dual-process morality and the personal/impersonal distinction: A reply to

- McGuire, Langdon, Coltheart, and Mackenzie. *Journal of Experimental Social Psychology*, doi: 10.1016/j.jesp.2009.01.003.
- Habermas, Jurgen. 1984. *The Theory of Communicative Action*. Vol. I: *Reason and the Rationalization of Society*, T. McCarthy (trans.). Boston: Beacon.
- Habermas, Jurgen. 1987. *The Theory of Communicative Action*. Vol. II: *Lifeworld and System*, T. McCarthy (trans.). Boston: Beacon.
- Habermas, Jurgen. 1996. *Between Facts and Norms: Contributions to a Discourse Theory of Law and Democracy*, W. Rehg (trans.). Cambridge, MA: MIT Press.
- Hanaki, Nobuyuki, Alexander Peterhansl, Peter S. Dodds, Duncan J. Watts. Cooperation in Evolving Social Networks. *Management Science*. 53:1036-1050.
- Hauser, Marc. *Moral Minds. How Nature Designed our Universal Sense of Right and Wrong*. (2006)
- Henrich, Joseph etc. Foundations of Human Sociality.
- Hoffman, Elizabeth, Kevin McCabe, Keith Shachat, and Vernon L. Smith. 1994. Preferences, property rights, and anonymity in bargaining games. *Games and Economic Behavior*. 7(3): 346-380.
- Holmes, Oliver Wendell, Jr. 1897. The Path of the Law. *Harvard Law Review* 10:457.
- Hume, David. 1742. *Essays: Moral, Political, and Literary*.
- Jolls, Christine. 2002. Fairness, Minimum Wage Law, and Employee Benefits. *NYU Law Review* 77:47.
- Jost, J.T., Banaji, M.R., & Nosek, B.A. (2004). A decade of system justification theory: Accumulated evidence of conscious and unconscious bolstering of the status quo. *Political Psychology*, 25, 881-919.
- Jost, John T. and Mahzarin Banaji, The role of stereotyping in system-justification and the production of false consciousness. *British Journal of Social Psychology*. 33:1-27.
- Kahan, Dan M. 1997. Social Meaning, Social Influence, and Deterrence. *Virginia Law Review*. 83:349-395.
- Kahan, Dan. 2003. The Logic of Reciprocity: Trust, Collective Action, and Law. *Michigan Law Review* 102:71.
- Kahaneman, Daniel and Jack L. Knetsch. 1992. Valuing Public Goods: The Purchase of Moral Satisfaction. *Journal of Environmental Economics and Management*. 22(1):57-70.
- Kahaneman, Daniel and Amos Tversky. 1979. Prospect Theory. *Econometrica*.
- Kennedy, Duncan, 1982. Antonio Gramsci and the Legal System. *ALSA Forum*. 6:32.
- Kennedy, Duncan. 1997. *A Critique of Adjudication [fin de siècle]*. Harvard University Press. Cambridge, Mass.
- Klarman, Michael. 2004. *From Jim Crow to Civil Rights: The Supreme Court and the Struggle for Racial Equality*. Oxford University Press. Oxford. New York.
- Knetch, Jack L. 1989. The Endowment Effect and Evidence of Nonreversible Indifference Curves. *American Economic Review*. 79:1277.
- Lessig, Lawrence. 1995. The Regulation of Social Meaning. *University of Chicago Law Review*. 62:943-1045.
- Lessig, Lawrence. 2008. *Remix: Making Art and Commerce Thrive in the Hybrid Economy*. Penguin Press. New York.
- Lieberman, Varda, Steven M. Samuels, Lee Ross. 2004. The Name of the Game: Productive Power of Reputations versus Situational Labels in Determining Prisoners Dilemma Game Moves. *Personality and Social Psychology Bulletin*. 30: 1175-1185.
- Llewelyn, Karl N. 1960. *The Bramble Bush: On Our Law and its Study*. Ocean Publications. Dobbs Ferry, NY.
- Madden, Mary and Lee Rainie. 2005. *Music and Video Downloading*. Pew Internet and American Life Project. <http://www.pewinternet.org/Reports/2005/Music-and-Video-Downloading.aspx>.
- Markovitz, Daniel. 2004. Contract and Collaboration. *Yale Law Journal*. 113:1417-1518.

