Choice Over Choice-Restricting Institutions

A dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy at George Mason University

By

Mark Adams
Master of Arts
George Mason University, 2009
Bachelor of Arts
Liverpool University, 2004

Director: Charles K. Rowley, Professor
Department of Economics

Spring Semester 2013
George Mason University
Fairfax, VA
Dedication

To my parents whose love and support made this possible.
Acknowledgements

I would firstly like to thank Professor Charles K. Rowley for his continuous support by meeting with me each week, his guidance throughout this process, and his many insights and suggestions which have shaped this text. I must also thank Professors Peter J. Boettke and Todd J. Zywicki for their support and the many valuable insights and suggestions offered during the proposal and that I have gained from their publications.

I would also like to thank my sister and the many friends whose suggestions and support helped make this happen.

Finally, I must thank my parents whose love and encouragement throughout this entire endeavor is the reason this is possible.
# Table of Contents

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>List of Tables</td>
<td>vii</td>
</tr>
<tr>
<td>List of Figures</td>
<td>viii</td>
</tr>
<tr>
<td>Abstract</td>
<td>ix</td>
</tr>
<tr>
<td>1. Introduction</td>
<td>1</td>
</tr>
<tr>
<td>I. Rational Choice: Its Critics and the Challenge of Finding a Replacement</td>
<td>1</td>
</tr>
<tr>
<td>II. Selecting Institutions</td>
<td>9</td>
</tr>
<tr>
<td>III. The Co-Evolution of Institutions</td>
<td>13</td>
</tr>
<tr>
<td>IV. Organization of Dissertation</td>
<td>14</td>
</tr>
<tr>
<td>References</td>
<td>17</td>
</tr>
<tr>
<td>2. Paternalism, Social Exchange, and the Law</td>
<td>22</td>
</tr>
<tr>
<td>I. Introduction</td>
<td>22</td>
</tr>
<tr>
<td>A. The Rational Theory</td>
<td>23</td>
</tr>
<tr>
<td>B. Cognitive Limitations and the Law</td>
<td>25</td>
</tr>
<tr>
<td>C. The Purpose of This Essay</td>
<td>30</td>
</tr>
<tr>
<td>II. Conceptions of Rationality</td>
<td>31</td>
</tr>
<tr>
<td>A. Rational Choice</td>
<td>31</td>
</tr>
<tr>
<td>B. Rational Addiction</td>
<td>33</td>
</tr>
<tr>
<td>C. Bounded Rationality</td>
<td>39</td>
</tr>
<tr>
<td>D. Biases and Heuristics</td>
<td>43</td>
</tr>
<tr>
<td>1. Representativeness</td>
<td>41</td>
</tr>
<tr>
<td>2. Availability</td>
<td>46</td>
</tr>
<tr>
<td>3. Anchoring</td>
<td>47</td>
</tr>
<tr>
<td>E. The Social Instinct</td>
<td>49</td>
</tr>
<tr>
<td>F. Prospect Theory</td>
<td>52</td>
</tr>
<tr>
<td>G. The Role of Regret</td>
<td>55</td>
</tr>
<tr>
<td>H. The Two Systems of Reasoning</td>
<td>57</td>
</tr>
<tr>
<td>I. Better than Rational</td>
<td>64</td>
</tr>
<tr>
<td>J. Summary</td>
<td>69</td>
</tr>
<tr>
<td>III. Social Exchange</td>
<td>71</td>
</tr>
<tr>
<td>A. The Rational Ecology</td>
<td>71</td>
</tr>
<tr>
<td>B. Social Contract as Choice Architecture</td>
<td>74</td>
</tr>
<tr>
<td>C. The Social Exchange Dilemma</td>
<td>77</td>
</tr>
<tr>
<td>D. The Adapted Society</td>
<td>82</td>
</tr>
<tr>
<td>E. An Evolutionary Model of Norm Selection</td>
<td>89</td>
</tr>
<tr>
<td>IV. Prescriptions for Policy</td>
<td>97</td>
</tr>
<tr>
<td>A. The Knowledge Economy</td>
<td>97</td>
</tr>
<tr>
<td>B. Creating Social Norms</td>
<td>101</td>
</tr>
</tbody>
</table>
## List of Tables

<table>
<thead>
<tr>
<th>Table</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Debt and Equity as a Source of Finance</td>
<td>131</td>
</tr>
<tr>
<td>2. External Capital Markets</td>
<td>132</td>
</tr>
</tbody>
</table>
List of Figures

<table>
<thead>
<tr>
<th>Figure</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Niche Theory</td>
<td>153</td>
</tr>
<tr>
<td>2. Niche Theory with Learning</td>
<td>162</td>
</tr>
<tr>
<td>3. Environmental Shift</td>
<td>166</td>
</tr>
</tbody>
</table>
ABSTRACT

CHOICE OVER CHOICE RESTRICTING INSTITUTIONS

Mark Adams, Ph.D.

George Mason University 2013

Dissertation Director: Dr. Charles K. Rowley

Human behavior is shaped by the rules we impose on ourselves and those that are imposed on us. These rules can overcome limited cognitive abilities but are themselves constructed by boundedly rational humans. This dissertation examines rule selection over three spheres: social exchange, corporate governance, and the law, and presents a framework for comparing approaches to rule formation.

The first essay develops a theory of voluntary choice-restriction through social exchange. I review evidence for neoclassical choice theory, behavioral choice theories, and alternative theories in economics, neuroscience, and evolutionary psychology. While biases exist there are multiple strategies for aligning choice and true preferences. Humans have specific evolutionary programming for social exchange which can be used to frame choices. I present a model of social exchange in which boundedly rational individuals can efficiently choose a system of error-reducing rules by selecting over bundles of rules. I further show that over a significant range of cases this is more efficient than using designed rules.
The second essay extends the analysis to firms. I respond to both agency-based and behavioral critiques of the market-financed firm, arguing that a theory of bounded rationality is implicitly a theory of knowledge. Firms must be capable of innovating and responding to external change. I propose a model of choice-restriction within firms which incorporates learning, radical innovation, and investment under uncertainty. I compare competing ownership structures and argue that the market-financed corporation is an efficient response to bounded rationality and dynamic uncertainty. This ownership structure uniquely permits the firm to alter its form (morphogenesis) to either rigidity or flexibility. This allows the firm to conserve knowledge and reduce transaction costs but still respond to changing external market conditions.

The third essay examines the role of federalism and competition in the law. Theories of bounded rationality admit a role for the law in regulating relations between investors and firms, but political markets are limited by the same features which constrain effective monitoring by investors. This paper proposes a model of choice restriction in which decentralized governments use both internal (constitutional) and external (federalist) institutions to commit to providing good law in unknown future states of the world. I present a model of jurisdictional competition with semi-immobile capital. Parties who commit to a jurisdiction under uncertainty rely on institutional commitments to provide efficient future law and guarantee property rights. I show that under certain conditions, competition under federalism will provide the most effective response to crises and may reduce some externalities associated with risk-bearing.
1. Free to Not Choose: The Benefits of Restricting Choice

Human behavior is shaped by rules: those we impose upon ourselves, those we enter into voluntarily, and those which are imposed upon us only sometimes by our implicit consent. These rules are at once the mind’s architect, a response to the limitations of the human mind, and, most importantly, a product of that mind. Though these rules restrict choice they are restrictions that enhance rather than reduce the scope of what people may accomplish.

This dissertation is a study of choice-restricting institutions: the rules which guide our activity. It studies the interactive process by which restrictive institutions shape our environment and ourselves, and in turn, how we shape those institutions. It seeks to understand institutions not as a product of pure design or of outside forces but as an emergent product of human intelligence and human interaction. To achieve these goals I look at choice-restriction in three spheres: the social sphere, the market sphere, and the government sphere and compare the process by which rules are created in each sphere. In doing so the dissertation lays out a framework for comparing the creation of rules in each sphere and determining where different types of restriction might best be created.

I. Rational Choice: Its Critics and the Challenge of Finding a Replacement

A growing experimental literature has found stark differences between economic models of rational self-interest and observed behavior in a laboratory (Kahneman et al., 1991; Thaler, 1988; Sunstein, 1997; Jolls, Sunstein and Thaler, 1998; Sunstein and Thaler, 2003). Instead of
behaving like the selfish *homo economicus*, the subjects of experiments tend to be social, preferring fairness, cooperation and sometimes acting spitefully. Although these findings have been criticized for the artificiality of the experiment, for the fact that subjects are most commonly students, and for the lack of incentives, behavioral arguments become an established part of economics.¹ Yet, the assumptions of rational choice are not, and never were, justified because they themselves are realistic, but because making these assumptions allows economists to say something useful about the real world without understanding first all its complexities (Alchian, 1950; Friedman, 1953).

By placing real humans in simulated abstract worlds, experimentalists ignore Smith’s (1991) entreaty to understand and explain the “manner in which institutions serve as social tools that reinforce, even induce, individual rationality.” That is, critics of rational choice model must do more than demonstrate that *homo economicus* is unrealistic. They must develop an alternative model that includes both the limitations of boundedly-rational individuals and an understanding of the social, economic, and political institutions that shape behavior in the real world. Finally they must explain how institutions that allow humans to imitate the mathematically abstract *homo economicus* can come into being in the first place.

This endeavor is worthwhile for several reasons. The first is that the rational expectations revolution in economics has greatly strengthened the assumption of rationality (Muth, 1961; Lucas, 1972). The second is that while previously there had been few sensible alternatives to

---

¹ See Wallis and Friedman (1942) for an early example of the problem of artificiality. Frank and Schulze (1993) find that economics students may be unusually poor subjects because of their economics training. Smith and Walker (1993) review the evidence linking incentives and decision costs in the experimental literature.
the assumption of rationality, it is now possible to measure, however inexactly, human behavior under controlled conditions. The third is that exponential advances in computational power permit an alternative form of modeling, reducing the need for extreme mathematical abstraction.

At present, some scholars have already begun the task of creating an alternative to rational choice economics. This effort has found some evidence that models based on experimental evidence can be better predictors of reality than the rational choice model (for a general review see Thaler and Sunstein, 2003, 2008). These initial successes are somewhat limited, and largely confined to the financial literature, where there is less ambiguity over preferences and empirical measurement of outcomes is more easily attained (see e.g., De Bondt and Thaler, 1985, 1987, 1995). Attempts to gauge the effect of institutional constraints on behavior through computational simulation have similarly been most successful in financial markets, suggesting that a convergence is possible (for a general review of the behavioral and computational strands of financial markets literature, see Lo, 2007).

Outside financial markets, the study of institutional arrangements builds largely of Coase’s (1937, 1959, 1960) argument that such arrangements can overcome limitations caused by real world transactions costs. When there are clearly assigned property rights and no transaction costs, rational individuals can make efficient contracts in the atomistic world of homo

---

2 For example, Becker (1996) argues that any alternative approach – be it founded on ‘cultural,’ ‘biological,’ or ‘psychological’ forces – comes close to providing comparable insights and explanatory power.”

3 For a more specific example, see e.g., De Bondt and Thaler (1985, 1987) find evidence for investor overreaction and seasonality as predicted by experimental evidence. Also, Courant, Gramlich and Laitner (1986) and Thaler (1990) find that savings and consumption habits deviate from Friedman’s (1957) rationally predicted permanent income hypothesis.
economicus. It is only by adding a theory of transaction costs that economics can explain the existence of firms or say something useful about the law.

Rational calculators can write extensive contracts, predicting each and every eventuality. However, there are significant transactions costs in trying to write such a contract, compounded by the impossibility of predicting every future possibility and face. Williamson (1975) argues that hierarchies (organizations) are necessary to make commitments when faced with the dual problems of bounded rationality and of conflict of interests. In team production, firms can overcome both uncertainty and transaction costs by allowing for relational contracts which are more flexible (Coase, 1937; Alchian and Demsetz, 1972, Williamson 1975). When coordination beyond the firm is necessary, efficient defaults in the law can replace complete contracts and thus reduce contracting costs (Coase, 1959, 1960; La Port et al., 1996).

Hierarchies allow for decisions to be made on the spot, but they also allow individuals to access tacit knowledge embedded in rules or routines (Hayek, 1973; Nelson and Winter, 1982). This knowledge is derived from experience, and from attempting different approaches, rather than from design or centralizing information (Ibid). While it does not conform to neoclassical rationality, using rules to manage knowledge may be procedurally rational in the sense described by Simon (1972), whereby a set of ex ante rules for decision-making are efficient under the constraints of bounded rationality.

Simon’s concept applies equally to organizations, or to individuals, and indeed Simon (1972) makes no conceptual distinction between the two levels of decision-making. As such, there are clear parallels between the use of heuristic rules at the individual level, routines at the organizational level, and laws at the societal level. Heuristics arise from a combination of
evolution and response to environmental learning. The first essay of this dissertation describes how the human brain has evolved to solve problems through the use of heuristics, alongside rational calculation of the kind assumed by neoclassical models which reduce the costs of bounded rationality.

When an individual encounters a problem frequently (e.g. crossing the road) she can rely on heuristics rather than calculation (in this case, a “look left” heuristic). The use of heuristics is not limited by the complexity of the problem; Kahaneman (2003) notes that medical professionals rely on highly-specialized heuristics to make decisions. In other cases an individual might encounter a problem too infrequently to develop a specialized heuristic, or an evolved heuristic may be maladapted to the individual’s current environment. For example, the individual may have no experience of what might happen if she fails to save for retirement or may suffer from hyperbolic discounting which causes her to inadequately weight future periods. Relying on an heuristic may lead her astray but an institutionalized rule can help her to incorporate the experiences of others into her decisions.

These rules must, by their nature, restrict choices. For example, the individual may be encouraged to save by a social norm encouraging thrift, by a legal default that she saves a proportion of her income unless she opts out, or by a traditional paternalistic scheme such as social security. It is clear that the latter option is choice-restricting (she cannot opt-out) but the former options also restrict choice, albeit to lesser degrees. Were the individual a *homo economicus* then indeed the cost of deviating from the social norm or default option would be minimal, but this is not true of the real individuals at whom the institutions are aimed (Rebonato, 2012). If the cost of calculating the alternative to the heuristic or institutional
response was truly minimal then the purpose of the heuristic or institution would be nullified. Thus, any institution that changes the relative cost over a set of decisions can be considered choice-restricting.

If humans are prone to err, then the more restrictive institution may be the most appropriate. If people are biased to save too little then merely encouraging thrift will help some people save more, but some will still save too little. Social security closes the door to error by requiring everyone to save (c.f. Weiss, 1991). Yet institutions, like heuristics, are imperfect human constructs which have the potential to introduce biases. While good institutions can achieve Smith’s goal of reinforcing or inducing rationality, bad institutions can guide individuals towards bad choices. The first essay describes the constraints that bounded rationality imposes on policymakers, such as knowing the right level of saving for heterogeneous individuals. The second and third essays expand this problem to include the prevalence of conflict which can exacerbate the problems of bounded rationality.

Making institutions less restrictive makes it easier for individuals to deviate from their preferred choices. Less restrictive institutions can also take account of individual preferences. Whereas heuristics incorporate feedback at the individual level, institutions aggregate information at a group level. A person can save too much as well as too little (even if it were possible, it would clearly not be beneficial to save all of one’s income). What constitutes the right amount of saving depends on the saver’s own preferences and circumstance.

Making institutions less restrictive reduces the possibility that the institution will introduce new error but transfers some of the decision costs back to the individual and increases the chance of individual error. This is the problem of optimal plasticity described by Cosmides (1989). In an
evolutionary context excessive “plasticity is death” (Ibid): the human mind cannot rationally solve every problem it encounters with the limited capacity. However, total rigidity is equally harmful under the constraint of radical uncertainty. The purpose behind choice restriction is to reduce the cost of decision making and to incorporate information not easily obtainable by the individual. An heuristic or institution would not be needed in the first place if individuals would have made the same choice at low cost without the rule. Thus, when it easier for individuals to do so they will deviate from good institutions as frequently as from bad ones.

More complex rules might allow institutions to adjust to a greater variety of individual circumstances: a social norm of thrift could include an exception where people borrow to gain education. Increased complexity cannot include private knowledge or predict perfectly every situation. Further, as the rules increase in complexity those rules themselves impose greater costs. Not only must the individual must know all of the rules but those rules must also be somehow enforced.

The structure of enforcement itself may provide for some level of efficient deviation. How easily a rule can be bent depends on the institution that created the rule. Small firms may regularly break from routines whereas large firms may find it harder to do so. A social norm favoring thrift can easily be waived for a terminally ill man but a regulation favoring saving cannot be waived unless some formal exception is added to the rule or the designated enforcer is afforded some arbitrary power. If the latter case prevails and rules are ignored ad hoc then the rules themselves become less knowable. Uncertain rules introduce new costs as people search to

---

4 For a summary of the literature see Becker (2004). Becker also notes that firms may find increased efficiency in more rigid routines.
discover which rules apply. Uneven enforcement may also permit enforcers to act capriciously or to serve their own self-interest.

The third essay describes how the common law evolves in this way: when a judge confronts a novel problem he creates a new precedent for that situation. For this evolution to be effective there must be some force driving the selection of new rules. A new law might be a proper response to a novel problem, an error introduced to the system, or the product of deliberate rent-seeking. Firms and social norms too can be influenced by error or rent-seeking.

Institutions can deteriorate over time and rent-seeking can drive ever greater complexity in rules as exceptions are placed over exceptions (Olson, 1982).

As an alternative response to the problem of institutional selection, people may select between competing institutions. In this conception multiple institutions serving the same purpose must exist at the same time. Such an arrangement exists with firms, governments (in jurisdictional competition), and social norms. People can compare a bundle of rules by observing different outcomes without necessarily understanding how each rule acts. Competition may also weed out institutions that deteriorate over time.

For competition to be effective it must be possible for the individual to exit an institution.

Permitting exit inherently reduces the ability of the institution to restrict choice and thus places some limits on that institution’s efficacy. Further, if institutions are deselected too quickly then

---

5 The result of the increased cost of knowing rules may be the additional effort the individual expends on discovering rules or the harm resulting from lower compliance with rules. When rules are uncertain and discovery becomes sufficiently costly individuals may prefer to accept the costs of non-compliance.

6 The common law also permits novel responses to existing problems although in both cases there must be mechanisms for ensuring the law is universal.

7 Rent-seeking is defined by Tullock (1967) as the wasteful expenditures to capture monopoly rents. In this context the rent arises from any law which inefficiently favors one group over another.
that institution ceases to serve as a repository of stored knowledge. Thus, although institutions may be improved by competition, the efficacy of institutional arrangements requires some limitation on the ability of individuals to deselect institutions.

II. Selecting Institutions

When people choose among institutions they cannot know the full consequences of selecting a particular institution. An individual who could know this would not only need to be capable of complete rational calculation for all the choices she would face, but would need to perfectly predict every eventuality and calculate her optimal response in advance. Such a calculation, even if it were possible, would entirely invalidate the purpose of the institution. Nor can institutions be selected or at the time of the choice without making the selection of the institution and the selection of choice the same. To be effective at reducing decision costs an institution must be somewhat binding over choices where the individual does not know what decision is optimal and where, at the time of selection, the individual may not what future choice she will face.

Although she must make this choice she is not endowed, at the time of selection, with any greater knowledge or cognitive faculty than were she making the choice absent institutions. Yet she will expect the institution, once selected, to lead her to make better choices: combining her own preferences with knowledge not available to her. That selection must take place through one of two mechanisms: institutions are designed by an expert (Jolls, Sunstein and Thaler, 1998) or emerge through spontaneous order (Smith, 1991). While institutions, like heuristics, are more likely a combination of these two factors, it is helpful to consider how institutions that arise purely through one or the other mechanisms would behave.
In the former conception Jolls, Sunstein and Thaler (1998) examine how choice architecture can help people align their choices, and the eventual outcomes that spring from those choices, with the individual’s true preferences. The authors argue that a benign paternalist (though not necessarily government) can help people make better choices by altering the way in which choices are presented to the individual. The idea of a choice architect appeals to specialization. People frequently devolve decision-making to experts such as hiring a stock broker or consulting a doctor. If an individual is unable to make rational choices in every area of life perhaps an army of qualified experts can. Nevertheless, experts present problems. The architect is not omniscient: she cannot know the private preferences of every person for whom she designs institutions. Further, unless she is completely benevolent and known to be the most competent person for the task she must somehow be monitored.

For stock brokers and doctors it is relatively easy for the expert to know the patient’s preferences (to make money or get better) and for the customer to monitor outcomes. Nevertheless, conflict and error still exist in these relationships. Investors may confuse high returns from excessive risk-taking with high returns form a particularly competent stock broker, or wrongly attribute their sickness to malpractice. Jolls, Sunstein and Thaler (1998) argue that in evaluating decisions people tend to use a hindsight bias: evaluating the quality of the decision based on the outcome rather than the information available to the expert at the time.

The individual may not even know the range of possible outcomes. The only observable outcome is that which occurs. The individual may not know whether that outcome was good or bad when compared with the range of non-occurring outcomes. Fama and Jensen (1983a,b) argue that when individuals devolve decision making to an expert, the very fact of that expert’s
specialist knowledge is a barrier to perfect monitoring. When the expert is expected to innovate, the problem of monitoring and contracting over future outcomes may become impossible.

In Fama and Jensen’s (Ibid.) example, managers are unable to perfectly monitor employees, leading to rent-seeking, or shirking, within the firm. Yet the problem faced by the firm (increasing profit) is a relatively simple one. A regulator acting on behalf of a consumer faces a much more complex problem: the firm at least knows that more profit is good but the regulator is unaware of the private preferences of those he regulates. Yet he is the expert: if the regulated individual knew the ideal outcome then regulation would be unnecessary. If neither the regulator nor the regulated knows with certainty what outcome is optimal then selecting the appropriate rule is much harder and if the selection of experts is purely political, the problem may be intractable.

Selection of ideal rules will be difficult when the regulator is purely benevolent but the regulator may have an incentive to select a rule which is not optimal. If monitoring costs are high then rent seeking will be more prevalent. Furthermore, the current discussion assumes that experts can be punished after a bad decision but contracts with experts will be necessarily incomplete. If the expert is hired to act as an agent to make decisions that agent must be given some discretion to make decisions. If there is risk or uncertainty over outcomes and experimentation is desirable, that discretion must include leeway to make errors.

Although the cost of detecting individual instances of rent-seeking may be high, individuals may take measures to avoid the risk. Both firms and governments have constitutions which not only limit the discretion available to experts but also the discretion shareholders and voters give
themselves. The constitution is itself a choice-restricting institution designed to improve the quality of other choice-restricting institutions (the law or organizational structure).

Although Sunstein and Thaler assume a choice architect, design is not necessary. Several scholars (e.g., Hayek, 1960, 1973; Rubin, 1977) argue that the common law is not designed but evolved over time. At first glance evolution appears to avoid the problem of selecting institutions by eliminating the designer. Further inspection reveals that eliminating design does not eliminate the need for human guidance: a fitness landscape must hold the preferences of each affected individual and good institutions must somehow be identified and promoted while deleterious institutions must be removed. If the criteria for selecting institutions are flawed then the eventual selection will also be flawed.

In fact the purely designed or purely evolved institutions are equally unusual. In the evolved common law rules may be selected by evolution but rules are not (as with genetic mutation) generated at random. Rather judges design rules with specific goals in mind and only those rules chosen by judges may be selected by the evolutionary process. In the designed code law rules can be changed if they prove unsuccessful and more than one set of rules can exist at once. Both common and code law systems copy one another and, consequently, frequently reach similar results (Tetley, 1999). Moreover, jurisdictions must always face some competition: even the Berlin Wall could not prevent exit or the eventual collapse of communism.

As with the law, firms and social norms are a mixture of design and evolution. Employees design procedures for handling problems but those procedures are replicated if they are successful. Clergy invent moral rules but religions may rise and fall. Although both evolution and design are present in all institutions in some mixture, this mix is frequently ignored by both conventional
economics and evolutionary economists but is central to understanding how institutions emerge.

III. The Co-Evolution of Institutions

In an uncertain world the optimal set of institutions will change. Indeed some institutions (such as firms) may exist to promote innovation and as such are themselves sources of uncertainty. Adapting institutions requires both design and trial and error. When institutions are capable of radical change failure becomes more likely. When institutions adapt too slowly they risk being left behind and replaced.

Institutions are human creations but humans are also a product of environment. We develop heuristics in response to our environment which consists in part of the institutions with which we interact (Smith, 1999, 2003). Those institutions are a response both to our cognitive limitations and to one another. Institutions are also inherently social in nature. Choice-restricting institutions do not only influence individual behavior in a vacuum but also mold interactions between humans, organizing social behavior or economic activity.

Subjects in experiments frequently cooperate even when the experimenters design incentives to discourage cooperation (such as eliminating repeated interaction and making it impossible to form contracts). If the same subject were forced lived in a society of the experimenter’s design (perhaps a society where enforcement of contracts was impossible) she would develop a different set of heuristics against cooperation. Such a society, absent outside enforcement of contracts might develop norms for cooperation but only in small groups. Members would then learn heuristics of cooperating in small groups but distrust of strangers.
Westerners are frequently trusting of strangers: a response to a set of laws which support anonymous interactions. Yet law is not the source of such norms but rather norms undergird society onto which laws are built (Pildes, 1996). Our law began by enforcing existing norms and continues because of current norms (Friedman, 1979; Schelling, 1960). Our changing world is not merely a response to law but a source of new law as judges and policy makers respond to novel situations. Thus the human mind is formed in part by our interactions with our environment and that environment is made in part by the human mind and molded by the interactions of institutions.

Yet changing institutions cannot be costless. If it were then the institution would fail in its goal of restricting future choices. Thus, institutional arrangements must include procedures for adapting to the changes that will create new and somewhat binding restrictions on the individual.

IV. Organization of Dissertation

This dissertation seeks to understand how individuals choose between choice-restricting institutions. I draw on a range of literature including experimental economics, law and economics, industrial organization, Austrian economics, and public choice.

The first essay will describe in detail the nature of bounded rationality in the real world, using concept of ecological rationality innovated by Gigerenzer (see e.g. Todd and Gigerenzer, 2007) and developed by Smith (2003). I will then use this evidence to compare the effectiveness of evolved social norms with designed rules. I will propose a model of how efficient social norms may be selected in light of the problems described above, examining how beneficial norms can
arise and enforced. In concluding I will look at specific policy areas to determine where social norms are likely to be most effective and where designed rules might be better. As part of this survey I will examine the case of corporate governance, where empirical evidence suggests a role for designed choice architecture.

The second essay examines corporations in more detail. I will look at how organizations balance the need for innovation with the need for control by owners, arguing that the joint-stock corporation is the most successful institution designed for this purpose. I will look at the institutional arrangements within corporations such as organizational structure, rules, and social norms. I will then examine cases where it necessary to break away from existing institutional arrangements and demonstrate how the legal structure of the corporation (itself an important institution) allows for major changes while preserving some of the organizational capital within the firm. I will argue that while some control rights associated with ownership may be transferred (and that this is sometimes efficient) others must remain inalienable. Furthermore, I demonstrate that while voting rights hold value, voting itself is not widely used and is not an effective control mechanism.

The final essay applies these lessons to government. I examine the role of corporate governance law as a solution to the problems of bounded rationality that affect firms and look at how efficient laws might emerge. I argue that while governments can provide theoretical relief to the problems that face the firm, many of those problems of are more prevalent with democratic institutions. I propose an alternative understanding of jurisdictional competition as a partial solution and incorporate a model of federalism as a commitment mechanism. As with the previous essays, there are competing demands for certainty and the ability to innovate. I
propose a model of jurisdictional competition where each jurisdiction seeks to commit to future
good laws on the basis of institutional arrangements. To succeed that the law must permit
beneficial innovation but prevent rent-seeking. In this conception, crises serve an important
function in jurisdictional competition by disrupting path dependent choices and permitting new
and more efficient laws to emerge. I use this model to evaluate the recent legislative action
known as the *Dodd-Frank Act* and argue that the law is likely to be counter-productive.

Each essay identifies an area in which individuals must make choices which will restrict their
future menu of choices. Individuals must balance the need for flexibility with the costs of
decision-making. By looking at such choices in three distinct areas I seek to demonstrate that
these are universal problems with clear parallels between each application. I further seek to
demonstrate a broadly applicable framework for comparing how competing institutions
function in a world dominated by uncertainty and cognitive limitations.
References


Friedman, M., (1953) *Essays in Positive Economics*, University of Chicago Press, Chicago


2. Paternalism, Social Exchange and the Law

Some of the nation's most prominent economists have tried their hands at refashioning American pension policy. Yet policymakers have generally ignored the economists' recommendations [...] for pension policy is paternalistic, and paternalism is something economists have a great deal of difficulty understanding.

Weiss (1991, p1276)

We [economists] have consumers with humanity, firms without organization, and even exchange without markets.

Coase (1988, p3)

I. Introduction

The rational choice paradigm, which still dominates mainstream economics, presumes that individuals act rationally in pursuit of their own self-interest (Becker, 1996; Sugden, 1991; Vanberg, 2004). By contrast, policymakers frequently assume that people commit errors (Weiss, 1991). Consequently, economic analysis is frequently disregarded by policymakers because it lacks a tool for understanding policies that are justified on paternalist grounds (Ibid.). If the
claims of rational choice were unimpeachable, or if paternalistic polices were so far outside the realm or economic thinking that economists had nothing to add to the policy debate, then this disconnect would be of little interest to either party. Yet economics is both compatible with weaker assumptions of rationality and well placed to evaluate the effectiveness of paternalist policies. I argue that economics beyond rational choice is not only possible, but necessary. This essay contributes to such an economics, building a model of human interaction based on less than fully rational individuals, and identifying the possible means by which those individuals may be led to better decisions.

A. The Rational Theory

Rationality is both an assumption and a prediction. The assumption claims that any individual behavior should be interpreted as rational as long as the behavior is internally consistent. If a man chooses to gamble his life savings or to live as an ascetic, the economist assumes that those decisions are intended rational acts designed to advance the man’s private preferences. Even apparently contradictory decisions can be made consistent with a rationality assumption. Jolls, Sunstein and Thaler (1998) observe that some people who like lobster, like it less when they have seen their meal alive. The authors argue that as all lobsters were once alive, whether or not the diner saw them in this state, the rationality assumption must be false. Posner (1998) replies that this observation is merely evidence that a lobster seen before its demise and an already dead lobster are simply different goods eliciting different preferences among consumers. The rationality assumption remains unchallenged.

See, for example, Choi et al. (2003), O’Donoghue and Rabin (2003), and Thaler and Sunstein (2003), for examples of policy proposals to correct less than fully rational behavior.
The ease with which almost any behavior can be reconciled with rationality makes the assumption difficult to test. If a man decides to destroy all of his wealth without apparent reason the rationality assumption simply states that he preferred to destroy his wealth than to keep it. As such the assumption alone is of little use to policymakers. To make falsifiable predictions, economists must make further assumptions about the nature of preferences. For example, by assuming that the man prefers greater wealth to less wealth the economist can predict that taxing cigarettes will discourage smoking. The rationality prediction thus makes stronger claims about human behavior than the mere assumption of rationality. The prediction is more useful to policymakers because it is empirically refutable.

It is the stronger rationality prediction that behavioralists seek to refute. In doing so the behavioralist must adopt her own assumptions about preferences: there is no use arguing that someone saves too much or too little without first assuming how much that person should want to save. Nevertheless, her assumptions may be as valid as the rational choice theorist’s and her arguments deserving of no less attention. Indeed, public policy and the law have often proceeded upon the assumption that individuals are less than fully rational (Weiss, 1991).

Despite the role rational choice has played in economic thinking, economics is still a valuable tool in evaluating policies which assume less than full rationality. Behavioralists can theorize about the benefits of paternalistic polices but can neither quantify those benefits nor predict which paternalistic policies will be most beneficial and which will be harmful. Economics has

---

9 A more complex set of assumptions would be required to explain how this leaves the man better off

been helpful in identifying how regulators can achieve their current goals at lower cost or improve upon outcomes at no extra cost (Tengs et al., 1995). Others have shown how economic analysis can be applied to pensions, savings, and sin taxes (Benartzi et al., 2004; Choi et al., 2003; O’Donoghue and Rabin, 2003). Sunstein (1996) and Thaler and Sunstein (2008) have demonstrated how policymakers can take advantage of our greater understanding of human behavior to achieve regulatory outcomes with less interventionist policies.

B. Cognitive Limitations and the Law

While economists remain reticent to treat people as less than rational, the law often does so. Rarely do disputes make their way to court where individuals have made no errors, or have written contracts which display perfect foresight. Behavioral evidence suggests that people cope with cognitive limitations by developing rules of thumb, or heuristics which, while useful, may result in biases which lead people to make errors.\textsuperscript{11} Cognitive limitations can also result in people making errors from not fully understanding agreements they enter into.

Traditionally, courts have mostly given short shrift to arguments based on these kinds of cognitive limitations. Courts grant exceptions on grounds of infancy\textsuperscript{12} or incompetence,\textsuperscript{13} and contracts may be discharged if some event unforeseen by both parties makes performance impossible\textsuperscript{14} or would lead to an outcome intended by neither.\textsuperscript{15} On the subject of unilateral

\textsuperscript{11} For a broad summary of the literature see Gilovich, Griffin and Kahneman (2002).
\textsuperscript{12} Cidis v. White, 71 Misc,2d 481, 336 N.Y.S.2d 362 (1972)
\textsuperscript{13} Ortelere v. Teachers’ Retirement Bd., 295 NYS 2d 506, 31 App. Div. 2d 139 (1968)
\textsuperscript{14} Taylor v. Caldwell, 122 Eng. Rep. 309 (K.B. 1863)
\textsuperscript{15} Sherwood v. Walker, 33 NW 919, 66 Mich. 568 (1887)
mistake, however, competent adults have traditionally been held responsible for their own failings.\textsuperscript{16}

In the last half century courts have relaxed these restrictions. Not only are debtors no longer sent to prison, bankruptcy law offers quite generous terms to those who have made serious errors (Jackson, 1985). Even greater protections are afforded through limited liability.\textsuperscript{17} The doctrine of impracticability affords greater leeway than does impossibility\textsuperscript{18} and courts usually deny punitive penalty clauses\textsuperscript{19} and give critical scrutiny to express conditions\textsuperscript{20} and liquidated damages clauses.\textsuperscript{21}

More recently, courts have become more critical of contracts that involve unilateral mistake. In the post-war era the doctrine of unconscionability has been developed and expanded to give reprieve to those who enter into contracts that are against their own interests.\textsuperscript{22} A term may be discharged if it constitutes unfair surprise again one party, thus assuming that a reasonable person may not be able to understand all the terms of a contract they have previously agreed to. In \textit{Gerhardt v. Continental Insurance Corp.}, the court refused to enforce a term within the fine print on the grounds it was “neither conspicuous nor plain and clear.”\textsuperscript{23}

\begin{itemize}
\item \textsuperscript{16} Harris v. Tyson, Supreme Court of Pennsylvania, 24 Pa. 347 (1855)
\item \textsuperscript{17} See Hart (2000) for a review of commercial bankruptcy law.
\item \textsuperscript{18} Uniform Commercial Code, Section 2-615 29; Mishara Constr. Co., Inc v. Transit-Mixed Concrete Corp., 365 Mass. 122, 310 N.E. 363, 367 (1974)
\item \textsuperscript{19} Patton v. Mid-Continent Systems, Inc., 7th Circuit, 841 F.2d 742, 750-1 (1988)
\item \textsuperscript{21} Lake River Corp. v. Carborundum Co., U.S. Court of Appeals, Seventh Circuit, 769 F.2d 1284. (1985)
\item \textsuperscript{22} Williams v. Walker-Thomas Furniture Company, D.C. Court of Appeals, 198 A.2d 914 (1964)
\item \textsuperscript{23} 225 A. 2d 328, 48 NJ 291, 48 NJ 2d 291 - NJ: Supreme Court (1966)
\end{itemize}
Eisenberg (1995) asserts that the court has gone further in recognizing cognitive limitations by reviewing the terms of contracts ex post to determine not only if a contract was unreasonable when it was signed but whether a term might be the result of a defect in the reasoning of parties. For example, Eisenberg argues that a party might agree to an express condition or a liquidated damages clause as a result of over-optimism. Thus even though both parties understood the meaning clearly, Eisenberg argues that neither truly understood the implications.

In rental law, the courts have also shown a willingness to invalidate terms that parties had understood and agreed upon. As recently as the early 1960s, courts assumed that a landlord was responsible for providing repairs only if he had expressly agreed to do so (Rabin, 1998). In *Javins v. First National Realty Corp.* the court found an implied warranty to ensure the rental property complied with building codes.24 This warranty was found to exist even if the parties had both understood that the property would not comply with code (Ibid.). Even if the tenant expressly assumed responsibility for repairs, the landlord can still be held responsible in many states. Thus, the courts went from assuming that tenants could predict the state of a rental property without expressly waiving the condition of code compliance, to assuming that even if tenants expressly waived conditions that they did not do so in their own interests.

In placing such restrictions on the ability to waive constraints, courts have followed legislators who created building codes with the same paternalistic intent. Programs such as social security and unemployment insurance do not give individuals the opportunity to opt-out. Instead, policymakers assume individuals will save too little and will underestimate the likelihood of

---

24 428 F. 2d 1071 - Court of Appeals, Dist. of Columbia Circuit (1970)
financial loss (Weiss, 1991). Indeed, the most heavily subsidized retirement plans are those which offer the least choice to savers (Ibid). At the time of writing, Congress has recently passed new healthcare legislation which will require people to buy health insurance.\footnote{Patient Protection and Affordable Care Act, Pub. L. No. 111-148, 124, Stat. 119 (2010); Health Care and Education Reconciliation Act, Pub. L. No. 111-152, 124 Stat. 1029 (2010).} Furthermore, the federal government has followed the lead of many states in preventing people from agreeing to high deductibles or to not include certain treatments.\footnote{Ibid.}

For some time regulators largely concerned themselves with economic problems such as natural monopoly or with social problems such as pollution, all of which are consistent with the rational theory of choice. Regulatory actions which did not fit the theory, such as price controls, were discarded some 40 years ago (Eisner, 2011).\footnote{27} Some actions opposed by economists have remained either out of sentiment or conservatism. The Food and Drug Administration (FDA) still prevents patients from contracting to buy drugs which are not (or not yet) deemed to be safe and effective, regardless of the patient’s willingness to bear the risk.\footnote{See Federal Food, Drug, and Cosmetic Act of 1938, Pub. L. No. 717, ch. 675, 402(a), 52 Stat. 1046 (1976)} In other cases regulators have become more willing to accept such arguments. The Federal Trade Commission now requires a three day cooling-off period for door to door sales.\footnote{Rule Concerning Cooling-Off Period for Sales Made at Homes or at Certain Other Locations; 16 CFR Part 429.}

Under Cass Sunstein’s White House stewardship, regulators have become more aware of the specific biases they are attempting to correct, and indeed have sought to exploit some of these biases for both social and paternalistic goals. In February 2010, Sunstein summarized some of
FDA plans to help smokers overcome biases which would lead them to overweight short-term enjoyment and underweight long-term health consequences by requiring cigarette manufacturers to place pictures of diseased lungs on cigarette packages, thus making risks more salient. Another proposal will put workers into their employer’s pension plan unless they explicitly opt-out. Sunstein also believes that proposals to make energy costs more visible and salient up front will help reduce costs for the federal government and reduce pollution. The Environmental Protection Agency and Department of Energy’s joint Energy Star program seeks to do something similar for individuals and businesses.

Some of the stances adopted by the court and by policymakers likely reflect a desire to use the law for redistributive purposes. Nevertheless, it is clear that recognition of limits in cognition plays an increasingly significant role in the development of the law. However, progress has been uneven and inconsistent and there is no rigorous theory to determine which paternalist policies are best. Economists can analyze the effects of bankruptcy law or limited liability on national income or growth but the analysis is purely empirical and in the aggregate. For paternalist policies which reduce energy use or reduce smoking, economists can calculate savings in energy or health costs but not say how much the individual recipient of paternalist policies benefits.

Nor do the proponents of placing cognitive analysis in the law agree upon a single approach. While Eisenberg (1995) argues that ex post analysis is and should be a part of the law, Jolls, Sunstein, and Thaler (1998, argue that the law needs to do more to eliminate the biases of judges and juries who give too much weight to ex post analysis.

---

30 Regulatory Change under the Obama Administration, Speech by Sunstein C.R., Feb. 17, 2010
31 Id.
32 Id.
33 Id.
C. The Purpose of This Essay

Despite substantial progress, the nascent field of behavioral law and economics is incomplete. Its practitioners attempt to place real people into economic models. Yet both rational man and the atomistic world of economic models are both simplifying assumptions designed to allow economists to say something useful about the real world (Friedman, 1953). That real people perform differently to rational man in a laboratory is to say nothing about the effectiveness of economic models in the real world. If economics is to say something useful beyond assumptions of rationality then it is not enough to simply extrapolate from the laboratory to the macro-economy. Economics must develop new assumptions about how people interact in the real world with social institutions, markets, and governments.

This essay offers a model of human behavior in the real world. As with the behavioralists, I draw on evidence from psychology and neuroscience to develop more realistic models of individual behavior. However, instead of living in an atomistic society and interacting with one another only through formal markets, I propose that people belong to social groups. Within these groups individuals develop norms of cooperation consistent with those observed by sociologists and anthropologists. I also propose an evolutionary mechanism where individuals can select norms which help them overcome bounded rationality by choosing between groups. I show that this mechanism can sometimes produce better results than the law but also indicate where soft paternalistic laws can help people make better decisions.

35 For example, Jolls, Sunstein and Thaler (1998) claim that once our biases can be understood, they can simply be incorporated into the standard neo-classical model. See also Supra note 2.
II. Conceptions of Rationality

A. Rational Choice

Rational choice describes both an assumption – people behave as they do because it is in their rational interest to do so – and a prediction – people will act according their rational self-interest. As an assumption, rational choice says that if a man chooses to smoke then he understands the health costs of his actions and has decided that he values the experience of smoking over those costs. The economist need only observe the man’s behavior and his budget constraints to know his preferences. Furthermore, the man will always choose to do what is in his own best interests regardless of the complexity of the choice presented to him. If any number of addictive drugs were available to him he could be no worse off because he could choose not to take those drugs.\(^{36}\) Similarly, if an investor is offered a complex array of options she can be no worse off than if she were only offered a simple default choice, and will be better off if any of those options is preferable to the simple default. Indeed, increasing and expanding her choices will always improve her situation or at worse, leave her unaffected (Friedman, 1962, 1980).

The assumption makes sense if the investor understands all the possible options. But the rationality assumption goes beyond this claim. The economist assumes that the woman knows her entire set of preferences over all possible worlds. If a college graduate decides to go to law school the economist will assume she can fully anticipate both her experience in law school and

\(^{36}\) Under such a paradigm the meaning of addiction must change somewhat to accommodate the principle that all choices are rational. See Infra II.B
her job prospects afterwards. She cannot be made better off by restricting her freedom to go to
law school or in making it harder for her to incur debts to pay tuition. Yet she may find that she
does not enjoy the experience at all and come to regret her choice. It is unlikely then that going
to law school could ever have been her true preference; or that an investor who put all of his
saving into some company that then went bankrupt, preferred to lose all of his money.

Whether it is through such major decisions as these or something as mundane as the decision to
try a new restaurant, the process of trial and error is fundamental to understanding human and
economic behavior (Hayek, 1960, 1973); but this process is not at all captured by rational choice.

It is possible that individuals have some preference for experimentation. A man might
frequently visit restaurants of unknown quality because he enjoys the new experience, even
though he does not expect to find many restaurants he would return too. Similarly, he may buy
lottery tickets for the same reason, even though he does not expect to win, because he has a
preference for gambling. He may even feel regret at having lost (or joy if he wins) but continues
to play. In this sense his choices may still reflect his true preferences.

It is less likely (although not entirely impossible) that the man invests all his money in a failed
company out of a preference for gambling. He may, however, have made such a decision
to play. In this sense his choices may still reflect his true preferences.

entirely rationally even if he has no taste for gambling. Rational expectations, a concept first
introduced by Muth (1961) and Lucas (1972), permits economists to introduce risk into the
assumption of rationality. When the man invests he knows the company stock may rise or fall
but the economist assumes he knows the exact likelihood of each event. Arrow (1966) argues
that rather than use this knowledge to calculate the average outcome, the man would imagine

32
himself in every possible situation and evaluate his happiness in each.\textsuperscript{37} He is risk averse because he fears poverty more than he prizes wealth and will demand a greater return to invest in riskier projects. With rational expectations he might invest in some experimental technology and even if the venture fails, his decision can still be rational because he understood the risk and thought it worth the potential payoff.

The economic theory of decision-making under uncertainty replaces the naïve presumption that any act which in hindsight appears to be a mistake must be market failure but does so only by making the equally improbable claim that no such mistakes are genuine examples of error.\textsuperscript{38} The theory is of no help to a regulator who wishes to protect investors from real errors while preserving the essential process of trial and error that produces innovation. Moreover, it is an incomplete theory. If a student regrets her decision to go to law school it is not because the actual experience of law school was unknowable but because it was unknown to her at the time of her decision. She may have some taste for experimentation but it seems more likely she made a mistake. It is quite impossible to tell whether an investor who makes a decision which looks unwise in hindsight, does so on the rational expectation of gain or has likewise made mistakes founded in his own lack of knowledge.

B. Rational Addiction

In general, the law assumes that addiction is not rational or in the individual’s own self-interest.

Many addictive narcotics are illegal and legal drugs are either restricted by prescription or are regulated as with cigarettes and alcohol. Controversially, many economists believe that

\textsuperscript{37} In economic parlance, he will calculate his expected utility (Bernoulli, 1738).
\textsuperscript{38} This is not to imply the standard view of economics rejects all arguments for market failure, only that such mistakes are not a type of failure.
Rational choice is consistent with drug use if the individual knows the health hazards of using drugs and chooses to do so anyway – such as a patient using morphine for pain relief – but Stigler and Becker (1977) argue that even addiction can be rational. They argue that their theory can explain how we can acquire some tastes, such as for fine wine or classical music, without claiming that our tastes have changed.

The authors argue that goods, such as fine wine and cigarettes, are only inputs to some higher commodity they call “euphoria” (a state of happiness which is reached by consuming the good alongside other inputs). Smokers possess “cigarette capital:” a talent for converting smoking cigarettes into enjoyment. A smoker builds his cigarette capital by smoking and in doing so enhances his enjoyment of future cigarettes. The man can be said to have acted rationally by wanting to smoke more once he has been exposed to cigarettes while retaining the assumption that his tastes remain unchanged. The theory can not only explain all forms of addictions and acquired tastes but other behaviors too, such as how advertising can increase demand without shifting consumer’s preferences. If consumers do not know which product they would most like to consume; the product and information about the product both become inputs to the commodity of euphoria.

Stigler and Becker justify using such a complex theory to explain an apparently obvious phenomenon with the claim that any theory that relies on unpredictable and immeasurable tastes will explain almost any form of behavior but will be entirely incapable of predicting which behavior people will ultimately choose. The science of economics seeks to explain how subtle changes in prices and incomes explain different behavior across individuals and across time.

39 For example, see Friedman (1972, 1984), Becker (1987, 2001), Miron and Zweibel (1995)
With changing states that science entirely loses its predictive power and the theory of rationality is once again reduced to a mere assumption that is neither verifiable nor helpful. To save the science Stigler and Becker (1977) propose instead that economists should assume that tastes are constant not only across time but across people.

The idea of constant preferences is not appealing because it is realistic – it is not – but because it is useful. It seems likely that differences in preferences play some part when two people choose to trade, or when some people seem drawn to one profession and others to a different profession entirely. Yet while such a theory undoubtedly has intuitive appeal, it is almost entirely without predictive power. By contrast, if the economist assumes constant preferences then he can answer questions about the impact of changing prices and income with useful and testable claims about the state of the world. A theory of the former type might explain the large scale entry of women into the workforce after the mid-century by reference to shifting cultural mores or changing attitudes toward women. A theory of constant preferences might instead look at factors which change demand for household labor relative to participation in the workforce. The latter theory might be less complete, or even provide a less satisfactory explanation, but in providing a partial explanation it can identify reasons why women have not fully entered the workforce and permit economists to make relevant recommendations to policymakers.

While the authors’ method seems eccentric it is not without merit. Becker and Murphy (1988) use the same framework to explain, and make predictions about, rational cigarette addiction. Becker et al (1991, 1994) test these predictions with some measure of success. Another study

---

Stigler and Becker (1977) at p.76
found that advertising is not merely a wasteful exercise in modifying preferences but rather provides consumers with useful information, increases competition and lowers prices (Benham, 1972). Yet while it is possible to maintain a useful framework where addiction to cigarettes is rational, the authors save only the word and not the implication. In Friedman’s (1962) view rationality means that individuals will always act in their own best interests, and will always benefit from having more choice. Stigler and Becker’s (1977) use of the word does not necessarily imply the same. If information about the product and the product itself are separate inputs then, although the consumer will act rationally in a world of imperfect information, she may also make errors. The consumer might indeed be better off with less choice rather than more, or if she were denied entirely the ability to become rationally addicted.

If Stigler and Becker’s theory is correct then preferences must at least be consistent with one another. If a woman prefers chicken to beef and beef to pork he must also prefer chicken to pork. She may still choose chicken one day and pork the next not because her preferences have changed but because she also enjoys variety. She may treat bacon and eggs in the morning as an entirely different good to bacon and eggs at night and have different preferences for the two. When preferences are so complex, as indeed they are, it becomes near impossible to verify the assumption of rationality. Yet a person may still hold apparently contradictory preferences.

Schelling (1984) suggests the example of woman who asks her doctors to refuse her an epidural while giving birth. She has previously given birth and, despite initially refusing her the epidural, requests pain relief after some interval. Her intransitivity cannot be explained by a lack of knowledge: she knows that birth will be painful and expects to change her mind (or her preferences) at some point. Her differing choices might be explained by treating her initial
choice as cheap talk: she claims she does not want pain relief when doing so is costless and may even afford her some respect, but her true preference is the one she reveals when she suffers a cost from being in pain. Still, Schelling claims, the woman holds the preference to be refused the epidural as superior to the decision she makes only in the heat of the moment to accept pain relief. She may even be angry at doctors after the fact for having given in to her request under pain. In this instance her true preference may not be her revealed preference in that moment. The same might also be true of a woman who throws away a packet of cigarettes to prevent herself from smoking, or a man who refuses a drink on the grounds that he might then want another.

If the woman would prefer to throw away a packet of cigarettes then she may prefer even more to be prevented from buying cigarettes or to have the cost of her purchase increased to discourage the activity. If conflicting choices indicate conflicting preferences then the assumption of rationality can do nothing to distinguish the higher (or true) preference. To determine if, or how, policymakers can make people better off by restricting their choices, economists must make additional assumptions about what those people might prefer.

In the rational choice framework these additional assumptions are materialism and soft egoism. An economist will assume that preferences are stable, that a person prefers greater wealth to less wealth, and that more choice is better than less choice. Yet the person throwing away a packet of cigarettes has clear preference to shed herself of material wealth and the associated choice. She may in addition wish to shed herself of the costs of decision making: the

---

41 Economics does not predict we never give to charity, only that we do so because we treat charity as a good. The implication that we only give for our own benefit rather than out of true altruistic sentiment speaks more to the convenience of that assumption and not the philosophical beliefs of the economists.
need for careful thought or the fear associated with the fear of potential error. It is possible to build a theory of rational choice which includes preferences over internal states such as stress and regret (see e.g., Loomes and Sugden, 1982) but doing so takes economics further from its goal of being a predictive science.

The prediction of rationality has limited value then as a description of individual behavior. Yet the theory may still have value not as a claim about individual behavior but as an abstraction from it. Friedman (1953) and Alchian (1950) argue that rational choice does not need to accurately describe human behavior at the individual level but only to explain human behavior in the aggregate. Human beings are complex and they interact with a complex world. If simple models are empirically valid then those simplifications may permit economists to say something useful about the real world. By contrast, richer models of human behavior may fail to produce useful and testable prediction about economic behavior.

If simplified abstractions can better describe human behavior in the aggregate then accuracy at the individual level may be unnecessary. Yet one can only extrapolate so far from abstraction. The claim of economics, that people are always better off when they are given free choice, follows from the claim that people are rational, not the claim that rationality is useful for explaining behavior in the aggregate. It is worth then, paying some attention not only to the predictions of rationality but also the limits of rationality and by extension the limits of those predictions.
C. Bounded Rationality

Simon (1972) identifies three limitations on rationality. First, people have limited information about the state of the world. Second, there is uncertainty over the current and future states of the world. Third, people have limited computational ability available to understand the complexity of the real world.

The first case corresponds reasonably with the woman who goes to law school without knowing in advance if she will find the experience rewarding. Alternatively she may need to buy a new car but does not know as much as she might about the various features of cars (e.g., safety, comfort, or quality of build). In both cases she may acquire more information by spending her time and money, learning more about law school or cars. As acquiring information is costly it is not rational for the woman to be perfectly informed about each decision she makes even though being better informed would help her reach better decisions. The automaker or law school has some incentive to provide her with information about the positive features of their product and even the negative features of their competitors, but they have little incentive to disclose information that would dissuade her from buying their product. Indeed they may benefit from her ignorance (Akerlof, 1970). Regulators might thus help the woman by mandating the seller disclose greater information or by providing that information as a public service.

Stigler (1961) proposes that a rational individual will treat the search for information as a profit maximizing activity; weighing at the margin the cost of searching for additional information against the benefit of making a better choice. Search models of this variety do not assume a truly ignorant consumer (Spencer, 1972). Typically the woman will look at each car in turn,
discover all she needs to know about the vehicle, and then determine the probability that the
next car she views will be better than best car she has currently seen. She will then either buy
that car or continue searching. In these models, the woman must know the general quality of
the cars she has not seen and how greatly those cars differ from one another, to predict
whether further search will be worth her while. Thus, if the first car she encounters is
considerably better than average she will immediately stop searching without expending any
additional effort. Likewise, if all cars are similar and search is costly she will probably not bother
to look at many cars. But if all the cars she sees are significantly below average she would know
this and continue searching until she finds a suitable vehicle.

A truly ignorant consumer would not know the likely quality of the next car she encounters.
Instead she might reason inductively based on those cars she has already seen. If the first car
she sees is above average she will form wrong conclusions about the average quality of cars and
only after additional and unnecessary searching dispels those illusions will she return to the first
vehicle. Likewise, if she encounters a stream of bad cars she will under-estimate the average
quality of cars and accept a lower quality model than had she known the state of the world.

Even to the truly ignorant consumer, the state of the world is knowable, even if in practice she is
uninformed and may rationally choose to remain that way. Yet the current, and more so future,
state of the world might be unknowable. If a man chooses to buy a new video recorder he
faces incomplete information of the type described above: he must decide which technology is
best suited to his needs. If there are competing standards there may be some benefit to using
the same standard as other consumers, such as VHS and Beta-max, or HD-DVD and Blu-Ray. It is
not enough to have all the information about which standard he prefers – if it were, the
manufacturer of the losing technology would have avoided entering the market in the first place – the man must also predict what choices others will make in the future. The preferences of the entire world’s people are unknowable and so the man must make a decision (which may be the decision to delay his purchase though this too holds costs) in an uncertain world.