- McAdams, Richard H. 1997. The Origin, Development, and Regulation of Norms. *Michigan Law Review* 96:338.
- McGuire, J., Langdon, R., Coltheart, M., & Mackenzie, C. 2009. A reanalysis of the personal/impersonal distinction in moral psychology research. *Journal of Experimental Social Psychology*, doi: 10.1016/j.jesp.2009.01.002.
- Mellstrom, Carl and Magnus Johannesson. 2008. Crowding out in blood donation: was Titmuss right? *Journal of European Economic Association*. In press.
- Ostrom, Elinor, Roy Gardner, and James Walker. 1994. *Rules, Games, and Common Pool Resources*. Posner, Richard A. 1973; 2007. *Economic Analysis of Law* (1st edition 1973; 7th edition 2007). Aspen.
- Posner, Richard A. 1992. *Sex and Reason*. Harvard University Press.
- Dreber, Anna, David Rand, Drew Fudenberg and Martin A Nowak. 2008. Winners don't punish. *Nature*. 452: 348-351.
- Raz, Joseph. 1979. *The Authority of Law. Essays on Law and Morality*. Oxford University Press.
- Rose, Carol. 1988. Crystals and Mud in Property Law. *Stanford Law Review*. 40:577.
- Rosenberg, Gerald N. 1991. *The Hollow Hope: Can Courts Bring About Social Change?* University of Chicago Press. Chicago. London.
- Scott, Robert E. 2003. A Theory of Self-Enforcing Indefinite Agreements. *Columbia Law Review* 103:1641 (2003)
- Scott, Robert. 2004. The Death of Contract Law. *University of Toronto Law Review* 54:369.
- Seabright, Paul. 2004. Continuous preferences can cause discontinuous choices: an application to the impact of incentives on altruism. CEPR discussion paper no. 4322.
- Sen, Amartya. 1977. Rational Fools: A critique of the behavioral foundations of economic theory. *Philosophy and Public Affairs*. 6(4):317-344.
- Simon, Dan and Keith J. Holyoak. 2002. Structural dynamics of cognition: From consistency theories to constraint satisfaction. *Personality and Social Psychology Review*, 6, 283-294.
- Simon, Dan, Chadwick J. Snow, & Stephen J. Read. 2004. The Redux of cognitive consistency theories: Evidence judgments by constraint satisfaction. *Journal of Personality and Social Psychology*. 86:814-837.
- Smith, Adam. 1759. *The Theory of the Moral Sentiments*.
- Stone, Katherine. 2001. The New Psychological Contract: Implications of the Changing Workplace for Labor and Employment Law. *UCLA Law Review* 48:519.
- Tajfel, Henri & John Turner. An Integrative Theory of Intergroup Conflict. In *The Social Psychology of Intergroup Relations*, (W.G Austin and S. Worchel eds. 1979).
- Titmuss, Richard M.. 1971. *The Gift Relationship: From Human Blood to Social Policy*.
- Tyler, Tom R. 1990. *Why People Obey the Law*. Yale University Press. New Haven, Conn.
- Tyler, Tom R. 2003. Procedural Justice, Legitimacy, and the Effective Rule of Law. *Crime and Justice*. 30:283-357.
- Tversky, Amos and Daniel Kahneman. 1981. The framing of decisions and the psychology of choice. *Science*. 211 (4481) 453-458.
- World Intellectual Property Organization, SME Division.
http://www.wipo.int/sme/en/documents/trade_secrets_employee_loyalty.html.
- Yamagishi, Toshio, Nobuhito Jin, and Toko Kiyonari. Bounded Generalized Reciprocity, Ingroup Boasting and Ingroup Favoritism. *Advances in Group Processes* 16:161-197 (1999).
- Zak, Paul and Jorge A. Barraza. 2009. *Empathy and Collective Action*. (unpublished; presented at conference on Context and the Evolution of Mechanisms for Solving Collective Action Problems, Workshop in Political Theory and Policy Analysis, Indiana University, Bloomington. Available <http://ssrn.com/abstract=1375059>.

Zimmermann, Reinhard, and Simon Whittaker, eds. 2000. *Good Faith in European Contract Law*. (Cambridge University Press).

Statutes and Cases

Uniform Trade Secrets Act. National Conference of Commissioners on Uniform State Laws.

Bowers v. Hardwick, 478 U.S. 186, 214-15 (U.S. 1986)

Lawrence v. Texas, 539 U.S. 558 (U.S. 2003)

Shepard v. Purvine, 248 P.2d 352 (Or. 1952)