The rational man may still know the range of possible states of the world and their likely distribution. In this case he has rational expectations of the future and will make his choices according to what outcomes are likely and his own preference for bearing risk.\textsuperscript{42} He lives in a world of knowable unknowns: how likely it is that he has bought a winning lottery ticket but not whether his ticket has won or how valuable flood insurance is to him without knowing specifically whether his home will flood. Yet there exists an alternative world of unknowable unknowns where not only is the future state of the world unknowable but so too are the range and likelihoods of possible alternatives.

It is hard to imagine a person having complete knowledge over states of the world that do not as yet exist. What probability would a man in 1700 have assigned to the oncoming industrial revolution and its consequences for society? Predicting future technology requires not only the logical mind of rational man but an imagination with which few are blessed. For the large majority, uncertainty is a reality with which they must cope. Simon (1957) argues that we purposely enter into incomplete contracts, such as those between and employer and employee, to cope with uncertainty. Employers are unsure what tasks they may require of employees and prefer vague contracts which leave the manager discretion to direct the employee’s day to day activities as needs arise. Indeed the difficulty, if not impossibility, of writing complete contracts

\textsuperscript{42} See Infra II.B
under uncertainty may be one of the main reasons for the existence of firms (Williamson, 1975, 1985).

Beyond these external challenges of incomplete information and uncertainty, individuals face internal limits of cognition. Even if the entire state of world, both current and future, is knowable at no cost, we may still be at a loss to comprehend its complexity. Simon (1972) suggests the example of chess. Though complex, the game of chess is far simpler than the real world. Each decision is constrained according to the rules of chess and the game must eventually end. Indeed there are a finite number of possible chess games that can exist (that number is in the region of $10^{120}$). By the simple principles of game theory the perfect game of chess can be deduced. Nevertheless, the most expert chess player or supercomputer cannot calculate this perfect game and instead relies on a mixture of logic and intuition (Ibid.).

Whether the individual faces informational deficits, uncertainty, or cognitional limitations, the effect is largely the same. The choices made by the individual are not those he would make in an ideal world of perfect information or unlimited calculative ability. Simon (1972) argues the solution is to redefine rationality somewhere between the economic and psychological concepts. Economics relies on substantive rationality where our actions and our preferences are one and the same. By contrast the psychological approach accepts a far broader range of behaviors as rational, including the practice of self-deception (Ibid). Although the two models are in some senses polar opposites, neither leaves much room for a category of irrational decisions: whatever a person does or thinks can be explained as rational. Somewhere in between these two models Simon proposes procedural rationality, where rationality is assessed according to our methods of choosing. It is no longer necessary that our actions are the ones
which serve us best, only that the way we choose our actions is that method which will serve us best over time.

Procedural rationality is closer to the psychological concept than to rational expectations which allows nothing for complexity or an inability to process information. Once again, redefinition appears to save the word rationality but fails to save the conclusion that only flows from the narrower economic meaning of the word: that each decision must be in the best perceived interest of the individual.

D. Biases and Heuristics

It is not enough for a lawmaker to know only that citizens are boundedly rational. If he also wants to help them he must also know how their thinking flawed (Rizzo, 2009). We may know that the woman has changed her mind about taking an epidural but to help her we must know which choice is her true preference. By contrast, if individuals are prone to make common mistakes of false logic that can be easily identified against positive criteria rather than some normative standard, then it may be possible to correct those mistakes (Arrow, 1982; Diamond, 1977; Grether, 1978; Kahneman, 2003; Kunreuther, 1979). A significant literature within cognitive psychology claims to have identified a large class of such errors, or cognitive biases (Stanovich and West, 2003).\textsuperscript{43} While these biases are often presented as being finite and knowable, there is no exhaustive understanding of bias and those that are known are too numerous to describe in full and only some are directly relevant to economics and the law. Kahneman and Tversky (1973) identify three judgment heuristics that describe how people

\textsuperscript{43} For a literature review see Baron, 1998, 2000; Dawes, 1998; Evans, 1989; Evans and Over, 1996; Kahneman and Tversky, 1972, 1984, 2000; Kahneman, Slovic and Tversky, 1982; Nickerson, 1998; Shafir and Tversky, 1995; Stanovich, 1999; Tversky, 1996.
make complex decisions under uncertainty. These heuristics are frequently successful but occasionally lead to “severe and systematic errors” (Ibid). The authors initially identified three distinct heuristics which subjects used to make risky decisions: representativeness, availability, and anchoring.

1. Representativeness

Representativeness is an heuristic whereby people classify objects or events as part of overall systems. Once the object or event has been classified the mind can proceed according to previous similar experiences. The mind does not use a strict system of categorization but rather a series of rapid mental leaps. For example, a chicken might be a representative animal but not a representative bird, although chickens belong to that sub-category (Tversky and Kahneman, 1983).

The heuristic ignores other information, such as prior probabilities. In one experiment Kahneman and Tversky (1973) told subjects that a lawyers and doctors were attending a conference. Some subjects heard that 70% were lawyers and 30% were doctors while the others heard the reverse. When the experimenters then described one attendee and asked subjects to determine if he was more likely a lawyer or an engineer, both groups gave the same

---

44 Kahneman (2003) has since sought to simplify this claim and explain these limitations within the framework of the automatic and deliberative system.
45 In calling these limitations “bounded rationality” Kahneman is using the term more broadly than Simon. I use the narrow sense of the phrase.
46 The authors use the term uncertainty rather than risk but their definition fits that of risk in the economists’ sense of the word.
47 Some scholars (e.g., Sloman, 1996) suggest representativeness may stem from associative processing by the automatic system (see Infra II.F. for a summary of the two system of processing) but this claim about the nature of the automatic system remains controversial (Evans, 2008).
answer. Even when the experimenter used a completely neutral description, the subjects ignored the initial proportions and focused entirely on the unhelpful description.

Another experiment asked people to assess the probability of a flood in a California and the probability of an earthquake causing a flood in California (Tversky and Kahneman, 1983). Subjects thought the flood caused by an earthquake would be more likely than the flood alone although this is evidently impossible. An earthquake-related flood is more representative of a natural disaster in California than a flood alone. If California homeowners were to use this heuristic they might conceivably underestimate the need for flood insurance.

Sometimes a successful company seems to be undervalued because of one poorly performing division even if it is a relatively small part of the company’s business. Based on a study where people were asked to value two sets of trading cards, this may be the result of a bias (List, 2003). One set included ten mint condition cards while another included the same cards and three poor quality cards. Subjects placed a higher value on the smaller set of cards than on the larger set, even though the larger set included all the cards in the smaller set (Ibid.).

In economics, portfolio theory shows that investors can avoid risk by diversifying their portfolio between many uncorrelated risks, but subjects in experiments often fail to appreciate the effect of combing independent risks. Psychologists told subjects that two hospitals, one large and one small, recorded the number of days when more than 60% of babies born were male. A large hospital where many babies are born is more likely to have a roughly equal number of boys and girls born on a given day than a small hospital. Yet most subjects predicted that large and

---

48 For example, if only one baby is born on a given day then 100% of babies born that day will be of the same sex.
small hospitals would have the same number of days while the remained of subjects were evenly split between the large and the small hospital.

Subjects are also likely to see patterns where none exist (Haselton and Nettle, 2006). When many stocks are up or down on a particular day, or when there several up or down days in a row, people interpret these as patterns rather than possibly random events. Even gamblers, where there is no underlying narrative to explain random behavior, believe in lucky streaks. The fact that stocks are not exactly random may be partially explained by investors misreading underlying randomness and reinforcing stock market trends.  

Coin tosses, with a fair coin, are indisputably random, but subjects believed that after a run of heads the next toss was more likely to be tails (Kahneman and Tversky, 1973). They also believed that the sequence H-T-H-T-H was more likely than H-H-H-T-T although both sequences are equally probable. Subjects understand that the coin tosses are random but they look for sequences which appear more representative of randomness. Here subjects infer a negative correlation rather than a positive correlation seen with stocks. Thus the heuristic error is not toward seeing one type of correlation but of seeing patterns were none exist.

2. Availability

In Kahneman and Tversky’s second set of biases, known broadly as availability, people misjudge probabilities based on how easily examples can be brought to mind. For example, a driver who sees a graphic car wreck temporarily overestimates the likelihood of being involved in a wreck (Kahneman, Slovic, and Tversky, 1982). Had the same driver merely read about the accident she

Stocks are not ordered but nor are they random. Extreme stock market events such as crashes and spikes occur much more frequently than if markets behaved randomly (Mandelbrot, 1967; Lo, 1999).
would likely have not made the same error. Reading about accident statistics, though these would be more informative, would be similarly ineffective, and the driver may even discard this information in favor of the emotively powerful, but purely anecdotal, wreck. Salience also matters: if a friend is involved in an accident you may drive more carefully but if a stranger has an accident your behavior remains unchanged, even though the distinction has no effect on your own likelihood of being in an accident.

People can also misjudge probabilities because of the ease of searching a set. It is harder to think of examples of words which have an ‘r’ as the third letter than words with an ‘r’ as the first letter and so people will wrongly judge the latter to be more common (Kahneman and Tversky, 1973). Subjects even report that more words end in ‘ing’ than have ‘i’ as the third to last letter, although the latter must occur more often. Similarly, people miscalculate probabilities based on how imaginable are examples. More people are killed each year by swimming pools than by guns but if it is easier to imagine a fatal accident with a gun than with a swimming pool then people will believe that there are more accidental deaths from guns than from swimming pools (Kahneman, Slovic, and Tversky, 1982; Levitt and Dubner, 2005).

3. Anchoring

The third set of biases in this category is known as anchoring. When asked to estimate a quantity people pick a starting point (which may be arbitrary) and adjust depending on whether they think the number they are being asked to estimate is higher or lower than the starting point; but that adjustment is usually insufficient. A resident of New York trying to estimate the population of Boston would start with the population of her own city and adjust down, likely
over-estimating the population of Boston. A resident of Bangor, Maine would adjust upward for his own town and likely under-estimate the population of Boston.

In one experiment psychologists asked subjects to guess how many countries there are in Africa but before the subject guessed, they watched the psychologist spin a wheel of fortune. The subject guessed first if the number was higher or lower than the random number generated by the wheel, and then what the actual number was. Although subjects who saw low numbers adjusted upwards and subjects who saw high numbers adjusted downwards, the subjects who saw high numbers still made higher average guesses than those who saw lower numbers. This result held even when psychologists paid subjects for accuracy.

Subjects also overestimate the likelihood of conjunctive events (e.g. tossing a coin ten times and getting heads each time) and underestimate the likelihood of disjunctive events (tossing a coin ten times and getting at least one head). Psychologists gave subjects the opportunity to bet on either a simple event (drawing a red marble from a bag containing half red marbles and half white marbles); a conjunctive event (drawing a red marble seven times in a row from a bag with 90 percent red marbles); and a disjunctive event (drawing at least one red marble out of seven from a bag with just 10 percent red marbles). The conjunctive even had the greatest probability of success (52%) and the disjunctive event, the lowest (48%). Yet subjects preferred the bets in

---

50 Specifically, the psychologist asked for the number of African countries recognized by the United Nations.
51 Kahneman, Slovich and Tversky (1982) do not report when the experiment was conducted but both subject groups reported fewer countries in Africa than exist at the time this article was written. If subjects shown the higher number still underestimated the number of countries then they must have over-adjusted rather than under-adjusted. It is possible that there are other biases as well as anchoring (for example, the availability of examples of African countries) to cause both groups to underestimate the number of countries. In another experiment subjects were asked to calculate the product of $1 \times 2 \times 3 \times 4 \times 5 \times 6 \times 7 \times 8$ and $8 \times 7 \times 6 \times 5 \times 4 \times 3 \times 2 \times 1$. Estimates were significantly higher on average when numbers were presented in descending order (2,250 versus 512) but still substantially lower than the correct answer (40,0320).
reverse order to their likelihood of success. Again the subjects failed to sufficiently adjust from
the starting point (the likelihood of drawing one red marble).

E. The Social Instinct

In addition to the three judgment biases identified by Kahneman and Tversky, psychologists
have uncovered many others. Many are irrelevant to law and economics while yet others do
not seem to be biases at all in the economic sense of the word. For example, people will refuse
to eat chocolate shaped like a cockroach or fudge shaped as dog feces, despite liking the taste of
both foods (Haselton and Nettle, 2006). Though this may in some sense be irrational, it is not
necessarily so in the economic meaning of the word. It is reasonable that people should hold
aesthetic preferences for foods that do not look like feces; and certainty neither economists nor
psychologists would recommend that lawmakers would make anyone better off by forcing him
to eat something so unappealing. Yet there is at least some evidence from evolutionary
psychology that this is indeed a bias and not representative of the individual’s true preferences
(Ibid.).

Likewise the desire for fairness: a number of economic experiments suggest that people have a
strong concern for fairness (Kahneman, Knetsch, and Thaler, 1986b). Whether the concern is
bias or preference it is so strongly and consistently held to the extent that the concern for
fairness overrides one of the key axiomatic assumptions of economics: the preference for
individual gain. Thus, even if the assumption of rationality stands, economic models which
assume only a preference for individual gain, and not for fairness, may draw consistently wrong
conclusions. Coase theorem, as restarted by Stigler, (Coase, 1960; Stigler, 1966) claims that so
long as property rights are properly defined and the cost bargaining is low, parties will always
bargain to an efficient solution; but if parties are concerned more with fairness than personal gain then such beneficial bargains may not be reached. Indeed an empirical study of bargaining after courts defined property rights found that such bargaining was far rarer than economic theory predicts (Kahneman, Knetsch, and Thaler, 1990).52

Economists have tested the competing assumptions of a preference for self-advancement and fairness in a series of experiments. In one of these experiments, the ultimatum game, two players who have never met and will not play one another again, must split a sum of money (Thaler, 1998). The first player is given ten dollars and told he may offer all, none, or some of the money to a second player. The second player then receives an ultimatum with no option to negotiate: she must accept the offer or neither will get any money. If both players were concerned only with personal gain then the second player would accept any rather than get nothing. The first player knows this and would offer the smallest possible sum of money; usually one dollar. With real people the first player will often offer more than the smallest possible amount while the second player will reject small offers at a cost to herself.

In a variation of the ultimatum game, called the dictator game. As before he may choose to give all, none, or some of the money to the second player, but the second player may not reject the money. Again the first player chooses to give some portion of the wealth though this outcome may vary if the first player is allowed to earn the money; in both instances suggesting a concern for fairness. In one version of this experiment, the experimenters also varied the social distance between the players by increasing the anonymity of the first player (Hoffman, McCabe, and Smith, 1996). When the social distance between players is small players seem to have the same

52 Though there are explanations other than spite. Individuals aloes suffer from an endowment effect which may impede bargaining (see Thaler, 1991)
desire for fairness as in the ultimatum game, even though the players were unlikely to meet one another again. As the experimenters increased social distance by making it harder for other players and experimenters to know how much each person had given, the apparent desire for fairness diminished though it did not entirely disappear. If the desire for fairness were born entirely out of preference then the protection of anonymity should have had no effect on the first player’s willingness to give. Thus it seems that there is indeed an element of bias involved, separating the player’s true preferences from his actions.

The authors attributed this finding to individuals bringing behaviors from the real world into an experimental one (Ibid). It may be that our biases are not entirely static but are affected by social norms and cultural conditioning. Yamagishi, Jin, and Kiyonari (1999) found that the Japanese, who may have stronger norms in favor of cooperation, are indeed more likely than Westerners to cooperate in experiments even though the experiment provides a supposedly neutral environment. Where Western psychologists once believed there was a tendency to overstate certain personality traits such as self-reliance, more recent studies have found that Japanese understate these traits but overstate collectivists traits.53

Economists have also found evidence that individuals fail to completely adjust to a changed environment (Fisman and Miguel, 2007). UN diplomats in New York used to be completely immune from parking tickets issued by the city. As economists would expect, diplomats are more likely to incur parking tickets when they do not have to pay the cost and, as behavioralists would expect, diplomats do not simply park wherever they like even though the cost of doing so is zero. Diplomats incurring parking tickets at a rate linked to corruption in their own countries.

Diplomats from countries where officials are held accountable were far less likely to get parking tickets than diplomats from corrupt countries. In other words, the level of bias (deviation from the economic model) was determined by the behavior diplomats learned at home.

In addition to static biases the evidence suggests the existence of pro-social biases. These are sufficient people will even replace the judgment of others with their own. Subjects will not only claim that a blue line is actually green (after secretly planted confederates of the experimenter purposely made the same error) but will continue to maintain they were correct while being quizzed by the experimenter (Thaler and Sunstein, 2008).

The bias towards social behavior may induce error and may be linked to increased rates of smoking, or obesity (Ibid.). When tied to the ability to punish, as in the ultimatum game, the desire for fairness can lead to the destruction of wealth. Likewise, the inability of parties to agree to mutually beneficial contracts (albeit contracts which seem unfair to at least one side) may be destructive. Kahneman, Knetsch and Thaler (1991) suggests that the unwillingness to agree to mutually beneficial contracts may be a source of unemployment; and that lawmakers might intervene by taking advantage of another bias, the tendency to focus on nominal rather than inflation-adjusted prices, and using inflation to cure unemployment.

F. Prospect Theory

Some biases are consistent across cultures. Among these, over-optimism has been clearly identified in laboratories (De Bondt and Thaler, 1995), in stock markets (Berg and Lien, 2005), and among entrepreneurs (Bernado and Welch, 2005). Humans are consistently over-optimistic about outcomes: even when a person has accurate perceptions about another person’s risk of becoming a victim of crime they will underestimate their own probability. In addition to making
unsound investments this may lead individuals to mistakenly enter into contracts (such as agreeing to excessive liquidated damages) in the belief they are more likely to meet their obligations than they really are (Eisenberg, 1995).

Yet we are also loss averse causing us to avoid risky situations. Not only do we hold these contradictory biases but we seem to hold them simultaneously. When Kahneman and Tversky (1979) offered subjects a risky bet with a high payoff or a certain bet with a lower payoff the subjects chose the certain bet. The psychologist then rephrased the bet so the subject started with the money they would have made from the risky bet. Now subjects had to choose between a certain loss or some risk of losing everything. The new bet was exactly the same as the first bet but subjects reversed their choices. Based on this result the authors argued that people tend to overweight guaranteed outcomes making them risk averse with respect to gains but apparently risk loving with respect to losses.

Simply changing the framing of the question elicits a different response, and an apparently different set of preferences. If people behave this way outside the laboratory we should expect to observe apparently irrational market behavior: people would avoid risky investments with high payoffs but Kahneman and Tversky (1979) identify two aspects to this bias. The first is loss aversion: people treat a forgone gain differently to a loss. The second is that people define decisions based on gains and losses rather than final outcomes, a theory originally developed by Markowitz (1952). Both these biases violate expected utility theory.

See Kahneman and Tversky (1979), p.288, Table 1: when subjects were offered a choice between a certain $3,000 and an 80% chance of $4,000 (which has an expected value of $3,200) they were more likely to choose the $3,000.
To create a theory of decision making which can both predict behavior and accommodate these biases, Kahneman and Tversky (1979) developed prospect theory. Individuals process gambles, or prospects, in two stages. In the first stage they simplify the problem. Whereas economists assume people imagine themselves in every possible outcome, Kahneman and Tversky assume people just look at gains and losses. Individuals will simply ignore very small probabilities although those small probabilities may significantly affect outcomes. They will also try to isolate the prospect as much as possible. Although this process of simplification may be procedurally rational for decision-making under cognitive limitations it will nevertheless induce inconsistencies which violate expected utility theory.

Depending on how individuals frame the bet in their own minds, they may reach different conclusions for the same problem. If a doctor tells a patient that an operation has a 90% chance of success the patient is more likely to opt for surgery than if he is told the operation has a 10% chance of failure. Read, Loewestein and Rabin (1999) show that it matters how narrowly we bracket problems. If a smoker narrowly brackets his decision to smoke he may ignore the small increased risk of lung cancer but if he uses a broad bracket and evaluates his decision to become a smoker against the total risk of lung cancer he may reach a different conclusion.\textsuperscript{55}

After simplifying prospects, Kahneman and Tversky assume that individuals evaluate those prospects by assigning decision weights. These weights might, in unusual circumstances, be the same as the probabilities used by economists but usually will not be. Although the authors are

\textsuperscript{55} When we break decisions down in this manner we may make other errors. See Thaler (1985, 1999)
not explicit about how people will pick decision weights, they show that such adjustments are necessary to explain the biases found by psychologists.

Despite seeking to replace expected utility theory, prospect theory retains the same basic form: each outcome has two distinct numbers associated with it. The first is the probability which Kahneman and Tversky replace with decision weights, and the second is utility over expected outcomes, which the authors replace with utility over gains or losses.

G. The Role of Regret

Several scholars have offered alternative theories to explain the anomalies identified by Kahneman and Tversky (see e.g., Allais, 1953; Machina, 1982; Loomes and Sugden, 1982). One of these theories, regret theory, argues that all anomalies explained by prospect theory may also be explained by a simpler modification to expected utility theory: when people make decisions they take into account their expected feelings of joy and regret over the outcome. If a woman receives a $1,000 rebate on her taxes she experiences a benefit from that income. The authors call this “choiceless utility” because her decisions played no part in her receiving that money. Had she made the money from an investment she would still experience that choiceless utility but she would also experience joy at having made a good decision. Likewise she would experience only the loss of the benefit if her taxes were increased but she would experience regret if she lost that money on a poor investment. Loomes and Sugden (1982) argue that when she makes her decisions she will not only take into account her preferences over real outcomes but also her preferences over emotional states; minimizing feelings of regret and maximizing feelings of joy alongside those other benefits.
Regret theory is able to predict most of the anomalies also explained by prospect theory while retaining the key aspects of expected utility theory. Regret theory also explains why people gamble but also purchase insurance, a phenomenon prospect theory cannot explain. Though the theory may save the rational choice model of economics by including preferences of feelings of joy and regret, like other explanations which rely on internal states, one can no longer conclude that individual choices preserve the individual’s best interest. A policymaker may intervene on behalf of the individual to save her the worry of potential regret and make the decision she would have otherwise preferred to make.

Several scholars, including Kahneman (2003), and Thaler (1980), support the claim that economics must consider these internal states in making predictions, and recommendations to policymakers. Neuroscience, however, offers only partial support for this claim. In the standard model of neuroscience the brain releases a chemical agent, dopamine, in response to positive experiences, and reduces the background level of dopamine in response to negative experiences (Caplin and Dean, 2008). Scientists were able to trigger dopamine releases in animals by giving subjects food, drink, and sex, and in humans with abstract benefits such as money (Ibid.). But, while dopamine appears to be linked to the experience of pleasure, including addition, it is not a simple expression of the economic theory of utility.

Studies by Schultz, Apicella, and Ljungberg (1993) and Mirenowicz and Schultz (1994) found that the brain does not release dopamine in response to the receipt of benefits but the expectation of benefits. The brain releases dopamine when the individual receives an unexpected benefit
such as food or money (Ibid.). When the individual expects the reward the brain releases dopamine at the moment of anticipation rather than at the moment the benefit is actually experienced; so a woman who wins the lottery experiences a dopamine spike when she learns that she has won but not when the check arrives in the mail. If the individual wrongly predicts a future benefit the brain reduces dopamine when the error is discovered; so a woman who wins experiences reduced dopamine levels when she discovers there were many other winners. When she makes a risky bet she experiences a dopamine increase in proportion to the expectation of benefit; that is, some increase but less than the increase from a sure benefit. The brain then adjusts dopamine levels depending on whether the bet pays off or not.

Our current understanding of the dopamine system suggests that economists should at least exercise caution before including internal states of joy and regret into their models (Caplin and Dean, 2008). There is no evidence to suggest that humans differently experience joy or regret at situations under their control than at those over which they have no control. To the contrary, humans are more prone to over-estimate their own degree of control over a situation (Hassleton and Nettle, 2006). Further, a person may feel regret at one moment but only because she previously over-estimated the benefit from her actions. Had she properly anticipated the benefit, or even been surprised completely, her overall experience would only change in the timing of her feelings.

Neuroscientists have concluded on this evidence that dopamine does not represent the economic concept of utility but is one a part of a learning system that helps people align their

---

56 The neuroscience literature uses the word reward, rather than benefit, but in a somewhat different sense to either economics or plain English. I use benefit here to avoid confusion.

57 In fact economics suggests she should experience utility only from consuming the items she can purchase with her winnings.
behavior with true preferences (Montague and Berns, 2002; Bayer and Glimcher, 2005; Bayer, Lau, and Glimcher, 2007; O’Doherty et al., 2003 2004). The dopamine learning system, and the feelings of joy and regret it creates, helps individuals to better align their actions with their true preferences by reinforcing behaviors which help them achieve their end goals and by correcting errors. Simply mitigating feelings of regret does not necessarily help individuals better achieve their true objectives. Further, by intervening in choice, even when the individual is in error, policymakers risk interfering with this learning system and potentially inducing new errors.

H. The Two Systems of Reasoning

Many modern psychological models ascribe decision-making to two separate but overlapping systems. The first of these is an automatic and unconscious system while the second is a conscious and deliberative system (Evans, 2008). While there is much agreement on the existence of dual systems there is far less agreement on how each system operates or on how the two systems interact to create a single action. Sloman (1996) reviews several competing dual systems models of decision-making while Evans (2008) observes that these major models fall into at least two separate, and partially incompatible, categories.

Neuroscience lends support to the principle of dual systems but the science is not yet sufficiently advanced to describe how each system functions. Psychological models provide

---

58 In fact there is substantial evidence that drug addiction operates by overriding the dopamine learning system and directly stimulating feelings of joy (see e.g., Chiara, 1999).

59 There are no two names that fully capture the properties that various studies have ascribed to the two systems and scholars in the field have used a number of different name pairings (Kahneman and Frederick, 2002; Evans, 2008). The least controversial names must be System 1 and System 2 which are also the least descriptive. I believe the slightly more friendly titles of automatic and deliberative which I have used here, capture at least some of the properties that are generally ascribed to each system and none of those properties that are controversial (Ibid.).

60 For a general review see Loewenstein and O'Donaghue (2004) and Ashraf et al. (2005)
this detail but only by observing the human mind operate as a whole. Just as economists
construct models of social behavior in the aggregate without first constructing a realistic model
of the human building blocks of that society, so too the psychologist constructs models of the
human mind in the aggregate without first understanding the neurological building blocks of
that mind. The behavioral approach may thus be opposed, or justified, with the same
arguments that apply to rational choice in economics. Without being able to fully understand
the human mind we can still say something useful about human behavior. Indeed, despite the
differences between the psychologists’ various models of decision-making there are also enough
significant similarities to produce a basic model of human behavior that is useful to
policymakers (Evans, 2003). In addition to the support neuroscience lends to the principle of a
dual systems there is also significant empirical support for ascribing certain distinct properties to
each system without fully understanding how that system works (Sloman, 1996).

The automatic system is unconscious, rapid, and high capacity; using heuristics to make snap
judgments (Sloman, 1996; Evans, 2008). Meanwhile the deliberative system is slower and more
logical, relying on general rules to determine solutions to problems (Ibid.). Both systems are
procedurally rational, using some form of reasoning to produce an answer. Neither system
exhibits the substantive rationality of economic man; but nor is either system irrational. The
deliberative system, while closer to the economic meaning of rationality, is boundedly rational
and faces constraints of limited information, uncertainty, and limited cognitive capacity. Indeed
the deliberative system is unable to process as many decisions as the automatic system or to
produce those decisions as quickly (Ibid.). Further, Comsides (1989) notes that we must often
make decisions on subjects that go far beyond our cognitive capacities.
Although the automatic system of reasoning is quick to make decisions it has limitations beyond those of bounded rationality. It is impulsive and tied to the emotional system (Smith and DeCoste, 2000; Evans, 2008). Rather than rely upon an abbreviated deliberative process the automatic system uses heuristics to reach split-second judgments. In one conception the automatic system does this by categorizing problems and applying solutions based on past experiences. By deviating from the deliberative process the automatic system can generate biases, hold prejudices that are not found in the deliberative system (Haselton and Nettle, 2006), and make logical errors (see e.g., Kahneman, 2003).

Despite the capacity for error, the automatic system is neither irrational nor rational only in that broad sense which includes self-deception. Rather it is a procedurally rational response to a world of complexity where it is sometimes necessary to make snap judgments. An experienced driver who can rely on the automatic system may be safer than the inexperienced driver whose deliberative system is distracted by the many complex operations of managing an automobile and fails to observe some hazard further ahead. Moreover, the driver who relies upon the deliberative system to decide whether to brake or swerve in an emergency may find herself in an accident before happening upon a solution. In such a case even a biased heuristic such as always braking, although the rule would inevitably lead to errors, would be preferable to contemplation.

Some heuristics arise from our genetic past (see e.g., Haselton, Nettle, and Andrews, 2005). We have no need to experience falling to be afraid of heights and indeed young infants exhibit this

---

61 For example, Loewenstein and O’Donoghue (2004) use such a model. See Evans (2008) for a more detailed analysis of the types of division identified.
62 Although it will of course matter where the experienced driver chooses to apply his attention.
fear. For the automatic system to be useful in activities such as driving it cannot be entirely
genetic; it must also learn from past experiences and apply those lessons to current problems.
Experience may lead a driver to avert an accident or a nurse to detect the subtle signs of heart
failure using only the automatic system (Kahneman, 2003). This experiential form of learning
may be triggered by an emotional event such as seeing a graphic car wreck at the side of the
road but reading about accident statistics would not trigger the same emotional response.
However, the automatic system is not limited to experiential learning (see, e.g., Strack and
Deutsch, 2004; Evans, 2008). The automatic system can reason solutions to novel problems and
can also learn from the deliberative system (Ibid). Thus a person who learns about the dangers
of cancer may associate smoking with harm, and may even experience a visceral aversion to the
smell of cigarettes, without any deliberation.

The two systems do not work over entirely separate domains but may set about solving
problems together, returning separate answers. When a rapid response is required the
deliberative system may return a response too slowly, leaving the automatic system in charge
(Logan, 1988; Smith and DeCoster, 2000). Given time to contemplate, the deliberative system
has the power to override the automatic system (Ibid). A driver may set off on a Sunday and,
unconsciously, begin to drive to her office. The deliberative system then corrects this error and
turns her about to her preferred destination. Giving her extra time to consider, or greater
incentives to pick the right answer will increase her likelihood of avoiding the error while
distraction will increase her chance of error (Metcalf and Mischel, 1999; Smith and Walker,
1993). But this act of conscious thinking, and overriding the automatic system, takes a
deliberate effort or expenditure of willpower.
Emotional states can also influence cognition (Metcalfe and Mischel, 1999). A man who is offered a cigarette in a “cool” state uses his deliberative system and declines the offer, but in an emotionally charged “hot” state uses the automatic system and chooses to smoke (Ibid). When he returns later to a cool state he may regret the decision (Ibid). To overcome his desire to smoke requires a conscious mental effort; an exertion of willpower. He will be less able to exert this willpower if he is tired or has recently expended mental effort on some other task (which may be another effort of willpower or an unrelated task such as a math problem).

Not smoking consumes willpower but throwing away the cigarettes in a cool state conserves that willpower. Thus the deliberative system can override the automatic system not only by mental effort but by the conscious decision to impose constraints on oneself. The man may also ask a friend to restrain him or, a policymaker to impose constraints on his ability to buy cigarettes in the first place. But it is unclear what the policymaker should do if he wants to help the man.

In Schelling’s (1984) example, a woman asks her doctor to deny her an epidural and then changes her mind in what is presumably the hot state of giving birth. By stating contradictory preferences she has breached the standards of substantive rationality. Yet in the dual systems model does not require that her preferences must have changed to explain this inconsistency. Rather, she has calculated the solution to her problem in two different ways and arrived at two different solutions. The doctor must somehow determine which of these stated preferences is the woman’s true preference.

---

63 Kahnemann (2003) does say that the two systems have different preferences but it is not clear why this should be. It may be that he is referring to revealed preferences (i.e. choices) rather than the individual’s true preferences.
Several scholars have argued that the decisions of the deliberative system should be taken as her true preferences (e.g., Metcalfe and Mischel, 1999; Sunstein and Thaler, 2003). Yet there is ample evidence that patients do not accurately remember the pain experienced during a medical procedure (Schwarz, 1999; Gorin and Stone, 2001). The woman presumably knows the pain she currently endures and thus her request for pain relief in the moment would seem to more accurately reflect her true preferences.

While the automatic system may be less like economic man it is also the more specialized of the two (Barkow, Cosmides, and Tooby, 1995; Gigerenzer and Goldstein, 1996). The deliberative system meanwhile is still boundedly rational, facing limitations of computing power and knowledge of the present and future. When a novice attempts to speak a foreign language he may use the deliberative system but the result is far from that produced by the specialized linguistic system. Nurses rely on the automatic system to make the best decisions (Kahneman, 2003). Over certain situations then, the automatic system may return a more accurate account of true preferences but neither system can claim to fully represent the individual’s true preferences in all circumstances. The systems often work together and on occasion both systems may deliver inaccuracies regardless of whether they are in agreement or not (Evans, 2008).

An intuitive approach suggests that a man errs by smoking rather than by exerting willpower. Indeed we might feel that this is what the man should do. We know that smoking is unhealthy and that it has addictive properties. We may imagine that a rational man would choose not to smoke but this misconceives the economic concept of rationality. The utility maximization principle embodies a hedonistic principle: the man will do whatever he believes will best serve
his private preferences and no social or moral considerations act upon him except as a modification to his preference function. Leaving aside for now whether such behavior is desirable, if policymakers are to interfere in the man’s choices in the name of advancing his own interests then they must pay heed to the his private preferences and not some normative external set of values.

I. Better than Rational

The heuristics and biases approach, when coupled with neuroscience and our understanding of the two systems of reasoning, describes the boundaries of human reasoning. Yet while psychology has recorded a significant number of heuristics and their capacity to introduce error, this approach only explains that heuristics do reduce the mental costs of decision-making; not why humans should use any particular heuristic (Buss, 2005). To fill that gap, evolutionary psychologist seek to understand why humans think the way they do.

Where cognitive psychology describes a limited and error-prone mind, evolutionary psychology has found a highly specialized mind that is “better than rational” when it comes to solving complex problems (Cosmides and Tooby, 1994; Goldstein and Girzerenger, 1996). Heuristics are not only quicker than rational deduction but often better over the specific domain of problems for which they are evolved. Goldstein and Girzerenger (1999) found that when subjects used simple heuristics their answers were equally or more accurate than answers by subjects who relied on deductive reasoning. Germans were more likely than their American counterparts, who presumably know more of U.S. geography, to correctly answer that San Diego has a larger population than San Antonio. While the Americans had more information that should have

---

64 Unless constraints are imposed on him externally.
helped them better determine the correct answer, the Germans were able to rely on the availability heuristic (they had only heard of the larger city of San Diego), which in this case delivered the correct response (Ibid.).

Whether or not heuristics produce better or worse responses depends on the question asked by the psychologist. Subjects perform much as economists predict when confronted with an explicit market exchange scenario (Kahneman, Knetsch and Thaler, 1990). The behavioralist critique is that such scenarios are unrealistic but if the evolutionist approach is correct, it is the behavioralists questions that fail to capture the challenges of the real world. Consequently, the biases experimenters observe may simply be misapplied heuristics that fail only by the psychologist’s normative standard, or artifacts of a mind evolved for some other purpose (Haselton, Nettle, and Andrews, 2005). Biases need not necessarily be errors so much as modes of thinking that have evolved because of the benefit they confer to humans.

When one type of error is more costly than another, the mind has evolved to err on the side of the least costly error (Kahneman, Knetsch and Thaler, 1990, Haselton and Nettle, 2006). A smoke alarm generates false alarms more often than it ignores real fires because the cost of the latter is so much greater than that of the former. Humans have an aversion to cockroach-shaped chocolate because the cost of accidentally consuming cockroaches far exceeds the cost of missing a chance to eat chocolate. This evolved behavior is, in part, consistent with rationally predicted behavior.

Nevertheless, subjects do not employ heuristics randomly. In another experiment, when the reason for subjects having heard of a city was unrelated to the city’s size (e.g. Chernobyl) they did not use the availability heuristic (Goldstein and Gigerenzer, 2002). Subjects were asked to exchange tokens which had been assigned some arbitrary value (different to each subject) by the experimenter. Market exchange effectively revealed the subject’s private valuation of the token.
If economic man is confronted with a cockroach-shaped object which, with some probability, is either chocolate or a cockroach, he would take into account the relative costs and benefits of error. In the economists’ models the probability and cost of error are separate variables in the man’s calculus. The evolved mind, by contrast, does not take such a mathematical approach but instead solves the problem by holding biases over beliefs. The difference lies not in the man’s eventual action, but in how he reaches his conclusion. That holding wrong beliefs may help him to an action which achieves his end goal is undoubtedly a subject of interest to students of the mind but is of little importance to lawmakers.

Whether or not heuristics are an effective shortcut or a source of error depends upon our environment. Todd and Gigerenzer (2007) notes that some environments – those that are similar to the environment for which humans evolved – can “make us smart,” while others lead us to error. When we see cockroach-shaped chocolate we act as though we cannot know whether the food is really chocolate or an insect. Behavioral experiments can only detect errors when there is a certain outcome to compare against the individual’s decision. In these environments humans are not rational. The behavioralist assumes that this carries forward to conditions of true uncertainty which characterize real life. If the human mind has evolved for real uncertainty, but not knowable risk, then that assumption is false and one cannot presume that errors are prevalent.

While experimental subjects fail to solve logical problems when those problems are presented in the abstract, humans have a sophisticated social exchange heuristic which allows them to easily identify breaches in a social contract. In the pure logical form of a problem, subjects are presented with a proposition that when “X” is printed on the face of a card, “1” must be printed
on the reverse side. Subjects are then presented with four cards, marked “1,” “2,” “X,” and “Y,” and told they must turn over just two to test the logical proposition. While most subjects turned over the “X” card, few correctly picked the “2” card and most wrongly picked the “1” card. When the proposition is rephrased as a social contract – a person buying beer must be over 21 – nearly all subjects correctly asserted that it was necessary to check if a person buying beer was over 21 and if a person under 21 was buying beer (analogous to the “X” and “2” cards) and nearly none thought it necessary to ID a person not buying beer (analogous to the “1” card).

The system of heuristics may thus perform, in many cases, just as well as rational man would do, and better than the boundedly rational deductive system. This finding alone is indeed important; were a policymaker to try to induce a man to follow only his deductive system she may lead him to greater error. Yet this is a far cry from the claim that the human mind is “better than rational.”

The human mind may be a more efficient decision engine than rational man because it reaches conclusions faster and does so using less processing power; but that does not mean the mind necessarily reaches better conclusions. Cosmides and Tooby (1992, 2005) also caution that the mind evolved in an environment very different to the one humans experience today. Cosmides (1989) and Tood and Gigerenzer (2007) both note that we are exposed to far more abstract statistical data than we would have been when humans first evolved. Stanovich and West (2003) argue that the concept of evolutionary advantage and economic rationality do not always overlap: evolved behavior should help us to survive, not maximize the economic concept of happiness.
Yet the human mind can still be better than rational. In the dictator game the desire for fairness left the player in the role of dictator worse off (and the other player better off). In the ultimatum game, both players were left worse off when the recipient turned down the ultimatum but the desire for fairness allowed the recipient to make a credible threat that a completely rational man could not. A lawyer may rue (or perhaps selfishly celebrate) the stubbornness of a client whose refuses to accept a deal that avoids a court trial, even though the cost of a trial exceeds whatever he is likely to gain, but must accept that such obstinance is sometimes effective. Johnson and Fowler (2011) found that over-confidence in entrepreneurs is a boon, making it easier for the entrepreneur to credibly commit to investors and improving the ability to negotiate.

In the broader scope of human activity, zero or negative-sum games are less common than are opportunities for advancement though cooperation. In the prisoner’s dilemma players can either cooperate, maximizing the size of the pie they share, or cheat each other, reducing the size of the pie but taking a larger share for themselves. If both players cheat then both are left worse off than if both cooperate, but both have an individual incentive to cheat the other and no way of committing to cooperate. Economists predict that both will cheat but some of the earliest experiments in economics found the opposite: players irrationally cooperated with one another and in doing so, made more money than if they had acted rationally (c.f. Sen, 1977).

Economists found similar results in the investment game. In this game the first player send an amount of money to another player, but the amount sent is multiplied by the experimenter. If the first player sends five dollars the second player might receive twenty dollars but if the first player sends ten dollars the second player receives forty dollars. The second player then
becomes the dictator, sending all, none, or some of the money back to the first player. In this
game the first player will usually send some (but not always all) of his initial endowment to the
second player who usually sends some amount back. Perfect strangers are thus able to
cooperate to get more money than if they were perfectly rational.

In a public good game, several players are allowed to send some, all, or none of their money to a
general pot where it is multiplied (by a number smaller than the number of players in the game)
and then distributed evenly among the players. Each player is better off the more money goes
to the general pot but individually players are better off by not putting any of their own money
into the pot. Again, players are made better off by avoiding the perfectly rational (and perfectly
selfish) equilibrium where no player puts money into the pot.

The fall from rationality does not disadvantage players in the most realistic games, the ones
where cooperating makes everyone better off. Trivers (1971) suggests that our emotional
desire for fairness, and our guilt when we act unfairly, is an advantage in eliciting trust. Frank
(2001) argues that our emotions allow us to make commitments. People can trust us because
we cannot help but be fair. Thus the trait of trustworthiness makes us better than rational
because it permits us to cooperate and to extract the gains from cooperation that rational men
cannot achieve.

J. Summary

Psychologists and behavioral economics have uncovered a range of biases, the majority of which
are not covered here, which undermine the assumptions of rational choice economics. Yet it is
hard for policymakers to act upon these biases. There is no complete list. There are known
examples and others that have yet to be discovered. Those biases we know have only been fully tested in the laboratory but it is impossible to isolate subjects entirely from their cultural and social experiences. Even when inconsistencies are observed, they do not reveal true preference or even error when the heuristics individuals use are understood as a collection of routines for solving problems that do not always generate the same result (Rebonato, 2012). Laboratory evidence gives policymakers little instruction on how people might behave in the real world, where biases may act together. We cannot know simply from the evidence whether a person might be over-optimistic or risk-averse in any given situation.

Some of the biases we find in the laboratory may be artifacts of our evolutionary past. Others may be cultural artifacts from the modern world. Heuristics are not always static but can adapt to the realities of our environment and may change over time. Some heuristics will do a better job than the boundedly rational deductive system which lacks the computational complexity needed to solve all the tasks we encounter on a regular basis. As both systems of reasoning can hold, at different times, the claim to represent true preferences, there is no way for policymakers to determine what an individual’s preferences really are.

Even if policymakers knew our true preferences it does not hold that we must try to make real people more like the simplified agents economists use for their calculations. Far from holding economic man as an idealized decision maker, Sen (1977) suggests that he is a “rational fool” who might mock boundedly rational humans as they cooperate when logic dictates they should cheat one another, but is ultimately poorer than his more limited brethren. Ostrom et al. (1992) takes a similar tack, arguing that economists should seek to understand, and appreciate, the human capacity to undertake collective action without a centralizing force. For the law to seek
to simply align human behavior with economic models would be a misguided effort that would make rational fools of us all.

III. Social Exchange

A. The Rational Ecology

The human mind has evolved to overcome cognitive limitations by combining rational problem solving with a system of heuristics: a combination of semi-flexible learned rules and rigid evolutionary artifacts. Whether heuristics are helpful or harmful depends upon the type of problem we are expected to solve. When applied to environment for which they are designed, heuristics can help economize on limited cognitive capacities. When improperly applied, they can lead us to significant error. As such it is not only our minds but also our environment that determines whether we are better, or worse, than the rational beings of economic models (Todd and Gigerenzer, 2007).

That environment, however, is not fixed but is a product of human activity. We create the social, legal, and economic arrangements which order our actions, and conceivably create institutions that make us rational. A smoker does this when he throws away a packet of cigarettes, as do his friends when they chastise him for lighting up, and as does a policymaker when she levies a tax on tobacco. Yet the challenge of designing our environment is far from simple. The smoker is just as boundedly rational when he throws away his cigarettes in frustration as he is when he buys and smokes them. The challenge then is to find the set of arrangements which help us align our actions with our inflexible preferences but do so under the same restrictions of bounded rationality that we seek to overcome.
Thaler and Sunstein (2003,2008) argue that the task of designing these arrangements must fall to a choice architect: a designated expert who can guide us towards our own best interests.

Turning to an expert makes intuitive sense. When we make decisions about our health or our legal affairs we turn to a doctor or a lawyer who knows the subject better than we do. Yet we only turn over our decisions in a limited sense. A doctor might advise the smoker to quit but he has no ability to coerce the man into action. Even when we know nothing we only ask a doctor to diagnose our condition, and to propose possible treatments. We expect, and the law requires, that the final decision must be made with our own informed consent.

Ignoring any ethical reasons for the patient’s rights, the division of labor makes practical sense. While the doctor possesses greater medical expertise, the patient is still the expert on his own preferences. Whether he prefers a prolonged life with pain to a shorter but more enjoyable life is a matter that no technical expertise can answer. If the patient does not know his own mind then the doctor faces a dilemma. In Schelling’s example the doctor must decide whether to administer an epidural to a woman who previously refused it. Her request for painkillers may be a heat of the moment error of the automatic system; her earlier decision might be a statement made at low cost that does not reflect her true preferences; or she may have made an error in remembering her previous pain.

For a regulator the challenge is even greater than that facing a doctor or a lawyer. Even knowing which decision is rational may not help when some deviations from rationality are beneficial. We would have to give the designated choice architect far greater power than we
suppose to give to doctors or lawyers. To make fully informed choices the regulator must know the individual’s mind better than she does. Not only does neuroscience not provide the solution to this problem but the task is impossible (Hayek, 1952).

Experts themselves are boundedly rational and expertise is merely a specialized heuristic: a sophisticated but nonetheless simplified model for solving problems without understanding the full complexity of the real world (Kahneman, 2003; Friedman, 1953). Such a model cannot capture the full complexity of the human mind and then exceed that knowledge (Hayek, 1952). Indeed experts will frequently disagree on the individual’s best interest depending on the model they use. An economist will use one model, and reach one conclusion, while a psychologist will use a different model and reach a different conclusion. One of these heuristics is inevitably less effective than the other depending upon the circumstance, but regulation frequently lacks a trial and error feedback mechanism to identify the most effective heuristic.

Thaler and Sunstein (2003, 2008) argue that because regulators are boundedly rational, choice architects must use a cautious approach. Regulators should use choice architecture to mold decisions but people should be able to reject the regulator’s counsel and choose otherwise. Rather than banning, or mandating, some activity, policymakers should use mandatory information disclosure or change defaults from opt-out to opt-in but leave individuals free to ignore the disclosure or still opt-out at a trivial cost. Yet information warnings can overwhelm boundedly rational people (Magat et al, 1988) and framing effects and defaults can have significant effects on outcomes (Kahneman, 2003; Thaler, 1985).

---

67 The above section on framing suggests that in fact doctors and lawyers play a greater role in making decisions than the law presumes.
68 Hayek’s exposition is somewhat complicated. Einstein’s description of the problem is simpler: “If the human mind were simple enough to understand we would be too simple to understand it.”
B. Social Contract as Choice Architecture

Although designed choice architecture can, and undoubtedly does, play a role in helping align our choice with our preference, it is not alone sufficient for the task. But conscious design may not be necessary. Our minds have evolved without any oversight, and the heuristics we use, both evolved and learned, are not chosen in any economic sense of the word. Just as our minds can evolve to make us appear more rational despite our cognitive limitations, so too might our environment emerge from an unguided evolutionary process.

Smith (2003) argues that this is indeed the case: institutions that make us rational are the expected outcome of social evolution. Expert guidance is unnecessary. Powell and Wilson (2008) find evidence to support this claim: subjects in a laboratory quickly converge upon a set of informal rules to facilitate social exchange without law with surprising propensity. The initial state of social arrangements matters: once a group of subjects converges on practices of cheating it is difficult if not impossible to then leap to cooperative exchange. Subjects can frequently avoid this by the natural human bias toward social exchange but outside the laboratory they have an alternative: leave groups with detrimental social contracts. Thus, in our cultural evolution, beneficial arrangements have an advantage over detrimental ones.

Anthropologists have made similar findings in small tribes (Cosmides and Tooby, 1992). Food sharing is important to survival in hunter-gather societies, but food sharing can also allow some members to shirk. Tribes not only develop social contracts for food sharing but these contracts differ between tribes. When the food supply is highly variable, as is usually the case with meat, tribal norms dictate a requirement for individual members to behave altruistically. When the food supply is stable, as with some crops, or when there is a stable food alternative so
starvation is less likely, tribal norms that demand less sharing create greater incentives to individual effort (Ibid).

On a larger scale Demsetz (1967) also finds evidence of efficient rules in larger groups emerging over time. When beavers were plentiful in the Hudson Bay, Native Americans thus had no need to claim ownership of the animal until after it had been captured. With the arrival of French fur trappers, demand spiked while supply dwindled. Over the years (in fact centuries) that followed, the Native Americans developed property rights in beavers to avoid unnecessary conflict over the now scare resource. The new norms did not completely eliminate violence but they created clear rules for ownership which at least avoided unintentional conflict.

Social contracts emerge from our propensity to cooperate and engage in social exchange (Cosmides and Tooby, 1992). Unlike market exchange, social exchange is based on a system of reciprocal altruism. Anthropologists have use this term to describe a system of barter without law, but also a process of delayed reciprocity. This latter system can be important in early societies (of the type in which the human mind evolved) when individuals relied on an uncertain food supply. If one person was lucky enough to have a food surplus he would share it with others on the expectation that when the situation is reversed, they will help him. But he will not give blindly. He keeps a mental account of who has helped him in the past. In his study of the social norms governing modern ranchers, Ellickson (1991) found the ranchers used a similar system of mental accounting of favors and preferred this to market exchange.

While, reciprocal altruism implies a form of trade with every member of the group keeping count of what every other member has done for him, an alternative social contract is possible. Riolo et al. (2001) shows that a large group can maintain altruism without direct reciprocity as
long as members can identify one another. Computer simulated agents helped other members of the group and were offered help in return, but no-one checked if the person they helped was the same as the person who helped them. Buchanan and Tullock (1965; see also Buchanan, 1975) show that individuals are willing to commit to such contracts because they do not know yet if they will be the likely recipient of aid or the likely donor. Thus an individual may agree to support a welfare program as a form of social insurance against risk.

A member of a church may aid another member who he does not know, not because he expects direct reciprocity, but because the social contract between members of the congregation dictate that he should help. The church is able to provide a simple form of insurance through altruism because members can recognize other members, even if they do not know the person directly, and because the individual member expects off through his membership of the church than if he refuses to help.

Ostrom and Gardner (1993) have shown that farmers in developing nations can maintain complex irrigation systems – public goods that economists predict cannot be provided through voluntary cooperation – without outside intervention or enforcement. Evidence from the laboratory suggests that maintaining cooperation requires the ability to enforce contracts within the group (Ibid., Ostrom et al., 1992). Not only do group members supply a public good, they altruistically punish (that is, punish at a cost to themselves) individuals who violate the contract. Outside the laboratory, groups may also ostracize non-cooperators. Members of the group refuse to trade with members who violate the contract (at a cost to both parties to the trade).

---

69 In Axelrod’s model it is not necessary the members of the group know each other, only that there is some clear signal of group membership.

70 Axelrod (1986) makes a similar finding with computer simulated evolution of norms.
The ability to sustain social contracts will depend in large part upon the cost associated when a member loses the advantages of group cooperation. In Ostrom and Gardner’s (1993) observations of irrigation systems, farmers achieved the most efficient and equitable distributions of water when the cost of diverting water was great. When the cost fell, or governments intervened to meet some of the cost, the need for cooperation was lessened and groups were less effective (leaving some farmers worse off).

The existence of the cooperative effort makes it possible for groups to provide other public goods, including the informal arrangements necessary to align individual behavior with true preferences. Indeed, the nature of social cooperation gives every member of the group a stake in the wellbeing of every other member. If a church has a practice of helping members who fall upon hard times then it is likely that it would also have norms in favor of thrift and work ethic, and against vices such as drug use. A person who violates these norms imposes costs on other members and so an efficient social contract would seek to eliminate this sort of behavior.

C. The Social Exchange Dilemma

While Smith’s (2003) claim of an evolutionary mechanism removes the process of intentional design it does not eliminate the need for a selection mechanism to choose efficient rules. Boundedly rational individuals must somehow identify and adopt the norms which best align their actions with their preferences. Moreover, the ideal set of rules must change over time. Whether or not norms in favor of home ownership or stock market participation are beneficial depends upon a range of economic circumstances. If norms are improperly applied they can work against the individual’s interests just as bad heuristics can.
Jolls, Sunstein and Thaler (1998) agree that social norms influence behavior but they argue that there is no reason to believe the overall impact is necessarily positive. There may be positive norms such as norms against smoking, or negative norms such as racism. Their analysis, and analysis by other norms scholars, stems from Sunstein’s (1996) concept of a norm entrepreneur: an individual whose specific job it is to develop new norms. The norm entrepreneur is essentially a policymaker albeit one who lacks the power of coercion. Like any other individual the norm entrepreneur is boundedly rational, although he may have some specialist knowledge. A norm in favor of thrift is thus just as likely to be misapplied as a government policy in favor of saving. Indeed, Jolls, Sunstein and Thaler (1998) argue that norms are less likely to be efficient than directed government policy. Instead, government should play an active role in influencing and promoting positive norms.

Eric Posner (1996) goes further, arguing that social constructs are not even benign. Norm entrepreneurs are driven by rational self-interest. While their self-interest might overlap with broad societal interest in some instances, other instances will see norm entrepreneurs promote their own interests at a cost to others. White norm entrepreneurs may promote racism because it advantages them but a norm of racism is still harmful to the broader society. Norms may also exist as a form of signaling (Posner, 1996) People develop in-group norms to make it easier to exclude or discriminate against non-members.

---

71 For a summary of the scholarship on norms see McAdams (1997)
72 For example, Ellickson (2001) assumes norm entrepreneurs are specialists but that there is also a role for opinion leaders who are not.
73 Signaling is the practice of adopting behaviors to distinguish oneself from others. For a signal to be effective it must be relatively less costly for the signaler to adopt the behavior than for the others one seeks to distinguish oneself from. Thus, if a person simply knows more about fashion then it will be less costly for them to dress fashionably than it is for a person who must first acquire a knowledge of fashion.
McAdams (1997) takes a more positive view. Like Posner, he believes that people are essentially rational but alongside other usual goods, they desire esteem. This demand for praise from others can lead them to supply public goods. McAdams notes that as providing esteem must be somewhat costly, the esteem mechanism merely creates a second-order collective action problem. For norms to survive it is necessary not only to esteem compliance or disesteem non-compliance; individuals must reward the act of giving proper esteem or punish failure to enforce norms.  

While McAdams argues that norms in general are beneficial, bad norms might still arise. People may esteem a particular hairstyle not because there is a collective action problem to overcome but because they happen to like it. These “nosy norms” have no public benefit, and indeed produce a needless conformity, but are enforced nonetheless.

Even if norms are positive they can still be over-enforced (Ibid.). People may esteem hard work or other activities which lead to success. Because McAdams assumes that people were rational beforehand, this esteem leads people to work harder than they should at the cost of more leisure or time spent with family. This can lead to unnecessary escalation as people wish to exemplify the characteristics associated with the norm.

---

This is considered a pure expense but signaling will confer some advantage on the signaler relative to the non-signal. Some portion of societal wealth (greater than the cost of the signal) is transferred to the signaler but total societal wealth is reduced.

74 For a more formal illustration, see Axelrod (1986).

75 It is McAdams’ view that nosy norms don’t provide benefits. In Posner’s (1998) model these norms might provide in-group benefits but are harmful to society in general. In rational choice economics, however, it is entirely plausible that people would have strong preferences over others behavior. If these preferences exist we should treat individual behavior as having externalities. An attractive mode of dress would have positive externalities while an unattractive hairstyle would have negative externalities. Supposing enough people were willing to pay something for you to change your hairstyle you might well be willing to make the deal. In reality these bargains are precluded by collective action problems and transactions costs. In this sense then, nosy norms might be wealth maximizing.
Bad norms (or previously good norms that now serve no purpose) are hard to change because once a norm is established people will esteem following the norm itself, generating a feedback loop. People either internalize the norm, or go along with it because they do not wish to be punished for opposing the norm (Kuran, 1997). It is possible for a bad norm to continue even if the majority opposes it because the norm itself has created a collective action problem (Schelling, 1978): people don’t want the norm but they do not wish to bear the cost of changing it. McAdams (1997) thus supports government interventions such as the civil rights act to change bad norms.

Ellickson (2001) approaches norms with a similarly guarded optimism but is more skeptical of government’s power to change behavior. His analysis of farmers in Shasta County, California found that farmers made wide use of norms to avoid conflict, share resources, and overcome collective action problems. Yet they rarely knew the relevant law and even when farmers did understand the law they ignored it in favor of commonly understood norms.

In Ellickson’s (2001) model there is both a supply and demand for norms. Norms are still supplied by change agents: three types of individual who drive norms. The first type act in their own self-interest (for example, someone quits smoking purely because it is in his own self-interest and others follow); the second type are norm entrepreneurs who have specialist knowledge (a doctor who campaigns against smoking); and the third type are opinion leaders who are generalists (a celebrity who supports the doctor).76 Whereas McAdams supposes that esteem works against changing norms, Ellickson suggests that people esteem successful norm entrepreneurs and opinion leaders. He introduces “norm customers” who buy into new norms.

76 Ellickson use of “norm entrepreneur” in somewhat narrower than other norms scholars.
that improve efficiency.\textsuperscript{77} This solves the collective action problem – whereby no-one wishes to bear the cost of changing bad norms – by introducing a norm that rewards people who bear the cost of changing bad norms. The demand side of the equation thus improves the likelihood of norms being beneficial but bad norms can still arise if customers themselves have an interest in promoting norms which benefit the group but not society in general.\textsuperscript{78}

Both optimists and pessimists of the New Norm School believe that bad norms exist and are persistent, and that these can be altered by government.\textsuperscript{79} Yet even the pessimists ignore the larger difficulty facing the selection of norms: the problem of knowing what norms are desirable. Boundedly rational individuals may sometimes be aware of their own failings and may take action to correct their behavior, such as when a person throws away a packet of cigarettes to prevent himself smoking. In this case the individual may support a social norm against smoking which would discourage him from the act. In other instances the individual does not know the optimal norms because of his own bounded rationality. If individuals’ bounded rationality leads them to save too little or borrow too much then as norm customers they are just as likely to support norms in favor of borrowing heavily.

The New Norms scholars assume away this problem by maintaining the assumption of full rationality but their solution creates more questions than it answers. In McAdams’ example, rational individuals are induced to work harder than they choose because of norms in favor of hard work. The problem exists because the norm is applied to a rational individual, but

\begin{itemize}
\item \textsuperscript{77} Ellickson uses efficiency in the utilitarian sense of the word.
\item \textsuperscript{78} Again, racism in the South is a good example.
\item \textsuperscript{79} Both McAdams (1997) and Ellickson (2001) make this point. For legislatures to want to change norms there must be broad assent which is possible in an esteem model when individuals become trapped in a bad norm equilibrium.
\end{itemize}
economic man is effectively a sociopath whatever preference for esteem we might award him. He has no use for norms precisely because he is rational. McAdams and Ellickson assume that he wishes to overcome a collective action problem but norms merely defer the problem. Providing norms becomes the problem and if we solve that by some norm of enforcement then that meta-norm becomes the problem.\textsuperscript{80} While other scholars have avoided the esteem problem by including the norm in the individual’s preference function this does not so much solve the problem as assume it away. If the norm is simply a preference then we do not have to worry about it leading the individual astray (Becker, 1996). Norms are neither preferences in the economic sense, nor are they biases of a rational individual.

Social norms are not merely similar to heuristics: they are heuristics. As with all heuristics, norms are rules we follow in place of complex rational calculation as a means of saving on limited cognitive capacity. Norms differ from other learned heuristics in that the rules they embody have external meaning. A norm of cooperation is an heuristic to cooperate with others who a party to the same social contract. They cannot be explained by a rational choice theory precisely because they are an answer to the fact we are not completely rational.

\textbf{D. The Adapted Society}

Like language, humans have evolved a capacity for absorbing culture because it permits us to engage in cooperative ventures and permits the easier sharing of knowledge.\textsuperscript{81} Indeed, these two adaptations likely arose at roughly the same time in our evolutionary history.\textsuperscript{82} Both are general heuristics. We do not have heuristics for specific words in German or English but we

\textsuperscript{80} Meta-norm is Axelrod’s (1986) term
\textsuperscript{81} Pinker and Bloom (1992) elaborate this point with respect to language.
\textsuperscript{82} At least in a geological time frame if not necessarily on a human time scale.
have evolved a capacity to learn language. Initially children who use the language heuristic will try to apply general rules, forming words that do not exist such as “hitted” or “cutted” but these will later be replaced by the specific grammatical exceptions known to adults (Pinker, 1992).

Likewise, we may have heuristics which permit us to easily adapt to culture, such an innate tendency to cooperate or conform to a social situation, but we will eventually learn the norms of our local culture.

Mental adaptations for language or culture are generalized structures because humans, even in early societies, had some need to adapt faster than evolution would allow (Pinker, 1992). Life on the African savannah is quite different to life in the rain forest or in northern Europe during the last glacial age. New words and norms are needed when humans migrate or climatic conditions change, or when the first humans abandoned hunter-gatherer lifestyles for farming.

As Cosmides (1989) notes, total plasticity leads to evolutionary death but so too does total rigidity. Humans are adapted to adapt, and cooperative exercise is one key to that adaptiveness. Once individuals within a group discover some aspect of the complexity of their environment they may not only share this information but develop complex social constructs to exploit this knowledge. Pinker (1992) reports that hunter-gather societies possess a sophisticated “folk biology” gleaned from the tribe’s collective experiences.

All heuristics embody a mental map of the world (Hayek, 1952). In the case of norms, the heuristics we hold include a map of the behavior of others. The norm itself may consist of varying obligations on different people (Sunstein, 1996). In a hunter-gatherer tribe norms might impose different obligations on varying members of the tribe depending on each member’s role, and corresponding expectations on the other members. In a modern firm there may be many
specialists with specific obligations. Other members of the firm likely lack the expertise required to understand the specific heuristics that govern the activities of each of their coworkers but would nonetheless have expectations for that worker’s behavior. Individually it would be hard to enforce each norm but the nexus of relations embodied by the implicit social contract can be enforced. By this method norms can allow for far more complex cooperative ventures than any particular party to the venture can understand.

Once individuals discover better methods for cooperative venture through trial and error they can transmit this knowledge to others. In this manner, the intricate rules and relations of a society can embody tacit knowledge of processes and outcomes without any individual having full knowledge of the process (Hayek, 1973). That the boundedly rational individual cannot understand the process is of no more concern than the fact that a single cell cannot conceive of how to form itself into an eye or a gazelle. Complex organisms are the result of generations of trial and error encoded into their DNA and complex societies are the result of trial and error coded into social norms.83

This process is simple when boundedly rational individuals can easily observe the benefits of a particular norm. If a hunter from an early human tribe finds success on new grounds, or adopts a new tool, others can easily imitate him. If routines that favor a work ethic improve the welfare of those who adopt them then those routines can be adopted as norms. It is harder to see how more complex norms might be adopted. The benefits of saving might take decades to accrue while the apparent benefits of a spendthrift lifestyle are immediately obvious to a boundedly

83 In the case of cells the individual cell in an organism does contain the DNA required to construct the entire animal, albeit without any learned behaviors. One could not, by contrast, construct all of society from an individual human member.
rational individual. When successful individuals adopt complex patterns of behavior it may be difficult to determine which norms are responsible for success and which are mere idiosyncrasies. Adam Smith (1759) noted the tendency of people to copy all the traits of the successful, including their bad habits.

Bounded rationality, the problem for which norms are the solution, can make it difficult for individuals to effectively select among individual norms. In practice, however, individuals are not required to make this selection. Individuals choose between groups, not individual norms. Indeed we may work very hard to penetrate some social sphere which matches our aspirations; adopting the norms and seeking to ingratiate ourselves with existing members. Through the simple task of identifying those groups which appear to the individual to create the greatest benefits, evolutionary pressure can identify the nexus of norms that constitute the optimal social contract. As we do this we may be entirely unaware of the process of cultural evolution in which we play our part.

It does not matter that we ultimately select for individual norms, or that the individual and not the group is the beneficiary. Biological evolution selects among genes, not organisms or species, but natural selection works upon the entire organism (Dawkins, 1976). It is only over time that evolution identifies the specific combination of genes that optimize for survival. Within the optimal (survival) set some norms may be arbitrary, such as whether we drive on the left or right. Some norms may be optimal in one cultural ecology and not in another.

---

84 For example, large claws are beneficial to a predator but not to a herbivore. Likewise, some norms may be beneficial in Anglo-Saxon cultures but detrimental in Japan
While the New Norms scholars have broadly accepted that exit and group selection can play a part in the formation of new norms, the process has been largely relegated behind other means of norm selection within groups. This is in part because they treat groups as being relatively stable entities such as entire societies, and in part because they regard exit as a means by which individuals oppose individual norms. The latter error likely stems from treating individuals as rational norm adopters rather than boundedly rational individuals using norms as social heuristics. The former error stems from a belief that because a norm is widely adopted that it is a global norm.

Human groups tend toward around 150 members – roughly the limit of our cognitive capacity for recognizing other individuals – though particular groups may still vary considerably in size (Ridley, 1996). Yet human cooperation extends far beyond the reach of the insular group, not only today but early in our evolutionary history. Even hunter-gatherer tribes enjoyed relatively fluid membership and openly traded with other tribes (Ibid.). Trade on this scale potentially permits the division of labor and the existence of far more efficient production techniques than would otherwise be available to a small group but also creates pressures toward the convergent evolution of language and culture.

Groups may converge on similar practices for more than one reason. If every firm in an industry adopts the same method of production it may simply be because it is the most effective. Trade between groups can allow cultural transmission of norms which benefit the groups that adopt them. Convergence may also occur to avoid conflict (Schelling, 1978). Just as common language and culture within a group can reduce conflict, so too can similarities between groups.

---

85 For example, see McAdams' (1997) review.
86 Ibid.
This is not the same as a norm in which all of society is the group. Rather it is the case that broader social environment is a part of the fitness criteria over which norms are selected. Natural selection has favored genes for language because those genes confer a benefit on individual humans in a society which has language. Neither the prevalence nor the social aspect of the gene implies that natural selection works on any vehicle other than the individual. Likewise, although evolutionary pressures may create benefits to standardization of language and culture, we should not assume that society is the relevant level of social organization.\footnote{I use the admittedly clumsy phrase “level of social organization” to mean the social analog of the organism in biology.}

In language the expansion of trade has reduced regional dialects in favor of standardization. In fact these dialects may merely be remnants of what once were different languages. Clearly there are benefits to speaking English if everyone else does, but this benefit selects over groups rather than determining the individual choice of language. We can see that when the benefits of deviation outweigh the benefits of standardization there is no convergent evolution.

Academics often have a need to describe constructs to one another and do so by creating new terms. As there is little pressure to standardize language across the disciplines we find that words such as “rational” have widely differing means in psychology and economics while words such as “heuristic,” “rule,” “norm,” and “routine” are sometimes used to mean the same thing. Furthermore, some groups, such as high school teens, purposely adopt differing modes of speech as a means of signaling membership of a group.

In culture too, the benefits of more open trade drive the standardization of those norms relevant to trade. We are less likely to be altruistic to strangers than to members of our own group because we have no social contract with that stranger, but we may still apply norms of
fair dealing. In a hunter-gatherer society if one member of the group mistreated a member of another group it might lead to broader conflict or the loss of trading opportunities. The group thus enforces norms of fair dealing with members of other tribes. Some groups may closely cooperate such as congregations who are members of the same denomination. They would be more likely to enforce altruism toward other members of the church than toward society in general but less so than to other members of the congregation.

South American traders once took the phrase “the word of an Englishman” to mean honest and fair dealings. English merchants evidently favored their good reputation, and the advantages in trade that conferred, to enforce honesty among their members. As with cooperating congregations, the merchant-wide norm does not require that all English merchants necessarily belong to the same group. There may have been cooperation between multiple groups to obtain this effect. Membership of any one group would not be necessary but there might still be consequences from a failure to align with any group. This seems to have been the case for Ellickson’s (1991) ranchers. There were many groups such as modernists and traditionalists, cowmen and ranchette owners, but these groups cooperated and traded with one another and enforced good behavior among their members. The groups also enforced against unaligned out-of-towners who ignored conventions.

A group of modern diners may enforce the norm in favor of tipping because the group’s norms internalize the benefits of reciprocity between server and diner. This does not mean that an individual diner will not leave a tip even if he is alone and has no intention of returning.

Boundedly rational individuals use heuristics to make decisions and sometimes, as in the laboratory, these heuristics will lead them to deviate from rationality. As with unnecessary
tipping, such errors do not always require correction but the existence of efficient mechanisms for norm development does not imply that boundedly rational humans become rational actors for all intents and purposes. Rather, such a theory can demonstrate when deviations from rationality are most likely.

Once we understand that all norms exist at the group level we can distinguish between norm formation and norm selection. Formation occurs within the group and understanding this process has been the primary objective of the New Norm school. However, understanding formation does not give us a better understanding of how norms are selected. Rather it is the equivalent of understanding cellular meiosis (the division of sex cells which produces random genetic mutations). This in itself is an important task but it cannot predict how natural selection drives evolution. Indeed, Darwin’s (1859) theory of natural selection preceded our understanding of genetic mutation by some decades (Morris, 2001). Though we should not ignore the formation of norms within groups, any theory which can accurately determine the efficiency of norms must take into account the dynamics of norm selection.

E. An Evolutionary Model of Norm Selection

I propose here a model in which groups are selected through evolutionary competition. It is important to distinguish this from theories of group selection. In these theories, an individual group member adopts behavior, including self-sacrificing behavior, because it maximizes the likelihood of group survival. These theories have fallen out of favor in biological evolution in the

---

Infra note ??: This is not a perfect analogy unless the group once formed is completely rigid. Unlike organisms, whose DNA remains, a group can usually adopt new norms during its lifespan. As with organisms the group’s ability to adapt to changing circumstances will affect evolutionary success.
face of strong evidence that evolutionary selection works on the phenotype i.e. the individual –
level (Dawkins, 1976).

In cultural evolution, Hayek (1973, 1988) proposed a concept for group selection where macro-
level groups adopted rules that favored their evolutionary survival (c.f. Zywicki, 2000). This
model presented here, however, is not a theory of group selection in this sense. Instead, I
propose that much smaller groups compete (in the same sense that organisms compete) for
members by offering beneficial rules. By allowing fluid membership it is possible for rules to be
selected which enable group survival (though not at the cost of the individual).

I define groups as inter-connected networks of individuals who organize for the purpose of
overcoming collective action problems, and sharing complex models of the environment for the
purposes of overcoming bounded rationality. Each group embodies an entire social contract
which consists of a nexus of norms. It is not necessary that every member of the group be
connected to every other member, but each should be able to identify other members of the
group.

A group can enforce this social contract only if it can punish group members who violate the
contract and prevent non-members from free-riding. In an early society punishment for
violation could have included physical sanctions as well as ostracism but modern societies
generally proscribe the enforcement of informal norms through violence. Thus exclusion is the
primary mechanism by which groups enforce norms; including meta-norms of enforcement.

89 By comparison, in a rational choice model of norms, individuals wish to enforce norms against all other
individuals.
90 Though violence presumably happens in some instances.
Groups can enforce norms only to the extent that the cost of exclusion exceeds the cost of defection. Individuals do not calculate these costs explicitly but it will be generally easier to compare the net benefits of group membership than to rationally evaluate individual norms against an objective criteria. People may make a fair approximation of group value through well designed heuristics evolved for this purpose. When the overall value of group membership is high the group will be able to enforce more costly norms with potentially higher benefits. The farmers in Ostrom and Gardner’s (1993) study were only able to enforce the optimal social contract when group cooperation was necessary to obtain water. Thus the scope of the social contract itself is factor in the group’s ability to enforce individual norms.

Modern societies have, at least superficially, trended toward less reliance on groups. The law enforces contracts, provides social insurance, settles disputes, and provides protection through a police and military. This formalization of relations may be advantageous in expanding opportunities for trade and division of labor but can potentially reduce the importance of group membership and undermine beneficial norms.91 Yet in practice group membership may remain of great importance in a modern economy. Networking is an important part of finding a job and of making and maintaining business contacts. Much of our knowledge of the world may still come from the advice and experience of others. By contrast formal market transactions can involve high transactions costs, such as writing formal contracts or obtaining knowledge of the relevant law. Enforcing a party’s legal rights through the courts can be especially costly. For most, group membership thus remains an important part of economic life and ostracism a significant cost.

91 See *Infra* Section IV for further discussion of this point.
As Eric Posner (1998) argues, groups will indeed engage in costly signaling to distinguish group members from non-members. Rather than being a pure inefficiency, however, signaling norms allow groups to prevent free riding by non-members. As such signaling is merely a cost of using voluntary action to overcome the collective action problem in the same way that maintaining courts, regulators, or even prisons, is a necessary cost of overcoming the collective action problem through government action.

Signaling norms may exist for other purposes too. The English merchants may well have known one another but wanted to signal their group membership to people outside the group. Again, while such signaling may be costly, it is a cost that must be compared to other costs of enforcement. Signaling might also serve as a means of raising the cost of exit so as to increase the group’s ability to provide benefits. Gang members may receive tattoos to increase the cost of defecting from the gang or members of a community might enforce a norm in favor of home ownership rather than rental. This introduces a punishment for defection and makes exit or ostracism more expensive than mere non-membership.

As ostracism is necessarily costly groups will usually seek lower cost methods of enforcing norms. Group members may play “tit for tat:” temporarily refusing to interact cooperatively with another member who violated a norm. Groups may use voice as a means of deterring violations. Rational individuals should never respond to costless cheap talk but boundedly rational people do in laboratories and in real life (see e.g., Ostrom et al. 1992). People may violate a norm in error or may identify a situation where norms are vague. Communication helps avoid costly punishment for accidental infraction or to establish clear norms where none previously existed.
Making exit more expensive increases the ability of the group to provide benefits to membership but can also reduce competitive pressure. Individuals will rarely seek to leave a group because of an individual costly norm precisely because such behavior would make it impossible for the group to enforce norms. For a member to leave the group the benefits of the entire contract must decline sufficiently that the benefit of leaving exceeds the cost. However, it is not necessary that the benefits of group membership decline to zero. A member may either join another group or a number of members may break off from the original group. Anthropologists have observed both behaviors in hunter-gatherer societies (Ridley, 1996) and computer simulations have produced evidence for break-off groups (Riolo et al., 2001).

When small groups break off from larger groups they retain much of the norms initially found in the larger group but they will be more flexible. With fewer people the costs of communication are lower and it is easier to adopt new norms wholesale. Small groups will thus be typified by more rapidly changing norms while large groups will be relatively rigid. This rapid pace of change does not assure that the smaller groups should have more efficient norms. Rather, small group formation allows for a process of trial and error, experimenting with new combinations of rules. Even boundedly rational individuals are somewhat rational so radical mistakes, such as suicide cults will be rare. Nevertheless there are significant potential advantages to making radical jumps. Small groups can avoid being trapped in local maxima.

---

92 Leaving a group amicably may sometimes have lower costs than defection.
93 In practice these are perceived costs and benefits which may incorporate some error.
94 Thus norms will generally be positive at the group level with the notable exception that when the group controls a political or physical resource it is possible for a social contract to impose an overall cost.
95 Evolutionary systems always risk being trapped in local maxima when they only make small jumps. The popular analogy is that of climbing a mountain. If a blindfolded mountaineer climbs by an evolutionary method he will experiment with traveling in every direction and choose the direction that takes him
These larger jumps can create substantial gains but inherent randomness will also produce many small groups which develop very inefficient norms. These groups will usually fail as the cost of exit will be low and members will have strong incentives to leave. However, if enough groups make errors this may generate a sufficiently large disturbance to alter the environment over which norms are selected. If most people drive on the right then norms for driving on the left will fail unless enough groups simultaneously make the same error in which case the environment will now support norms for driving on the left.

While small groups are flexible and error prone, larger groups are necessarily more rigid. Near universal assent will be necessary to replace an old norm once it has been internalized and large groups have higher communication costs. The higher cost of changing norms will generally be beneficial. The growth of the group is due to its having efficient norms and new norm formation is subject to the limitations of bounded rationality. If members could easily change those norms then the effect of errors would be likely to reduce the value embodied in the social contract. The protection rigidity provides against the invasion of bad norms will exceed, at least initially, the cost of making it harder to adopt better norms.

Despite increased rigidity, large groups will occasionally adopt new norms. Boundedly rational members can still make use of both systems of reasoning. External stimuli, such as new information about the dangers of smoking, can trigger a revaluation of existing heuristics and

---

slightly higher. Over time he will reach the top of the mountain by this method but he may be unaware of nearby higher peaks (the highest peak being the global maxima). If he takes superhuman jumps he reduces the risk of being trapped on a lower peak (the local maxima) but also increases his chances of calling into a crevice.

96 Changing norms requires assent but not consent. Nearly everyone must know the norm and either abide by it or exit, allowing the few who do not comply to be punished. It is not necessary everyone agree with the norm.
norms. Such changes should be relatively rare because norms can embody knowledge not available to the deductive system. Our tendency toward conformity, the apparently irrational bias that leads us to make mistakes, can protect us from becoming rational fools. However, new norms may arise relatively frequently when group members are forced to respond to novel circumstances. Groups must develop new norms whenever they are confronted with an entirely new problem for which cooperative solutions exist, or circumstances merit an exception to an existing norm.

While large groups tend to be large because they started out with efficient norms, social contracts will generally drift further from optimal set of norms. The group at one point induced better than rational behavior in the wider environment but over time that environment is changed by economic and legal innovation, and by the changing behavior of other groups. Previously efficient social contracts can become less efficient both because inefficient norms are locked in and because efficient norms are accidentally changed, while groups can adopt suboptimal responses to novel situations. While combination of these factors makes the norms in large groups less efficient over time the actual change in both norms and membership may be relatively rapid. Only when the social contract drifts far enough from optimality will members bear the cost of exit of forming a new subgroup.

The model thus predicts the same norm cascades predicted by rational choice models and observed by sociologists. In fact entire groups can behave much like the individual rational agents described by the New Norms scholars, although they tend to be more rigid and there will

\[97\] In this respect group norms will form in much the same way as does common law. Judges respect precedent except in unusual circumstances but will freely form new precedent. Hayek (1973) offers a similar reason that the common law embeds tacit knowledge that may not be available to individual judges.
be fewer groups than there are individual agents. Like those individuals, groups may benefit from adopting similar norms to other groups. There may be periods of stability when large groups retain relatively rigid norms but when some of these groups break up and adopt new norms this break-up changes the environment for all other groups. If groups generally have efficient norms then this perturbation may have little effect but when change over time has eroded the value of the social contracts, these changes can cascade through the system triggering more break-ups and further altering the landscape. These new groups would generally be smaller and, although there would also be more groups, it will generally be easier for those groups to adopt new norms making the landscape temporarily fluid. These cascades would not only apply to individual norms but many norms may change rapidly together before successful groups begin to grow again, reintroducing rigidity and settling norms back into an evolutionary stable equilibrium.

The system of social norms is thus a complex and dynamical one. Norms are not taxes and subsidies on our otherwise rational behavior but a part of a separate system of reasoning capable of giving a better response than the finite capacity of our rational minds can achieve. Nor do individual norms exist separate to one another; they are part of an intricate social contract connecting together whole groups. This does not mean that norms the norms we have must represent the best of all possible worlds or that policy can never improve on the outcome

---

98 There may also be some benefits to non-conformity.

99 I borrow the idea of the evolutionary stable state from biology (see e.g., Dawkins). For the concept of cascades I turn to complexity theory where a small perturbation to a system can produce either a negligible effect or a significant disturbance (the “butterfly effect”). For example, a small amount of sand added to a pile may have no effect or it may create an avalanche during which period the solid sand takes on some of the characteristic properties of a fluid. See (Buss, 2005) for a further discussion of these principles. The overall system I describe is somewhat similar to the idea of punctuated equilibrium in evolution described by Gould and Eldridge (1972) where long periods of stasis are interrupted by geologically brief periods of rapid evolutionary change when organisms belong to small groups.
of norms. Rather it suggests a middle road. Bounded rationality will sometimes lead us to error but our evolutionary and cultural experience embedded in norms and heuristics can also sometimes help guide us to better solutions than we can rationally deduce.

IV. Prescriptions for Policy

A. The Knowledge Economy

The challenge facing a paternalistic policymaker is to identify the institutions that can most successfully align choices with true preferences. Those institutions must embed knowledge about the state of the world in a manner that is readily accessible to bounded rational people. That is, the institutions we pick must make us smarter than we really are, not less smart.

Rational choice economics, by reducing the role of markets reduced to a simple mathematical calculus, ignores the intermediating role played by institutions. Economists have given us, as Coase (1988) is quoted as saying at the beginning of this article, “consumers with humanity, firms without organization, and even exchange without markets.”

Behavioralists have given consumers their humanity but remain quiet on the issue of exchange. Thaler and Sunstein (2008) assume that firms are still the black boxes of economic theory: ruthless and profit maximizing, and fully equipped to exploit the bounded rationality of real people. They model boundedly rational people as interacting directly with these firms in the atomistic world of economic models. In fact exchange operates through two venues; market exchange and social exchange. The latter might be characterized as exchange with humanity.
People cooperate to produce public goods without formal markets and beyond the shadow of the law. But there are limits to humanity.

Market exchange, the cooperative endeavor between complete strangers without central coordination, allows for the far greater division of labor than social exchange will permit. It is not merely the exertions of physical labor that are shared but mental effort too. Even boundedly rational individuals can develop specialist heuristics over their own narrow fields. List (2003) observes that traders do not exhibit the same heuristic biases as other experimental subjects, although they might well be just as error prone in other areas. Markets, like norms, can embed knowledge about the state of the world (Hayek, 1937). The simple and narrow heuristics that each individual uses in his role belie the complexity of the entire system.

Hayek (Ibid) argued that markets exist, not in spite of, but as an answer to bounded rationality. Each individual need only know his own businesses and the information carried in prices can inform him of the wants of others (c.f. Boettke et al., 2012). Yet if the individual does not even know his own mind then markets cannot tell it to him. Thus the market must be intermediated either by social institutions or by government.

Government and social exchange both manage expertise but differ in the manner in which they organize knowledge. Each expert does not have a complete mental map of the world but a rather a set of specialized heuristics for solving problems. Whereas social exchange relies upon the experience embodied in the relations between individuals to translate individual efforts into cooperative action, governments must somehow manage this expertise hierarchically. Firms must try to solve the same problem but firms have narrow a purview and use a simple profit.

Shadow of the law is yet another term I borrow from Coase (1960)
heuristic to evaluate their behavior. Regulators, by contrast, have no equivalent to the profits of firms or the dopamine feedback system employed by humans.

Far from an efficient heuristic for managing information, hierarchical management can often resemble something more akin to the children’s game of Chinese whispers (Tullock, 2005). The regulator must attempt to weigh the arguments from economists, engineers, psychologists, or other competing experts. Yet the regulator lacks the specialized knowledge of the experts and must use her own heuristic to select between the views.\textsuperscript{101}

The solutions experts arrive at through rational deduction will not always be superior to current norms. Rational deduction with finite information and under conditions of uncertainty can still lead to serious error which can be hard to detect in the absence of feedback mechanisms. The problem regulators seek to solve is to align the infinitely varying preferences of individuals over the vast array of choices available to them. The regulator must make rules sufficiently flexible to allow for different preferences while subtly guiding each individual to his own best interest.

In doing so policymakers must avoid the temptation to treat expert models as perfect representations of how humans do behave, or should behave. The rational choice model of economics is merely an heuristic for roughly describing human behavior. People will not only deviate from the simplified model but will often benefit from doing so. Nevertheless, government can still play a role in helping individuals align their behavior with their best interests.

\textsuperscript{101} See e.g. Fama and Jensen (1983 a,b)
The behavioralists have wisely focused much of their critique on financial markets. These have obvious appeal as a test of economic theory. Financial markets are closer than most human interactions to the mathematical constructs of economists. They are formal trading venues where strangers meet and where participants can reasonably be expected to be entirely self-interested. Moreover, there are no complex preferences to disaggregate: traders care about money. If even here, traders are not entirely rational – which they are not, although as noted above they may be more rational than most of us – then rational choice economics must require radical revision.

The behavioralists have been successful in their effort to detect deviations from the rational choice model within financial markets. They are also likely correct in their diagnosis. However, they are wrong in their assumption that this implies greater deviations from rationality must exist elsewhere. Financial markets, by very virtue of their closeness to economists’ models, also lack the intermediation from social exchange mechanisms.

Participants are frequently not members of the same group and do not enforce norms against each other. This is not to say that individuals can develop no heuristics or norms about the manner of stock market participation. However, we should reasonably assume that the more formal and impersonal the manner of exchange, the greater the role of government in intermediating relations. There are similar benefits to regulating behavior on roads, where exchange tends to be impersonal and between strangers.
B. Creating Social Norms

Almost all of the New Norms scholars, regardless of whether they are optimists or pessimists, suggest that government should play some role in the formation of norms. Sunstein (1996) suggests that we may hold to different norms as citizens than we do in our normal lives: a man might support government recycling programs but fail to recycle himself. He might do this rationally if he would prefer that everyone recycle but finds that the benefit when only he recycles does not justify his effort. He would happily recycle if there was a social norm in favor of recycling but if no norm exists then it may be extremely difficult for every member of society to simultaneously jump to an equilibrium where everyone recycles and enforces the norm (Schelling, 1978). The man would therefore prefer government act to establish such a norm.

There is no reason why only good norms might emerge in this way. A person might prefer that a norm of racism exists but does not want to publicly argue for one. Moreover, boundedly rational individuals should make worse decisions when they try to choose between specific norms rather than selecting between groups. In addition to the inherent difficulties in rationally calculating the ideal set of norms, the individual gains little from making a correct choice. In experiments, higher reward generally induces greater rationality in individuals (Smith and Walker, 1993). In voting, the likelihood of altering a result is extremely small and thus material interest is unlikely in itself to induce rationality. Further, because voters generally do not all know one another there is no group that can internalize all the benefits of voting. The act of

102 While the formation of norms may be less important than those scholars suggest relative to the selection of norms, the law undoubtedly forms a part of the selection criteria for norms. Thus government certainly can influence norms.
103 A norm may also not form because no group exists that is large enough to internalize the benefits of the norm.
104 Buses only became segregated in the South after the Jim Crow laws.
voting is therefore unlikely to be intermediated by social exchange except in small communities.\textsuperscript{105}

Brennan and Lomasky (1993) suggest that because the individual has so little chance of changing the outcome, his vote is largely an expressive act. Voters enjoy voting for a policy and can be confident that their own vote has little effect on the outcome. Caplan (1997) goes further in arguing that voters are “rationally irrational,” juxtaposing the Freudian sense of rational with the economic sense. Voters enjoy believing their favored policy whatever the actual evidence suggest. In this model the voter is rather like a boorish guest at a dinner party, freely imparting his own advice. Not only does it cost him nothing to hold this belief, discovering contradictory information is costly. Caplan suggests that the confirmation bias, the tendency of humans to see new information as supporting one’s existing prejudices, makes learning of our errors even more costly.

The idea of willfully ignorant voters may stretch credulity further than most economists, or even behavioralists, would be willing to accept. However, the nature of voting makes it harder to use dopamine response system. The effect of voting is often not felt for some significant time after the act. Even then, voters are usually unaware of how voting differently might have changed the outcome. Even trained economists can only speculate whether spending more or less would have averted a recession because the proffered alternative clearly did not happen. These are the sorts of abstract questions subjects most frequently fail to answer correctly. If the voter only knows that the recession has not ended then confirmation bias may well cause him to think the evidence supports his own view, whatever that may be. Absent norms or dopamine

\textsuperscript{105} Sunstein (1996) suggests that higher levels of government should actually intervene against social norms interacting too closely with local governments to avoid groupthink.
feedback he has nothing to challenge his wrong belief. Indeed, if he does get some enjoyment from simply expressing a view then his error correction would enforce his views, be they right or wrong, each time he expresses them.

Given the low stakes of the game (for the individual) voters should be more likely to rely on heuristics when they vote but not to learn detailed or specialized heuristics for voting. Workers and consumers are more consistent when they weigh risks against reward than are expert regulators, who tend to put greater weight on emotionally charged risks; presumably in response to political factors (Morrall, 1986). This suggests that individuals are more likely to apply the availability bias when they vote than they are in ordinary interactions. Loss aversion can explain regulators documented tendency to prefer sins of omission to sins of commission. The FDA tends to give greater weight to lives lost from drugs that are already on the market than it does to lives lost because lifesaving drugs were not allowed onto the market (Gieringer, 1985). Similarly, regulators tend to be as conservative about banning existing practices as they are about permitting new practices (Huber, 1983). Again, loss aversion seems to explain why old and new risks are not afforded the same treatment.

Once established, loss aversion can make it harder to change a rule where a social norm might have collapsed over time. Once government creates a rule it alters the environment in which social contracts are formed. Social norms and the law may then work together to lock-in a particular rule. If government passes a law supporting discrimination against blacks then norms supporting discrimination are less likely to die out and voters are less likely to support repeal of the law. Although governments may often be tempted to believe that they can act more quickly than slowly evolving norms, there is even greater potential for locking in bad rules. Regulators
must thus apply special care that they do not apply normative values to their judgments and freeze current norms in place.

C. Law and Norms

If individuals could costlessly work around the law, then soft paternalism would have little effect. If the regulator bans a particular product or levies a tax upon it then that must change behavior, but warning labels and opt-outs should have little substantive effect on rational individuals. Indeed, Thaler and Sunstein (2003, 2008) rely upon this point to make the claim that the costs of soft paternalism are trivial. As psychological evidence reviewed above demonstrates, people are not rational and the institutions that intermediate between individuals and markets can be highly beneficial or, by the same measure, highly costly. However, if social exchange intermediates perfectly between individuals and the broader environment then individuals should be behave as if they were rational.

Soft paternalism is potentially costly not because individuals are perfectly rational or social exchange perfectly intermediates interactions with markets, but because people are boundedly rational and social exchange is an imperfect solution. Norms take time to form and people will often carry over the behaviors they have learned in other environments. If somewhat efficient norms have evolved over time to combat a particular problem then by changing the environment the regulator risks shifting individuals further from the ideal behaviors. A college education is a valuable investment but the benefits take a long time to accrue. Behavioralists

\[106\] Coase (1960) makes this point with regard to establishing defaults in the law with and without transaction costs. The theory translates somewhat easily into bounded rationality if we treat discovering information and expending finite cognitive capacity are transactions costs of a sort.

\[107\] Infra Part II
have uncovered a bias towards under-weighting events in the far future (Laibson, 1997, 1998). A policymaker may therefore conclude that people would acquire too little education and implement programs, such as subsidized student loans, to offset that tendency. But if there are already norms favoring education then this policy is potentially harmful.

The subsidy might initially cause people to seek excessive education and some graduates would find themselves without the jobs for which they have trained. This situation might well persist for some period of time in which norms and subsidies combine to generate an education bubble. Eventually, however, the bubble bursts and norms adjust to the new information. If many graduates are unemployed and languishing in debt then norms are less likely to support education. But, because of the subsidy, the private benefits to education have increased. Norms that accurately align behavior with private interest would still favor education.

Whereas the new paternalists tend to see heuristics and norms as static and exogenous, both change over time in response to the environment. If the legislature recognizes that people fail to understand complex contracts and passes a law limiting the consequences of not understanding a contract, then people are less likely to read contracts. A study of the effects of inflation over several countries, (Langdana, 1994) found that businesses in each country always reacted in the same way to increased spending by customers regardless of whether the effect was brought about by economic boom or inflation, but the response varied between countries depending on the usual mix between the two factors. In countries with high inflation businesses treated increased spending as mostly inflationary whereas in countries with low inflation they treated the spending as mostly a sign of growth. The government could confuse individuals with

108 This resembles Sunstein’s claim that norms can be treated as taxes and subsidies on behavior
one-off monetary stimulus but overusing the mechanism could not stimulate growth because the heuristics adapted. The process by which the individuals in each case come to change their behavior is not the same as that of a fully rational individual but it reaches a rather similar conclusion. Thus the rational choice model can still be helpful in understanding the relative effects of a change in costs even if it is not helpful in explaining the absolute state of the world.

In some cases, however, the interaction between changing rules and changing behavior is more complex. Heuristics, including internalized norms, are a response to our limited cognitive faculty. For the same reason that we need rules in the first place, we cannot have an infinite number of rules to match every possible situation. Instead, a procedurally rational individual should hold the most detailed rules where the benefits of doing so are greatest while retaining only the most general heuristics for uncommon and low cost problems. Adding too many rules, either from government or from social norms can overwhelm our limited cognitive capacity.109 Some disclosure rules can guide people to make better decisions but too much information can lead us to error. Too many warnings make us disregard all warnings (Magat et al., 1998). In some cases, like the Americans who were less able than German subjects to identify the larger of two U.S. cities, the additional information replaces an efficient heuristic with an error from the deductive system (Goldstein and Girzerenger, 1999).

Because it must apply to all and because lawmakers do not possess knowledge about each individual’s circumstances, the law must be broad but also shallow. In some cases the law will be too expansive or introduce too much detail and overwhelm our cognitive faculties. In others, the policy is a far blunter tool than are norms. In the education example, the subsidy cannot

---

109 See Hale, Boys, and Adams (2011) for a more in depth review.
take into account information that would be readily available to members of the individual’s group: whether he has an aptitude for college or whether the subject he has chosen is one that will increase his potential or is merely a distraction. If he does poorly then members of the group can more accurately measure if the individual has tried his best but struggled or if he has failed through lack of effort. In the former case his failure may be attributed to an error in attending college in the first place, for which he would not be punished. In the latter case he would be pressed into trying harder.

It is possible that groups still enforce these norms. Indeed we should imagine that in many cases they do, though not as strictly as is if the individual or group bore the full cost of the error. But the individual has been given by the subsidy a degree of freedom to defect against the norm without losing his financial support. He may be pressured to try harder but he understands his wellbeing is not affected in the short run if he does not. The norm has been weakened and the subtle distinctions of norms have been lost. Furthermore, Pildes (1996) notes the more nuanced approach to policing the norm has also been lost. Whereas groups can rely on low cost speech to restrain and correct defection, the formal law most often relies on more substantial punishment.

In some situations no norms or heuristics will exist because the individual has had no chance to learn (we only save for retirement once) and there exists no group which can incorporate the lessons of the past. In these instances policy can help people make better choices. However, it is also possible that people lack norms or heuristics because the problem is novel. Thaler and Sunstein (2008) use the example of a change in Swedish pension law (giving individuals more freedom to invest their own money). In the rational choice model of economics people should
have immediately improved their investments relative to the decisions made on their behalf. Instead they made several errors economists regard as irrational. Over time, however, these errors subsided. Time would allow people to learn from initial errors. Sufficient time may even correct the tendency to overreact to information provided individuals are able to learn and transmit that knowledge through norms.

Changes in the law may also affect norms or heuristics entirely unrelated to the law being passed. This happens to some extent when adding to an excessively complex warning label leads consumers to ignore the other, unrelated, warnings. Or it may happen when rules disrupt the social contract as a whole. With the student it is difficult for the group to enforce norms because he relied little on the group as a whole. When the government provides public goods, including laws, that displace the role of social groups, they reduce the cost of violating the social contract, and as such make it more difficult to enforce any norms.

Agricultural societies could only enforce an equitable and efficient distribution of water when every member of the group was needed to build the dam and irrigation network, but any provision of public goods can potentially replace social norms. The provision and enforcement of abstract property rights allows strangers to do business more easily without intermediation by social exchange. This need not greatly impede group formation if social exchange is still sufficiently valuable. The provision of some public goods, chief among which the provision of law, is necessary for the considerable wealth and diversity of modern society and as such can

---

110 The investors spread their money between multiple broad market indexed funds. These funds generated different risk exposures by differently weighting stocks so the individuals would have been better off simply picking one fund with the appropriate risk exposure. The general desire to divest the portfolio was correct but investors failed to recognize the funds already did this.

111 Social exchange would allow groups to develop norms on the subject of investing but would not facilitate cooperation between investors in large organization.
contribute to the opportunities for social exchange. Only in an entirely atomistic society does a person have no need for social exchange and such a society rarely exists outside of economic theory.

In trying to organize human activity for the benefit, policymakers risk disrupting existing norms. In a study of cities Jacobs (1961) observed that city planners inadvertently disrupted social norms because they failed to recognize the role norms played in maintaining communities. Planners ignored, or were unaware of, the importance of public areas for communication and the enforcement of social contracts. In eliminating these areas so raised the cost of group formation and maintenance. These groups, more than police, were important for maintaining public order and with their decline cities became more dangerous and city life more atomized.

Traditionally, groups possessed the means to provide welfare for members. At one point (largely during the late 18th and early 19th centuries) workers may have lived nearby one another and belonged to Friendly Societies which provided insurance for the group (Gosden, 1967). Before then and in smaller rural communities, churches provided a similar locus for social activity. Similarly, the middle and upper classes had their own institutions be they churches or other civic organizations.

With an organization such as a Friendly Society it was possible for groups to enforce norms in favor of a strong work ethic. Indeed, members had good reason to want another member to remain healthy and avoid behaviors that may have made him dependent on the group. With the advent of the welfare state the lowest income groups gained the benefits of redistribution

112 For a brief summary of the relevant elements of Jacobs’ book see Pildes (1996)
113 Workers would likely have been largely male so it is more accurate to say that the heads of households worked together and that this tied families to the same group.
and those not aligned with groups gained access to similar services that group members had previously enjoyed. By the same measure, it was no longer necessary to belong to a group to gain these benefits.

In providing assistance regardless of group membership the welfare state weakens the structure of the social contract; with the potential of making violations more common and ultimately disrupting the norms entirely. Because the welfare state combines redistribution with social insurance, it provides the greatest benefit to the poorest. Equally, it the most significant impact on norm formation should also be among those with the lowest incomes.

Beaulier and Caplan (2007) note that low income is associated with cognitive error. Although low incomes are associated with poor health, increased spending on healthcare has very little effect on health (Reinhardt et al, 2004). In fact healthcare spending is higher for the poor than for their healthier and wealthier counterparts (Sutherland et al, 2009). The apparent paradox can be explained by the greater tendency for the poor to smoke, or suffer from drug addiction or obesity; causes of poor health (Beaulier and Caplan, 2007). Murray (2012) ties this more explicitly to norm collapse whereas Beaulier and Caplan allege that low intelligence is to blame.

In a somewhat similar vein Jones (2008) argues that low intelligence can explain low income because lower intelligence individuals are more likely to be impatient less likely to cooperate in laboratory experiments. Even if IQ testing measures only cognitive capacity and nothing else, low intelligence should indicate greater reliance on rigid norms rather than the absence of norms.\textsuperscript{114} It is more likely that Jones’ finding is a consequence of the tendency to carry

\textsuperscript{114} Jones suggests that policymakers focus on raising measured IQ which would indicate he does not see IQ as a pure measure of cognitive capacity.
heuristics from general society into the laboratory, indicating that poverty is indeed associated with weak social contracts.

The persistence of social norms may have led policymakers to think that the dire predictions of rational choice economists were flawed. Yet people simply continued to behave rationally for their previous environment. While rational choice economists may have been overly pessimistic about the initial effects of welfare they have been excessively optimistic in the long run. Although it has provided aid to many, the welfare state has also weakened the social groups which had once intermediated between the individual and the marketplace.

Those institutions are now gone and abolishing the welfare state, were it possible or desirable, would not reinvent successful norms overnight. Attempts to induce rationality through broad brush rules - whether through forcing welfare recipients to participate in drug tests or by trying to control the diet of fast food customers - will tend to lack the subtlety and nuance in that norms can provide. Moreover, such attempts presume the law can be used to inculcate new norms into a society. Indeed several of the new norms scholars have suggested policymakers should be willing to pass laws they do not intend to enforce. The scholars argue that by using the law as a form of expression, policymakers can communicate the wishes of society and thus inculcate a new norm. A law against allowing dogs to foul public areas or parking in a disabled spot should make it easier to enforce norms against those activities.

This view misconceives the fundamental relationship between law and norms. It is not the law that gives rise to norms but the background network of social norms that permits the effective functioning of the law. This became evident with the conversion of the former communist bloc to free markets. Whereas rational choice economists had imagined that simply defining the
institutions of property, the rule of law, and democracy would be sufficient, the practitioners of markets discovered that capitalism required “moral and cultural infrastructure.. [that] is learned through sustained commercial practice, and lost when those practices deteriorate” (Rose, 1996). In other words, civil institutions require norms that are long established and often persistent, but once destroyed are hard to rebuild.

Pildes (1996) argues that norms must undergird the formal state for that state to be successful. Laws that support existing norms of cooperation are more likely to be successful, though they may lock in those norms, but laws that violate norms of cooperation are not only likely to be ignored but can undermine the formal apparatus of government. Existing institutions of government may serve as focal points for individual cooperation. Whether the law says we drive on the left or the right or that we go on red or green, these laws help strangers to cooperate. Two strangers who wish to cooperate, where both would rather wait for the other than collide, can easily agree without communication to follow existing laws. Because most people follow the law it is relatively easy to punish occasional violators. Likewise, abstract property rights for shareholders can facilitate the cooperation in complex endeavors so long as the law is consistent with existing social norms. Moreover, good law can embody knowledge either from expert findings or, as in the common law, from the experience of courts (Hayek, 1973). In this instance laws can be internalized as norms.

When those laws interfere with social activity, when the speed limit is far below the preferred speed of traffic, or when developing countries impose property rules that violate existing principles of cooperation, the law ceases to act as a focal point. It becomes impossible to enforce speed limits when everyone speeds. When the government fails to enforce its laws or
deviates so far from existing social norms that it cannot enforce its laws, it not only reduces the impact of that law but it reduces the moral force of the law – the extent to which laws are internalized as norms – and undermines the formal apparatus of government.

V. Conclusion

The behavioralist critique of economics rests upon the unreality of the assumptions of rational choice. Such is the evidence against these assumptions that it is hard to argue that people could ever be regarded as the rational calculators of economic models. Yet this critique ignores the fair protest that rational choice was never intended as a description of psychology or even of markets. It only describes the outcome of all those complex processes – the psychological and the social – that comprise market activity.

The behavioralist critique errs not in its attempt to introduce humanity to economics but in the limited manner in which it has done so. Like rational choice economics it treats humans as utility functions in an algebraic problem. In laboratories, it has injected real humans into the atomistic world of economic modeling and found them to be as out of place as a the hypothetical economic man would be in human society. In doing so the behavioralist model inadvertently ignores the role of social institutions that rational choice captures, albeit just as inadvertently.

Heuristics are not taxes or subsidies on the thinking of otherwise rational calculators, but a distinct method of reasoning employed by humans to overcome cognitive limitations. Likewise, norms are a part of a complex social contract which induces us to make those decisions that are in our own best interests. The norms are imperfect and will sometimes lead us to error. They are persistent and stubborn in the face of change but they also embody a far greater knowledge
than we can individually access. A new economics, if we are to have it, should be based on an understanding of a fuller model of human behavior that includes our social instincts.

Heuristics and models are valuable not for the rightness of their approach but the aptness of their answer and a more circuitous route to the right answer is not more virtuous than a shorter one. As a descriptive theory rational choice is imperfect but so are all models and that imperfection is not fatal. The theory’s convenience and ability to generate testable mathematical predictions have rendered it a useful heuristic. It can be called upon more readily than richer psychological models and economists should continue to use it so long as its predictions, if not its methods, can describe the real world.

Yet there are times when rational choice fails and in those times we must turn to a richer model of behavior despite the challenges that may pose. In proposing such a model I suggest that the law must sometimes treat us as less than rational so that we may have the benefits brought about by impersonal exchange. That idea is only moderately new in economics and is firmly embedded in the law. At the same time, policymakers must be wary that it is social norms that undergird the law and not the other way around. By trying to rewrite social rules policymakers risk disrupting the complex of rules that bind our social interaction, leading us to folly and not reason.
References
References


______ (Sep., 2001) It’s Time to Give Up the War on Drugs. Business Week, September 17.


Eisner, M.A., (2011) “Beyond the Logic of the Market: Toward an Institutional Analysis of
Regulatory Reforms,” In Handbook on the Politics of Regulation, Faur, D.L., (Ed.), Northampton,
MA: Edward Elgar

gradualism,” In Models of Paleobiology, Schopf, T.J.M., (Ed.) , pp.82-115, San Francisco, CA:
Freeman, Cooper and Co.

Harvard University Press

Vol.3 No.1, pp.1-49.

Evans, J.S., (2008), “Dual-processing accounts of reasoning, judgment, and social cognition,”

Evolution and the capacity for commitment, New York: Russel Sage Foundation

“Company Stock, Market Rationality, and Legal Reform, University of Chicago Law & Economics,


Fama, E., Jensen, M., (1983a) Agency problems and residual claims, Journal of Law and
Economics, Vol.26


[A] view of economic growth that depends so heavily on an exogenous variable, let alone one so difficult to measure as the quantity of knowledge, is hardly intellectually satisfactory. From a quantitative, empirical point of view, we are left with time as an explanatory variable. Now trend projections, however necessary they may be in practice, are basically a confession of ignorance, and, what is worse from a practical viewpoint, are not policy variables.

- Kenneth Arrow (1962)

In the wake of the 2008 financial crisis critics have argued that the United States’ market-based finance system is fundamentally flawed. Against this, a significant literature suggests that market-based economies out-innovate and outgrow the major alternatives. Theories of market failure share with theories of growth and innovation the fact that both are theories of knowledge. Any theory of knowledge must also be a theory of its absence: a theory of bounded rationality. Organization - how firms make decisions - and finance - who owns the firm and how they control their agents - are linked to the challenges of innovation and growth.

115 See Infra section 1.
116 Ibid.
I propose a morphogenic model of the firm to combine these elements under a single framework. In biology, organisms (and organizations) are shaped by phylogensis – changes to the organism over multiple generations – and morphogenesis – changes as the organism adapts to its surroundings, as well as other changes that affect the organism’s form over time. I argue that the evolutionary stability of the joint-stock corporation can be predicted by the ability of that organizational form to adapt to the constraints of knowledge, as these later change over the firm’s lifecycle.

I. Finance

The modern critics join a long tradition, beginning with Adam Smith’s (1776) claim that “negligence and profusion... must always prevail” in the management of the joint-stock corporation, through more recent proponents of the German-Japanese model and stakeholder theorem. To this they add a behavioral critique in which irrational markets are responsible for serious misallocations of capital and investor short-sightedness permits unchecked malfeasance by directors of firms and financial intermediaries at the expense of not only shareholders but of employees and consumers (see, e.g., Porter 1992; Bainbridge, 2012).  

The critics have their favored alternatives. Jensen (1989) predicted the public corporation would be eclipsed by the management-owned, debt-financed close corporation. While Jensen, following the Berle- Means critique, thought that shareholders are too weak, proponents of stakeholder theory suggest that shareholder interests must be balanced against the interests of other parties. Firms might also be organized as worker or consumer-owned cooperatives.

117 For an opposing view on the market myopia hypothesis see e.g., Office of the Chief Economist (1987), Hall (1987) and McConnell and Muscarella (1985)
Alternatively corporations might choose to assign voting rights to employees. Yet while these corporate structures are legal, and stakeholder voting is required by law in countries including Germany, partnership and cooperatives are rare in the United States and other common law countries and no large corporation has voluntarily adopted employee voting rights. Common law countries continue to rely heavily on the market-financed joint-stock corporation (public corporation) for the organization of commercial enterprise.\footnote{There may now be greater opportunities than previously for trading shares in close corporations which may serve to lessen the difference between the private and public corporation.}

While the joint-stock corporation may be prevalent in the U.S. several scholars have noted that the Berle-Means form is hardly ubiquitous (Demsetz, 1983; Demsetz and Lehn, 1985; Schleifer and Vishny; Morck et al., 1988). La Porta et al. (1999) demonstrate that most countries do not rely heavily on the widely-owned corporation but on concentrated ownership rights. The authors find that the Berle-Means firm describes large firms in the United States and other common law countries. Using a 20% threshold for close ownership the same is true of the majority of medium-size firms. Nevertheless, while the Berle-Means firm exists in only a few countries, those countries make up most of the world’s stock market value (Ibid).

Most investment in US firms is financed by internal cash flow (Myers, 2001). Even in market-based systems, debt remains more important than equity as a source of external finance (Ibid). In a survey of debt and equity issuances from 1980 to 2006, Covass and Den Hann (2011) found that the relative importance of debt and equity varies considerably with firm size as well as market cycles. The latter finding is shown in Table 1 which indicates that the largest firms (those that fit the Berle-Means model) are largely debt financed. It is smaller firms, which tend to have more concentrated ownership, that also tend to rely more on equity (Ibid; Myers, 2001). Indeed
equity often declines as firms redistribute funds to shareholders while increasing leverage (Myers, 2001).

Table 1 – Debt and Equity as a Source of Finance

<table>
<thead>
<tr>
<th>Size classes</th>
<th>[0, 25]</th>
<th>[25, 50]</th>
<th>[50, 75]</th>
<th>[75, 90]</th>
<th>[90, 95]</th>
<th>[95, 99]</th>
<th>[99, 100]</th>
<th>All firms</th>
</tr>
</thead>
<tbody>
<tr>
<td>N</td>
<td>795</td>
<td>769</td>
<td>772</td>
<td>473</td>
<td>155</td>
<td>132</td>
<td>32</td>
<td>3,128</td>
</tr>
<tr>
<td>Asset Change</td>
<td>26.1</td>
<td>15.0</td>
<td>11.3</td>
<td>9.0</td>
<td>7.7</td>
<td>6.7</td>
<td>4.1</td>
<td>6.9</td>
</tr>
<tr>
<td>Equity Share</td>
<td>0.870</td>
<td>0.532</td>
<td>0.263</td>
<td>0.087</td>
<td>-0.025</td>
<td>-0.106</td>
<td>-0.338</td>
<td>-0.025</td>
</tr>
<tr>
<td>Issue Share</td>
<td>0.724</td>
<td>0.407</td>
<td>0.246</td>
<td>0.188</td>
<td>0.163</td>
<td>0.138</td>
<td>0.145</td>
<td>0.187</td>
</tr>
<tr>
<td>Debt Share</td>
<td>0.109</td>
<td>0.201</td>
<td>0.273</td>
<td>0.310</td>
<td>0.323</td>
<td>0.279</td>
<td>0.201</td>
<td>0.268</td>
</tr>
<tr>
<td>LTD Share</td>
<td>0.236</td>
<td>0.580</td>
<td>0.878</td>
<td>1.034</td>
<td>0.997</td>
<td>0.874</td>
<td>0.784</td>
<td>0.891</td>
</tr>
</tbody>
</table>

Source: Covas and Den Haan (2011). Size class indicates the percentile size class occupied by the firm. Asset change indicates the percentage change (growth) in total assets. ‘Equity share’ indicates the increase in shareholder equity minus dividends as a portion of this change. Issue share indicates the portion to new issues of common and preferred stock; ‘Debt share,’ the portion to change in debt; LTD share the portion to new issuances of long term debt.

Despite the importance of debt in even the United States, the U.S. and other common law countries rely far more on public issuances of equity than other nations, including other developed nations. Table 2 shows the ratio of stock market capitalization shown in (held by minority investors) in the United States and selected other nations in 1996, reported by La Porta et al. (1999).

The authors’ data shows that US equity to GNP is 0.58 while the ratio of debt to GNP stands at 0.81. Yet while debt predominates, even in the U.S., the respective ratios in France are 0.23 to 0.96, in Germany, 0.13 to 1.12, and in Japan, 0.62 to 1.22. The data agrees with the later finding by Myers (2001) and Covas and Den Haan (2011) that debt predominates but it also indicates that common law countries including the US rely far more heavily on external equity with dispersed ownership than other developed nations.
Table 2 - External Capital Markets

<table>
<thead>
<tr>
<th>Country</th>
<th>External Cap/ GNP</th>
<th>Domestic Firms/ Pop</th>
<th>IPOs / Pop</th>
<th>Debt / GNP</th>
<th>GDP Growth</th>
</tr>
</thead>
<tbody>
<tr>
<td>Australia</td>
<td>0.49</td>
<td>63.55</td>
<td>-</td>
<td>0.76</td>
<td>3.06</td>
</tr>
<tr>
<td>Canada</td>
<td>0.39</td>
<td>40.86</td>
<td>4.93</td>
<td>0.72</td>
<td>3.36</td>
</tr>
<tr>
<td>UK</td>
<td>1.00</td>
<td>35.68</td>
<td>2.01</td>
<td>1.13</td>
<td>2.27</td>
</tr>
<tr>
<td>US</td>
<td>0.58</td>
<td>30.11</td>
<td>3.11</td>
<td>0.81</td>
<td>2.74</td>
</tr>
<tr>
<td>France</td>
<td>0.23</td>
<td>8.05</td>
<td>0.17</td>
<td>0.96</td>
<td>2.54</td>
</tr>
<tr>
<td>Italy</td>
<td>0.08</td>
<td>3.91</td>
<td>0.31</td>
<td>0.55</td>
<td>2.82</td>
</tr>
<tr>
<td>Germany</td>
<td>0.13</td>
<td>5.14</td>
<td>0.08</td>
<td>1.12</td>
<td>2.60</td>
</tr>
<tr>
<td>Japan</td>
<td>0.62</td>
<td>17.78</td>
<td>0.26</td>
<td>1.22</td>
<td>4.13</td>
</tr>
</tbody>
</table>

Source: La Porta et al. (1999)

It may be, as Roe (1993; Roe and Gilson, 1993; Bebchuck and Roe, 1999) claims, that each country’s choice of finance depends upon legal, cultural, and historical accident. However, La Porta et al. (1997) find that this greater reliance is the result of stronger investor protections afforded under the common law. Mahoney (2001) has found that common law countries grow on average 0.71 percent per year faster than code law countries. La Porta et al. (1996) found that better investor protections are responsible for deeper capital markets in common law countries. King and Levine (1993 a,b) find that depth of capital markets is a determinant of growth and that countries with more developed finance systems experience faster GDP growth, physical capital accumulation, and more efficient allocation of capita. While the ubiquity Berle-Means equity-financed corporation may have been exaggerated, equity and dispersed ownership remain an important part of the economic landscape of the United States, and may be associated with higher economic growth.

II. Organization

While Smith (1776) introduced the division of labor he remained skeptical of one of that most significant division between capitalist and manager. Adolf Berle and Gardiner Means (1932)
reintroduced the problem of separation of ownership and control to modern economics, claiming that diffuse ownership left control of the corporation in the hands of management. Managers have an interest in avoiding work the owner would have them do, or in using the firm’s resources to their own benefit.

If the manager simply steals from the company then this is embezzlement and clearly illegal. In countries where investor protections are weak, this sort of direct theft by managers or majority shareholders is common and firms cannot easily raise capital from markets (Schleifer and Vishny, 1997; La Porta et al., 1996, 1999). Indeed, much of the earlier legal precedent in the West addresses this problem (Schleifer and Vishny, 1997). Yet while modern American corporate law affords investors substantial protection compared with many foreign regimes, the business decision rule still gives greater deference to the decisions of managers and boards than is typically afforded by US law (Easterbrook and Fischel, 1991). If a manager spends the firm’s resources on a private jet rather than a new factory, when the latter would generate a better return for investors, the law will generally protect him (Ibid.).

This conflict might be avoided if it were possible for investors and managers to perfectly contract with one another and if such contracts could be easily monitored and enforced. But it is often unclear ahead of time what tasks a manager will be expected to perform. It is prohibitively costly, if not impossible, to produce extensive contingent contracts for every circumstance. One possible alternative is for firms to use spot contracts for managers and other employees.

When investors and managers, or firms and employees, write incomplete contracts, one or both parties must make specialized investments. This creates an extractable quasi-rent: the
difference between the rental value of capital in its specific use and the next most valuable use (Klein et al., 1978). Firms may require specialized machinery or facilities which are less valuable in their next best use. Both employer and employee invest in training to develop human capital which makes the employee more valuable but they cannot agree in advance how that capital will be utilized by the firm or at what rate of pay. Once parties have made specific investments this incomplete contracting leads to potential conflict (Ibid.).

Williamson (1975) argues that it is impossible to of write complete contingent contracts, forcing firms to hire employees through relational contracts. These contracts permit managers to direct the activities of employees within reason. Managers and other employees must still be compensated but instead of attempting to write ex ante contracts for performance, firms can instead use ex post settling up. A manager or specialist who performs poorly in an unexpected situation may lose his job or be denied promotion. By contrast an employee who exceeds her contractual obligations may be granted rapid promotion or a bonus. If labor markets are efficient and current performance is the best indicator of future performance then ex post settling up can overcome the incomplete contracting problem. But it is harder to compensate an employee for an exceptional performance; such as the effective handling of a crisis that is unlikely to be repeated. Moreover, employees, and especially managers, may have firm-specific human capital making it harder to relocate. Firms must rely then on reputation and repeated interactions with employees to guarantee rewards.

Relational contracting can potentially reduce the cost of writing contingent contracts, but investors and managers must still monitor subordinates. That problem is likely to be most acute for investors who, in a public corporation, are a usually a diffuse group. Monitoring managers is
costly and, whereas the investor who monitors the manager bears the entire cost, she only reaps a small share of the benefit. Moreover, managers have greater local and specialized knowledge about their company than do arms-length investors. While the owner of Adam Smith’s pin factory could easily see if his employees were producing pins, Fama and Jensen (1983 a,b) argue that it is far harder to determine whether or not a manager is performing his function. Indeed, if the manager has been hired for his specialist knowledge then it is self-defeating to attempt to direct his every action.

The same conflict is repeated between managers and the specialists they appoint. Again, this problem would be solved with complete contracts but such contracts are frequently impossible. The specialists’ output is combined with that of a team and individual performance may be hard to measure (Alchian and Demsetz, 1972). If the team is tasked with innovating or other problem solving tasks where output is not a perfect indicator of effort, then it will not be possible to write or enforce complete performance-based contracts.

Klein et al (1978) argue that when such conflicts occur between firms, they can be resolved by vertical integration. Williamson (1975) similarly argues that hierarchy can reduce conflict within a firm. Firms cannot own their employees but employees can own firms in the form of partnerships or worker cooperatives. This may potentially resolve the conflict between workers and management but leaves open a conflict between workers and the owners of capital. The owners have also made specific investments over which they must now relax control, leaving workers free to expropriate quasi-rents. Klein et al. therefore argue that for such arrangements to become feasible workers must also become capitalists. But this may result in an inefficient distribution of risk: not only do workers lose any potential gains that may accrue form
comparative advantage in risk bearing, but if labor markets fail to clear then workers bear additional risk if the firm is forced to retrench.

A similar logic must apply to the relationship between consumers and firms which may often be characterized by asymmetric information. Firms know more about the quality of their product than do consumers and as before, writing and enforcing complete contracts is usually impossible. Akerlof (1970) notes that this may sometimes completely inhibit the formation of markets. The relations between firms and consumer may also be asymmetric when the firm has market power. This allows the firm to charge above the competitive equilibrium price. Again this conflict may be resolved if consumers own the firm (Enke, 1945).

Klein et al. (1978) argue that government ownership of utilities is a particular case of vertical integration in which consumers indirectly own the firm. It is not only the consumers who must worry about natural monopoly but the investors in utilities who must worry that once they invest in plants to provide water or power, that consumers will vote to regulate the price and the investor will be unable to recoup the cost of his investment. Consumers could become the owners without government, raising the cost of capital through their own savings or on debt markets, but if this arrangement exists it is not popular. Both government and other forms of consumer ownership suffer from the same problem of diffuse owners who may fail to properly monitor managers; but government at least has the power of taxation to fund debt obligations.

In the mid-1980s, scholars proposed an alternative, albeit somewhat amorphous, alternative corporate structure: the stakeholder corporation (Taylor, 1971; Freeman, 1984; and more recently Walsh, 2005; Sisodia et al., 2007; Harrison et al., 2010). Instead of responding only to the interests of shareholders, management should balance owners’ interests with the interests
of consumers, workers, and other stakeholders in the corporation.\textsuperscript{119} While the idea has been popular with some academics and policymakers, it is has proven difficult to implement; not least because its proponents cannot define precisely what balance should be struck (Jensen, 2000; Bainbridge, 2012). Jensen (2000) argues that stakeholder theorem should be interpreted to imply welfare maximization. Under the standard conditions of neoclassical economics this becomes indistinguishable from profit maximization although Jensen has elsewhere noted that it is unlikely that perfect monitoring would ever exist (Fama and Jensen 1983a,b; Jensen and Meckling, 1976).

By contrast with the ambiguous stakeholder theory, bank-ownership appears to offer a viable alternative form of finance. The two most notable incarnations of this finance, Germany and Japan, not only appear to solve the monitoring problem through financial intermediaries but suggest that alternative forms of ownership involving labor and customers are also possible. Those countries once seemed poised to wrest America’s dominant economic position from her and consequently received intense scrutiny from American academics. Banks in both countries can exert influence on corporations in two ways: as lenders and as stockholders in the firm. Although German and Japanese banks hold a much larger share in major companies than the 5% maximum allowed the American law, they do not usually hold enough stock to hold complete control over the company (Roe, 1993).

In Germany, most corporations, with the support of banks, prohibit any individual shareholder from voting more than 5% of the firm’s stock. However, this restriction does not apply to a bank voting multiple stock holdings held in trust by the bank’s brokerage arm (Ibid.; Baums, 1992). By

\textsuperscript{119} Harrison et al., (2010) emphasize the distinction between managing the firm for the benefit stakeholders and broader concepts of social responsibility.
contrast, Bank Holding Companies in the United States are not only restricted to a 5% holding but must be passive shareholders. Concentrated holdings give the bank more power to influence managers and potentially to eliminate conflicts. The ability to vote stock not owned by the bank eliminates one problem of diffuse ownership: the individual small voter has a low probability of successfully ousting management. It does not obviously solve the other problem: the bank only benefits from improved management up to the extent of its own holdings. Banks may be able to bundle good monitoring into the brokerage services to mitigate this problem. However, as a lender, the bank might be more risk averse than the other stockholders, including the bank’s own clients.

By law, German corporations have two boards. One is controlled by management while the other – the supervisory board – is controlled equally by shareholders and workers. Banks are usually well represented on the supervisory board. The role of shareholders on the supervisory board is replicated by independent directors on American boards, although it is not clear how well those represent shareholder interests (c.f. Bainbridge 2012). Employee representation has no analog in the American corporation and may eliminate some conflict. However, German corporations have been unable to produce the same shareholder value as their market-financed counterparts. Despite its name the supervisory board is purely advisory, and is often kept intentionally weak by shareholders and their representatives who can use concentrated holdings to directly influence management.

---

120 The banks do not actually charge for voting a client’s shares but do charge for brokerage services. This model might be justified because voting the additional shares has no additional cost to the bank and benefits other customers. In this way the bank avoids a free rider problem among shareholders but this does not explain why customers cannot go to brokers who do not offer this service.

121 If tied, shareholders may cast the deciding vote. Shareholders can also override supervisory boards.
The Japanese financial system also relies heavily on bank financing. While multiple banks usually hold shares in the corporation and are lenders to the same, monitoring falls directly to a main bank with the support of the other banks. As with the German corporations, the banks do not hold enough shares to directly control management. The Japanese banks are generally less involved unless there is a crisis and monitoring falls to Keiretsu: coalitions of firms with cross-holdings in one another. While ownership is concentrated no firm holds enough shares to oust management but a coalition can do so more easily than can diffuse shareholders. A Keiretsu can intervene before crisis would typically force the American corporation into costly bankruptcy proceedings. Gilson and Roe (1993) argue that the Keiretsu sit somewhere between contract and vertical integration. The members are often also business partners and cross-holdings can protect members who make specific investments from defection by another member without the need to make detailed contracts. Unlike complete vertical integration, membership is somewhat fluid: a poorly performing company (division) can be ousted while an outperforming company can elect to leave.

The Keiretsu resolve conflict between firms, and may have allowed Japanese corporations to refine just-in-time manufacturing processes, but at the cost of making outside shareholders weak. To be successful Keiretsu must allow Japanese corporations to resolve some differences between the interests of shareholders in the firm and the interests of other firms in favor of other group members, but there is no guarantee that management will not use this feature to their advantage. Whatever coordination gains may arise from Keiretsu the overall impact is lower return to shareholders (Prowse, 1992).\(^\text{122}\)

---

\(^{122}\) Measured as lower profitability to book-value.
Despite the apparent advantages in monitoring, neither German nor Japanese firms fulfilled their promise of dominating American rivals. Nor did they deliver the same returns to shareholders (Ibid.). Their rapid growth seems to have been a consequence of idiosyncratic circumstances. Both emerged from World War Two having experienced significant loss of capital but also with newly imposed stable liberal democratic regimes. Opportunities exceeded available investments and, either because market institutions were undeveloped or mistrusted, corporations preferred to rely on banks to deliver the necessary capital inflows (Ibid.). It is possible even that the system was made necessary by weak property rights which made shareholder concentration necessary (c.f. Schleifer and Vishny 1997).

As investment opportunities diminished, the corporations became less reliant on banks whose monitoring capabilities likewise diminished. The inability of the corporations to deliver the same returns as their American counterparts created pressure for reform though neither have come fully to market-based systems. Both, though more Japan, continue to experience an economic sclerosis. If there is any convergence between the American and German-Japanese systems it has come largely from reforms by the latter. The activist financial intermediaries which emerged in the United States in the 1990s, declined in the 2000s while Germany witnessed the emergence of the hostile takeover (Bainbridge, 2012; Boehmer, 1998; Jackson and Höpner, 2001). Unlike the German and Japanese banks, where the firm’s need for capital created an ongoing lending relationship with client firms, the American intermediaries could not solve the free rider problem. With minority shareholdings intermediaries could only capture a small portion of the gains of monitoring management. Investors who followed the more traditional axiom of supporting management or selling could give investors a better deal.
In fact much of the shareholder activism of the 1990s was driven by minority shareholders precisely because the shareholders were only marginally exposed to the cost of policy (Bainbridge, 2012; Romano, 2001; Choi and Fisch, 2008). Public pension funds such as CalPERS and labor unions sought to use their position as minority shareholders to pass ballot measures which aided their members to the detriment of other shareholders. Activists also used ballot measures to promote a social agenda which would similarly have harmed the financial interests of other investors. When corporate constitutions allow minority shareholders access to the proxy machinery it can prove expensive for shareholders even when the measures are not approved.

This may explain why most corporate charters, and default corporate law, place some restrictions on outside shareholders using the proxy machinery while allowing the incumbent board to use corporate funds to fight off shareholder challenges (see e.g., Easterbrook and Fischel, 1991). Yet managers are far from being perfect representatives of investor interests. Scholars have identified a number of ballot measures, usually anti-takeover devices, which reduce the value of the firm (Jarrell and Poulsen, 1987). Institutional investors often break the conventional rule – vote with management or sell – to oppose anti-takeover measures, but such measures frequently pass anyway (Easterbrook and Fischel, 1991). Jarrell and Poulsen (1987) found that measures are more likely to pass when insiders hold more of the stock, and when institutional investors hold less.

Although managers are more likely to succeed when they own stock, managers do not own enough stock to vote down all challenges. If they did there would be no reason to erect procedural barriers to takeover bids. Shareholders who approve these measures do so at a
detriment to themselves are presumably less informed than institutional investors and do not actually mean to reduce their wealth. Smaller investors can free ride off market prices when buying and selling but not when voting. Investors may then be willing to pay more for stock in corporations whose charters restrict their ability to vote and bring measures.

Given the ineffectiveness of voting as a measure of control, it is reasonable to ask why shareholders should vote at all. Jensen (1989) predicted that Management Buy-Outs (MBOs) and Leveraged Buy-Outs (LBOs) with concentrated ownership supported by debt, and potentially securitized debt holdings would replace the public corporation in most venues. In these corporations, management’s share of residual claims is considerably higher. Rather than an unspecified share of profits coupled with control rights, the debt holders are given guaranteed returns but no say in the firm’s management. The combination of managers’ increased ownership, and the stringency with which debt obligations are enforced, holds managers accountable to a far greater extent than the traditional public corporation. Yet while debt continues to play important role in corporate finance, history has yet to bear out Jensen’s prediction of the “eclipse of the public corporation.”

Debt may be too stringent. Managers cannot control all variation in the firm’s profitability and while shareholders can be denied their dividend, defaulting on debt has legal consequences. The debtor firm may be able to meet some of its obligations, but not all. Creditors can become engaged in a wasteful race to stake their claim ahead of other creditors (Zywicki, 2008). In some instances this race leads to the firm’s assets being broken up and sold for less than the value of the firm as a whole. Chapter 11 bankruptcy can provide a solution to this problem but shareholders lose substantially in bankruptcy (they are not usually wiped out) and have little
incentive to start proceedings until necessary (Aghion et al., 1992; Hart, 2000). Creditors have no power to compel bankruptcy before a default at which point the firm’s assets may have already been needlessly squandered in the attempt to rescue shareholders. Consequently, debt is most frequently available only when the firm can provide collateral. Riskier ventures require a cushion of equity (Schleifer and Vishny, 1997).

Managers might provide this additional equity but this exposes them to greater risk. This reduces the benefits of diversification and specialization in risk bearing found in capital markets. The remaining outside owners must compensate managers for this risk with inefficiently high wages. Instead shareholders have increasingly turned to independent boards (Bainbridge, 2012). In addition to increasing the board’s independence from management, corporations have used stock options to align directors’ interests with those of shareholders. Stock options provide incentives while somewhat insulating directors from downside risks. Jensen (2000) argues that this trend is efficient and ultimately achieves the effect he predicted. Nonetheless, while investors can observe the firm’s overall profitability and this may be reflected in the share price, managers can artificially increase profits by taking on excessive risk or by depleting the firm’s capital. If this behavior can be hidden from shareholders then options may lead boards to excessive short-termism and create systemic risk.123

Yet while boards monitor, shareholders still vote. They vote overwhelmingly in favor of management and to reappoint incumbent directors (Easterbrook and Fischel, 1991; Jarrell and Poulsen, 1987; Choi and Fisch, 2008; Bainbridge, 2012). Despite this loyalty, shareholders pay a

123 As insider trading laws only prohibits trading on significant events, not overall confidence in the firm, legal insider trading may mitigate this problem by allowing management to pass information to markets (Manne, 1967; Bainbridge, 2012).
premium for the right to vote and the law makes voting rights inalienable from ownership.\textsuperscript{124} Voting makes no guarantee of protections for small shareholders. If the law or corporate charters did not forbid it a shareholder with the barest majority could loot the firm of its assets.\textsuperscript{125} Indeed, Schleifer and Vishny (1997) note that this happens in most of the world and was relatively common in the early history of the American corporation. As such, corporations cannot easily raise finance, it is in the interests of all shareholders to ensure against such looting. Thus corporate charters should include restraints against both large and small shareholders, but still to provide some protection against management excess.

In the banking based systems, banks rarely intervene except in crises. So too should American shareholders prefer passivity only occasionally punctuated by activism. This is achieved through the takeover bid which may be used in crises but also in the situation which Germany and Japan did not encounter until the 1990s: a firm sitting on capital which may be better deployed elsewhere by investors. While the takeover is more versatile than bank intervention, it is not without difficulty. The bidding firm must expect to profit from the deal but if shareholders are free to reject the tender offer then they can free ride off any improvements under the new regime. If the bidding firm could expropriate the firm’s wealth with a simple majority it could easily capture all the benefits of monitoring but such a law would also permit looting. Instead, the law or corporate charter must allow bidding firms to freeze out small minorities once some portion of shareholders accepts the bid.

\textsuperscript{124} That is, in the most common case of a share that carries one vote that vote cannot be permanently sold. Alternative stock issuances with several or no votes are legal.

\textsuperscript{125} Delaware law does indeed impose a fiduciary duty on majority shareholders and corporate charters may make additional requirements.
III. Evolution

Williamson (1975) argues that the organizational challenges facing the modern firm can only be explained by a combination of conflicts between stakeholders, bounded rationality, and uncertainty. Without uncertainty it would be possible for each party to perfectly predict the future and write contracts accordingly. Without bounded rationality it would be possible to write extensive contingent contracts for all possible scenarios and to perfectly monitor all parties ex post.\textsuperscript{126} The theory of rational expectations theory (Muth, 1961; Lucas, 1972) suggests it is reasonable follow Simon’s (1972) lead in treating uncertainty as a form of bounded rationality.

Continuing with Simon’s treatment, bounded rationality may be separated into three (somewhat distinct) categories. First, individuals and organizations may not have full information about the state of the world. Consequently, any decision must be separated into actual choice, the selection of options, and the decision to bear the cost of search (March and Cyert, 1963; Nelson and Winter, 1982). Williamson (1975) also includes the inability of language to convey information as precisely or completely as may be desired. For the individual this represents a limitation in information availability but for the organization qua economic agent it represents a limitation of the second type: the limited processing power available to decision-makers (Simon, 1957).

Simon (1957) imagined a limitation on the processing power available to individuals: even grandmasters cannot perfectly solve the game of chess. Later studies in mathematical

\textsuperscript{126} It seems plausible that such extensive contracts would involve physical transaction costs also (Coase, 1937). Nonetheless, this only shifts the cause of frictions and the remainder of Williamson’s argument remains intact.
economics demonstrate that not all economic problems have computable solutions (Lewis, 1985, 1992) meaning that there is no algorithm than must solve the problem polynomial time. Similarly, the problem of uncertainty can be explained now in terms of dynamic complexity such that rational expectations cannot provide a complete response (Rosser, 2009). Under bounded rationality decision-makers cannot optimize in the manner of economic man. Simon (1957) argues they must instead satisfice: pick the best option out of the available alternatives. The idea of satisficing has found some measure of support in both advances in computational complexity theory (Velupillai, 2009) and in studies of social and evolutionary psychology (Gigerenzer and Goldstein 1996, 1999).

This broad definition of bounded rationality can not only explain why firms may vertically integrate but also the limits on integration and why firms are organized as they are. As firm size increases, it becomes harder for shareholders to monitor management and for management to monitor lower levels of organization (Williamson, 1975). Not only is it more difficult to measure the output of specialists (Fama and Jensen (1983 a,b) but the costs of communication within the firm rise with additional levels of bureaucracy.

It also becomes harder to centralize decision making. Hayek (1937, 1945) argues that central planners lack time and place knowledge: specific knowledge that cannot be centrally collected. Though Hayek sought to describe the limitation of a government the theory applies equally to the constraints facing firms as the grow larger, and boards lack specific knowledge about arbitrage opportunities, unexpected costs, or delays, that lower-level employees may be aware of. Management may take advantage of this knowledge by devolving some decision-making to those employees but doing so increases monitoring costs. Where the logic of conflict seemingly
supports considerable economies of scale, Hayek’s argument implies that bounded rationality makes central control more expensive than markets, where decision-making is entirely devolved to the individual. There must be a balance between rising costs of central planning, and the falling costs of conflict, that come with firm size (c.f. Coase, 1937).

While firms cannot avoid this tradeoff they can nonetheless seek to minimize it. As firm size grows consumers and employees are more likely to have heterogeneous interests. It is difficult to measure consumer and worker surplus. Consequently, worker- or consumer-controlled enterprises are liable to have higher monitoring costs than shareholder controlled enterprises and to miss potential gains from integration. While some proponents of stakeholder theorem have claimed that non-owner stakeholders are more likely to reveal private information to the firm when the firm is managed for all stakeholders (Harrison et al., 2010), this ignores the problem of conflict between those stakeholders. Shareholders have more (though by no means entirely) homogenous interests and it is easier to measure profit than consumer or worker surplus.

Williamson notes that if all employees are to communicate with one another the costs of coordination will rise exponentially with firm size. Hierarchical control structures can mitigate these costs but not entirely. Boards who lack time and place knowledge will want to devolve decision making to lower levels, but these employees will lack even the perspective available to the board about the broader firm’s operations. Thus while conflict alone cannot explain inefficiencies, some misallocations may be attributed to bounded rationality even in the absence of conflict.
Given the limitation of bounded rationality Simon (1957) argues that organizations must seek to align the behavior of individuals with the broader interests of the firm by editing the choices available to the decision maker. Again, more recent studies support Simon’s proposition that institutional design can align individual behavior with economic rationality (Jolls, Sunstein, and Thaler, 1998; Sunstein and Thaler, 2003; Thaler and Sunstein, 2003, 2008; Gigerenzer and Goldstein, 2002; Todd and Gigerenzer, 2007). In the case of the firm, the goal is not to align the individual’s behavior with her own interests but with those of the firm. Simon (1972) argues that just as humans can be treated as organizations at the biological level, so too can organizations be treated as individuals. Thus Simon (1957, 1972) and Smith (2003) both argue that organizations themselves can be judged against economic standards of rationality.

Nelson and Winter (1982) propose a model of the firm built upon routines, or rules, which govern interaction between employees. These rules, which are discovered and not designed, are akin to an organizational technology. Being perfectly specific to the firm they cannot be sold or claimed by debtors. Thus the associated quasi-rent constitutes the entire value of the rule. Hayek (1973) claims that rules themselves can embody tacit knowledge from experience. Thus it is possible that no individual in the firm knows how the entire firm works, or why a particular process is used, or indeed anything beyond their own specific role and not even the reason for that. The idea that routines can store knowledge has since been demonstrated in the empirical literature (see Becker 2004 for a general review).

Knowledge need not be stored as specific actions but as decision-heuristics which align individual behavior with the interest of the firm. Individual employees may still act within the framework of routines to take advantage of time and place knowledge but they are not asked to
consider the sum consequences of their actions on the firms. Instead salesmen may have the
goal of increasing sales while safety officers have the goal of promoting safety. In this manner,
firms can seek to monitor specialists through job-specific heuristics. Management routines must
balance these goals, at least implicitly, such that the overall effect is to maximize the likelihood
of firm survival.

Although the firm may appear to have abandoned the goal of profit maximization, relying on
rules may confer an evolutionary advantage on firms. In human evolution, total plasticity –
making perfect rational calculations in each moment – equates to evolutionary death (Cosmides,
1989). Not only would this method require expensive calculation but much cooperation and
coordination is built upon deviations from rational calculation. In humans, commitments to
values such as fairness, can overcome the need for the costly enforcement mechanisms implied
by rational choice economics (Sen, 1977; Ostrom and Gardner, 1992; Trivers, 1971; Nesse,
2001). These commitments permit humans, not only in laboratories but also in field studies, to
overcome problems such as one-shot prisoner dilemma games without relying on repeated
interaction.\textsuperscript{127}

Many of the coordination problems faced by the firm exceed the limits of possible calculation
Firms, like humans, can also make commitments which allow for \textit{ex post} settling up with
employees, or increase trust among consumers. Indeed, the actual practice of the firm may be
better described by the amorphous and shifting values found in stakeholder theorem, which
emphasizes the importance of trustworthiness in reducing conflict and promoting the transfer
\footnote{\textsuperscript{127} With perfect rationality, reputation might explain some of this behavior but at least in laboratories
even the reputation mechanism seems unnecessary.}
of private information (Harrison et al., 2010), than the profit maximization principle found in law. This gap between actual practice and legal theory may be explained in part by the difficulty faced in translating informal agreements into formal contract or law, and in part by the difficulties presented by unanticipated contingencies.

While Simon (1957) argues this rationality can be achieved by design, such a designer might lack the complete knowledge needed to create such a schema. By contrast, Alchian (1950), Becker (1976), and Smith (2003) all argue that evolutionary forces can select firms which behave as though they were rational, even if there is no designer available. Biological evolution acts through phylogenetic processes: genes are selected based on the ability of the organism, or phenotype, to survive and breed. Evolution may work upon routines by selecting over organizations but, while there is no organizational analog to breeding, evolution may also work on the morphogenic process within organizations (Axelrod, 1986). Firms may select between routines, provided some comparison heuristic is possible, or successful routines may be imitated and spread in this manner. Both forms of selection are somewhat limited in that the optimality of a routine may depend upon the other routines adopted by the firm such that only selection over the entire firm can determine the optimal bundle.

As Alchian (1950) notes, evolution selects for survival and not for the economists’ optimization function. Once established a firm may receive quasi-rents from some innovation, or even from having discovered valuable routines to align its own behavior with rationality, which would allow its survival even with some deviation from rationality. Furthermore, the surviving routines

---

128 Similar processes may work on the selection of heuristics by individuals.
129 To adapt a rather simple example from Dawkins (1976), a routine for communicating technical information in German will be successful in a firm in which most other business is conducted in German but not a firm where business is conducted in English and few employees speak German.
must have first survived through disequilibrium before the market converged upon its static state. As firms are not usually created as large organizations, large firms will tend to arise from small firms that were most successful in disequilibrium (Nelson and Winter, 1982). Hannan and Freeman (1977) note that those larger firms will not be isomorphic to the original successful firm but have been altered by the process of growth to become a fundamentally different organization. The morphogenic process by which firms modify their routines with growth does not guarantee that the alterations to the firm through growth will necessarily may the firm fitter.

Arthur (1989) notes that technological growth tends to be path dependent: that is those technologies that are initially successful are developed in small leaps while alternate, and potentially superior, technologies remain undeveloped. Routines, as a social technology, follow a similar pattern (David, 2000; North, 1990; cf. Becker, 2004). Levitt and March (1988) argue that firms may be caught in competency traps, where once they have developed a specific set of routines it becomes too costly to change. Empirical evidence suggests that firms can indeed become locked into processes (Becker, 2004; Levitt and March 1988). While firms can gain from imitation during their formation, gains from imitation are small once the firm is established and first movers can suffer a disadvantage by becoming locked into inferior processes.

The firm need not be replaced despite its inefficiency because routines are selected from the available pool rather than an infinite series of options. Indeed, any evolutionary model that omits a process for the discovery of new routines should converge to a static equilibrium

---

130 Though large firms may sometimes create entire divisions to enter new markets.
without innovation (Nelson and Winter, 1982). In the case of selection over firms this also implies a significant tendency toward market concentration (Ibid.). By contrast, if the cost of imitation is low and there are some scale economies then large firms can shut out smaller rivals by imitation but in the process destroy the incentive to innovate in the first place (Ibid.).

IV. Innovation

If economic development were to be achieved by evolutionary selection alone then firms should trend toward monopoly and stagnation. While this trend may have once appeared plausible such is no longer the case. Firms continue to fail and be replaced by new firms embodying new routines (Beinhocker, 2006). Some modern evolutionary theorists have thus shifted away from static regimes towards dynamic models of innovation (Ibid). In doing so they do not so much discard the understanding of rules and routines but seek to understand how routines can adapt to change.

The tacit knowledge developed from experience and embedded in the firm’s routines stands at odds with the time and place knowledge available to each employee. Routines are necessary to allow firms to be managed effectively but overly rigid routines will fail to take account of uncertainty. The firm must thus strike a balance between these two knowledge problems, allowing sufficient plasticity to adjust without discarding the value bound up in routines.

---

131 For a relatively simple description of this process see Fagerberg (2002)
In biology, niche theory predicts that a specialist will out-compete generalists over a narrow range of outcomes but that generalists will have the advantage when there is greater uncertainty. This claim is illustrated in figure 1 where the fitness of a generalist and a specialist are both represented over a single environmental variable. The specialist is more fit over the range $m$ to $n$ (the specialist’s niche). When environmental conditions are stable, the specialist will prevail but under uncertainty the advantage falls to the generalist. The terms generalist and specialist are relative: a mouse cannot be so generalized that it can become a hawk. As Cosmides (1989) notes, complete plasticity even in problem solving would be evolutionary death. Hannan and Freeman apply this model to firm’s routines arguing that organizations should similarly respond to incentives to become generalists or specialists depending upon uncertainty. The authors further note that by making routines more rigid the firm may improve
efficiency but strong inertial pressures (path dependency) which may prevent a firm reforming even to avoid complete failure.

Nelson and Winter (1982) argue that innovation can itself be routinized. Simple innovation may occur by a worker or manager identifying a problem and resolving it through his own expertise, or through trial and error. Once he finds a solution it becomes routine should the problem occur again. In addition to determining how to solve the problem routines may determine who solves the problem and so avoid conflict within the firm. Routines can aid learning but empirical evidence suggest that routines themselves are not inert (see Becker, 2004 for a general review). Instead routines are a source of both stability and flexibility.

Scholars distinguish between single-loop learning in which the firm seeks to improve the efficiency of current processes and double-loop learning in which the firm may actually alter its routines (Becker, 2004). To do so the firm must have higher routines which determine the process by which a routine is altered. While Nelson and Winter suggest that routine would be ineffective when the firm is faced with an entirely novel problem, more recent research in neuroscience has found that similar rule-based processing in humans can cope with novel problems though the result is not guaranteed to be optimal.

Higher routines will first determine whether existing practices merit change. Absent perfect information and processing power, Simon and March (1958) argue the firm must have some heuristic to determine whether existing results are satisfactory. If the firm decides to change it must then engage in search, again determined by higher routines, and finally in choice using yet another set of routines. Indeed, March and Cyert (1963) find empirical support for such a
process albeit from a small survey. By altering routines only when confronted with problems the firm can avoid damaging beneficial routines or expending unnecessary resources on search.

With routinized innovation the firm can adapt to changes in its external environment or alter the environment by the development of new technologies and routines. Managers and employees within the firm can be considered entrepreneurs in the sense used by Hayek and Kirzner as arbitrageurs of time and place knowledge. As the firm travels through disequilibrium they respond to changing external circumstances and refine the firm’s routines. Empirical evidence suggests that this process of refinement, or learning by doing, can substantially improve efficiency but that diminishing returns set in over time. This was first demonstrated in the aircraft industry (Wright, 1936; c.f. Alchian, 1963) but has since been demonstrated in a range of other industries (see reviews by Yelle, 1979; Argote and Epple, 1990).

The firm’s learning curve typically takes the form of a power function:

\[ h_n = h_1 n^{-\gamma} \]  

[1]

Where \( h_n \) is the number of hours to produce each good; \( n \) is the number of goods produced; and \( \gamma \) is parameter for the rate of learning. When \( \gamma < 1 \) there are diminishing returns to learning. The rate of learning may also be expressed as a progress ratio (p).

\[ p = 2^{-\gamma} \]  

[2]

In their review of studies Argote and Epple (1990) find a wide range of values for different industries with a modal value of \( p = 0.815 \), which corresponds to an approximate value of \( \gamma = 0.3 \). The authors also find that learning can depreciate over time when knowledge is not used. However, it is not always necessary to maintain constant personnel. The authors note that
World War II shipyards achieved reductions in labor of 45% and reductions in build time of 75%, despite turnover rates in excess of 10% per month (Ibid.). By contrast, Reagans et al. (2005) found distinct components for individual learning and organizational and team learning among doctors. The latter suggest that this difference in the importance of individual learning may depend upon the complexity of the job, the level of individual skill required, and the need to adapt under greater uncertainty. Argote and Eppele (1990) suggest that the ability to transmit learned information to new workers is an important variable.

The ability to transmit routines to new operations may be highly valuable. In a study of the printing industry, Knott and McKelvey (1999) compare franchises with independent stores and company-owned establishments. The authors find that both residual claims and routines affect firm efficiency but the coefficient on routines is 12 times higher than that for residual claims and that this result holds when the model controls for system-wide economies of scale.¹³²

Firms may seek to routinize the discovery of new technologies though research and development (R&D). Yet while March and Cyert (1963) find that firms ordinarily institute reform to solve problems, and hence are most likely to institute change during poor performance, spending on research and development tends to increase when firms are more successful. The authors attribute this to an increase in organizational slack brought about by improved performance. In other words, the firm has greater access to quasi-rents which are not redistributed to shareholders but are consumed by the internal bureaucracy. Some portion of this slack may be taken by consumption of perquisites or simply by retention of inefficient

¹³² 4 times higher when controls are omitted from the model.
procedures but firms may also expend resources on R&D or developing new markets or corporate divisions.

These expenditures may be beneficial but are not necessarily so. That managers prefer to concentrate on corporate empire building rather than return additional revenues to shareholders does not imply that shareholders benefit from this. Indeed, the financial history of the latter portion of the twentieth century suggests that markets prefer to limit organizational slack, though not completely, by returning money to shareholders and by increasing debt funding relative to equity.

Schumpeter (1934) argues that innovation is more likely to be performed by entrepreneurs outside of existing firms though he also claims that increased routinization of innovation will eventually bring that process within firms. He later argues that this indeed has happened (Schumpeter, 1956). Monopoly rents may create incentives to innovate (as patent laws assume is necessary) but capitalists can redistribute these returns afterwards to the most effective innovators. Incumbent monopolists need only conduct new research if those firms have a comparative advantage in research or, more plausibly, if the monopolist would be the sole possible buyer for the new technology.

While some innovation clearly exists within firms, Schumpeter (1934, 1956) defines entrepreneurship (or innovation) as the discovery of new combinations. The role of the entrepreneur is to break with routine rather than follow it. In contrast to the process of continuous refinement, the melding of new combinations yield a creative destruction. Even if the firm is able to overcome path-dependency, it would be forced to discard a substantial portion of its routines and any accompanying quasi-rents. Not only does this destroy the
knowledge contained within the routines, it may cause conflict. Where routines represent a
trace, breaking routines also breaks the truce (Lazarik and Denis, 2001; Lazarik et al. 2003).

Innovation of any form is difficult to monitor as output may vary considerably from input, either
creating conflict between employer and employee or placing additional risk on the innovator.
When innovation yields such a destructive influence on the firm’s routines, and thus breaks the
trace, it becomes more difficult to compensate employees *ex post*. Indeed the contracting
problem for this type of entrepreneurship may be so severe as force the entrepreneur to act as
capitalist and bear the entire risk of the venture. Casson (1982) argues that the entrepreneurs
must also be a capitalist but for the reason that the entrepreneur is betting against the market.
Casson’s entrepreneur is, by definition, in “a minority of one.” Rather than an invention, that
might be sold if intellectual property is protected, Casson envisions the entrepreneur as the
creator of an intangible idea – inside information or tacit knowledge about markets – that
cannot easily be traded.

While routinized activities which may be carried out within the firm, Schumpeterian creative
destruction must frequently take place outside the firm. Whereas Schumpeter claims that
entrepreneurship is an entirely separate function to that of capitalist, as is the case in the Berle-
Means corporation, Casson’s argues that the two functions are intrinsically connects. Path-
dependent learning suggests that Casson is partly correct when innovation challenges new
routines. While routinized innovation of this nature is difficult, it is possible that existing firms
use routines which are widely recognized to be less efficient. Firms may fail to challenge
routines because past rigidity brought efficiency but now creates inertia. A challenger need not
always be in a “minority of one” but may find like-minded individuals. This leaves room for specialization in risk-bearing and the managerial and technical aspects of running a business.

Casson argues that in this case the capitalist becomes the entrepreneur and the notional entrepreneur (the manager) has in fact become an employee. This definition is unnecessarily narrow. When the operator remains part capitalist, even if she sacrifices majority ownership, both the capitalist and the operator serve an entrepreneurial function broadly defined. Indeed the operator is likely less diversified and may thus bear a considerable risk. The operator-entrepreneur will face greater novelty than the director of the firm until problems can be solved.

By contrast, Schumpeter (1934, 1956), argues that the entrepreneur and capitalist are separate roles and although the entrepreneur may play many additional roles such as manager, laborer, or capitalist, these are distinct from the entrepreneurial function. This difference in approach may be explained in part by the inability of scholars to agree upon a clear definition of entrepreneurship. While Schumpeter’s entrepreneur solves problems, Casson’s entrepreneur identifies problems. Rather than choose between the various definitions of entrepreneur I adopt March and Cyert’s (1963) and Nelson and Winter’s (1982) hierarchy of choice, and the parallel concept of single-loop/double-loop learning. Each choice is composed of recognizing a problem or opportunity, identifying the available options, and choosing among those options.

The extent to which entrepreneurial decision-making can be carried out within the firm depends upon the whether a suitable decision-heuristic can be found. A firm cannot write a complete contingent contract telling its sales team how to behave in every circumstance but routines can embody a decision heuristic that the team should increase sales and the firm may create
appropriate incentives. The individual representative is left to decide how to react to customers. Managers of the sales department may have authority to identify new markets but it is harder to create an heuristic which consistently generates this outcome. A firm may only realize a missed opportunity when a competitor moves first.

Similarly a buying department can be told to minimize costs or an engineering department can be told to increase output. The latter choice may involve identifying possible production technologies. These heuristics may not maximize profits but they can meet Simon’s criteria of procedural rationality. By contrast it is far more difficult to create decision rules for entirely novel innovations. A software designer may have decision heuristics for improving interfaces or adding desirable features to its product but not for creating a new social media product in an unknown market space. This is not to claim that innovation never occurs in large firms but that it will only do so if the firm can create a decision heuristic.  

Over time a successful firm’s responses to novel problems will instantiate routines and simultaneously expand routinized innovation and ‘learning by doing.’ This will generally expand efficiency at the expense of flexibility but efficiency may suffer if the firm has excessive slack. Instead of redistributing gains from efficient routines to shareholders the firm can consume this slack. Managers and employees may extract perquisites or reduce their effort. The firm may engage in empire building or other unnecessary programs. Over time innovations outside the firm will alter the evolutionary fitness landscape and reduce the efficiency of the previously fit routines. While the firm has some power to change these routines excessive slack may reduce the incentive to do so. Furthermore, general problem solving will continue to produce new

\footnote{Firms may also purchase innovations either an intellectual property or by buying a small firm wholesale and thus buying the routines as well.}
routines but may not eliminate the old routines. This then leads to excessive complexity and bureaucratic rigidity.

V. Morphogenesis

Innovation adds a further dimension to the behavior of the firm. Previous sections distinguished between the selective, or phylogenic processes, and the adaptive, or morphogenic, processes. In the latter case, the firm’s adaptiveness represents a trade-off between efficiency, found in specialization, and flexibility, found in generalism. No firm can be entirely rigid or the ongoing morphogenic and phylogenic processes would quickly render its routines obsolete. Nor can any firm be completely adaptive for this too would signal evolutionary death. Ultimately, the firm’s survival depends in part upon this trade-off making adaption itself a criterion for selection.

The development of the firm is path dependent. The firm must first survive in disequilibrium but the later life-cycle firm will not be a larger isomorphic version of the initial firm. Both learning and the constraints of organizational growth will modify the firm’s routines. The firm is modified by its ecology too, including innovation. This innovation can be divided into learning-by-doing, including double-loop learning, and disruptive innovation. Firms must, at alternate points in their life-cycle, be capable of both but must trade-off one for the other. This life-cycle creates differing requirements for monitoring, finance, and ownership. This section presents a model of firm innovation which demonstrates that the joint-stock firm outperforms other modes of ownership in allowing firms to modify their control structures to best suit their innovative capacity at various points in their life-cycle.
I assume that learning-by-doing occurs continuously while disruptive innovations will only occur when the benefits of adopting the technology outweigh the advantages of tacit knowledge and the truce embodied in routines. I propose that that the likelihood of a technology shock can be modeled as a probabilistic function of time.

H1: The benefits to specificity increase over time.

Learning-by-doing reduces the cost of performing activities as the firm gains experience. This experience does not accrue evenly over time but is concentrated on those activities the firm performs most frequently. The firm does not gain experience as rapidly with irregular activities and the cost of maintaining organizational slack for these contingencies would not fall as significantly as the unit cost of performing frequent activities.

Figure 2 – Niche Theory with Learning
Figure 2 adapts the illustration in Figure 1 from niche theory to show the effect of single-loop learning on both a generalist and a specialist. Again both firms compete on a single environmental dimension contained in the set \( E = \{e_1, e_2, \ldots e_n\} \) where instances \((e_i)\) are normally distributed about an average \( e^* \). I further adapt the model to incorporate learning by assuming firms can improve their fitness at state \( e_i \) by repeated performance of that of similar tasks. Fitness at \( e_i \) is given by

\[
    f_i(n) = f_i(1)n_i^\gamma + \sum (n_j^\gamma)d(|i - j|)
\]

Where \( f_i(n) \) is fitness with respect to environment \( e_i \) on the \( n^{th} \), \( f_i(1) \) is fitness on the first iteration, \( n_i \) is the number of operations performed, and \( \gamma \) is learning constant. \( n_j^\gamma \) is learning for a closely related activity and \( d(|i - j|) \) is a decreasing function of the distance between \( i \) and \( j \).

With experience both the specialist and the generalist improve their overall fitness across their entire niche but gains are greatest around \( e^* \) where both accumulate experience more rapidly. The specialist has a narrower niche and accumulates experience more rapidly for most values of \( e_i \) within that niche and will also increase fitness within that niche by performing more closely related tasks. Only at the periphery does the generalist gain more rapidly. This implies that the specialist also increases the width of the niche over which that firm enjoys a competitive advantage.

**H2:** The benefits of flexibility decrease over time.

Benefits to flexibility are greatest in an uncertain environment. Following from H1, as benefits to specialization increase, industry will increasingly be dominated by specialized firms which
should generate less uncertainty. Firms may increase become more specialized through double-loop learning or the selection processes will eliminate firms which are too generalized. Selection will also eliminate firms which are optimized for a different environment. Over time increased market concentration will eliminate uncertainty from industry behavior.

Increasing certainty in one industry may also increase certainty in closely related industries creating a feedback effect. For example, if one practice comes to dominate the steel industry then selection processes will identify automakers which are best adapted for that process creating greater certainty in both the auto manufacture and steel manufacturing industries. As uncertainty decreases, variance in e, decreases, again favoring increased specialization. Novelty is introduced only by disruptive innovation and will decline between such periods.

H3: Firms will tend toward greater rigidity over time

An industry may become more rigid because phylogenic processes select for specialists or because morphogenic processes promote rigidity. This becomes desirable when the benefits of specialization increase and the benefits of flexibility decrease. Further, in the early portion of the firm’s life-cycle the firms will have little value embedded in tacit knowledge and will rely more on active response to problems. As routines are established over time it becomes costly to actively interfere with established processes at the risk of breaking the truce.

To alter its adaptiveness in response to changing market ecology the firm must rely upon higher routines. These higher routines will make it harder for any individual to alter existing procedures and will thus preserve knowledge and the truce. In doing so it becomes harder for the firm to adapt to changes in the environment. The new adaptiveness function represents
increased specialization combined with greater learning. This differs significantly from the biological concept of niche theory in that a species cannot choose to change its adaptiveness. A carnivore cannot become an omnivore because it so chooses. An omnivore may be better adapted to eat meat and choose only to vary its diet when facing hunger, but it must have been born with the biological capacity to do so.

This deviation is possible here because only single-loop adaptiveness is represented in figure 3. The higher routines can be less easily adapted while top-level routines cannot be adapted at all. This is reasonable for a learning organism. A person who is born with great intelligence and manual dexterity may choose to become a neurosurgeon. Once trained in that field she is less likely to choose to become a lawyer or an engineer even if those courses were initially available to her. She may also learn particular traits or habits, peculiar to her specialization, that actually reduce her aptitude to those professions without changing her genetic abilities.
Figure 3 – Environmental Shift

In figure 3 gradual changes in the environment are indicated by a shift from $e^*_1$ to $e^*_2$ (again representing average outcomes). The industry incumbent has initial fitness function $F_I$ which, over time, shifts to $F'_I$. A new entrant, with fitness $F_E$ which is optimized to the present value of $e^*_2$ is unable to compete until $e^*_2$ drifts far enough from $e^*_1$ to outweigh the incumbent’s experience advantage. The condition for entry is:

$$\sum_{i} [p_i f_i^I(n) - p_i f_i^E(1)] < 0$$

Where $p_i$ is the probability of any given $e_i$, $f_i^I(n)$ is the experience adjusted fitness of the incumbent, and $f_i^E(1)$ is the fitness of the inexperienced entrant. The possibility of entry is decreasing in the incumbent’s level of experience, and increasing in the value of $e^*_2 - e^*_1$, and the variance of $E$. Entry into an industry is likely both shift the value of $e^*_2$ (though not
necessarily away from $e_1^*$ and to increase the variance in $E$, both in the affected industry and in closely related industries. Thus while sudden disruption may result from uneven change, gradual environmental change may also produce cascades. The result may be similar to the condition of punctuated equilibrium in biology, where long periods of slow change may be disrupted by short (10,000 years or so) periods of rapid change. Similar concepts in complexity theory suggest stable environments are hard to penetrate but gradual environmental change may eventually give way to mass extinction or arrival of new species (c.f., Beinhocker, 2006; Lewin, 1992).

H4: Firms will shift from active to passive monitoring over time

As firms shift more to dependence on routines, the firm will also shift toward passive monitoring with little direct intervention by owners. In the initial stage the firm is small and managed closely by an operator-entrepreneur who is part capitalist and part manager. Although there may be other backers (or such may enter early in the firm’s growth) these are likely to take an active interest in the management of the business. During this phase there is little reliance on routines or tacit knowledge and the benefits to monitoring are significant.

As the firm grows larger, concentrated ownership and control becomes less desirable. The firm must routinize much innovation and devolve decision-making to subordinates managed through efficient decision heuristics. Active monitoring becomes more costly. If the firm wishes to raise new finance this active monitoring creates potential conflict between large and small shareholders or between owners and debt-holders. By protecting the firm’s routines from intervention this rigidity reduces conflict and improves efficiency. Owners will still wish to exercise control and will turn to passive monitoring methods such as leveraging the firm, which
reduces slack and thus improves efficiency, but does not impose direct control over management.

H5: Joint-stock firms are designed to optimize top-level plasticity.

Where lower routines can be altered by higher routines, including the adaptiveness of a set of routines, the top level routines form the organizational DNA and cannot be altered. A firm may elect to become a generalist or a specialist with respect to an industry niche but it cannot easily convert its ownership form. Alternative forms of organization may be better optimized for specific environmental conditions that occur during the firm’s life-cycle but are insufficiently adaptive over the firm’s entire life cycle. This is achieved by the existence of property rights which can be separated to limit intervention but also combined where necessary to impose greater control over the firm.

In the early stage both control and ownership is highly concentrated. The nature of the firm in this initial stage does not require joint-stock ownership but highly diffuse ownership structures typical of customer or worker ownership would be inefficient. These groups would experience higher organization costs in raising capital and would have difficulty monitoring management.

As the firm grows the capitalist-entrepreneur is replaced with management and it becomes inefficient to retain concentrated ownership. This frees the firm from a tight wealth constraint and allows risk to be spread among those with a comparative advantage in risk-bearing. Yet while ownership rights are sold widely the link between ownership and residual control is retained. Owners cannot separate their voting rights from their rights to profit even though they may remain passive.
In this stage the law, and the nature of voting rights, enforces passivity. If a shareholder is unhappy with management then she will benefit more from selling her shares to someone who is less unhappy than in attempting to institute change. Only if many shareholders are unhappy, reducing the stock value and exposing the company to takeover risk, should management be concerned. This passivity protects the knowledge contained within routines and maintains the truce. Similarly, the law makes it harder for minority shareholders to raid the firm’s assets at the expense of the majority.

Passivity protects the firm but it can also allow organizational slack to grow. While some slack is beneficial, too much will be spent in ways that do not benefit shareholders. Firm’s can reduce their slack by increasing their debt holdings and returning capital to shareholders but they will do this only if it is in management’s interests to do so. Takeover provides that incentive. While bank-owned firms in Germany and Japan were successful in routinized innovation they were far less successful in managing slack or in the creative destructive portion of innovation. Indeed these elements are linked: drawing excess capital out of large firms will facilitate greater capital availability for small firms which need finance to grow.

While the threat of takeover may curtail organization slack this does not imply that there should be no takeovers. The semi-rigid structure of the firm is well designed to prevent overarching change that destroys the value locked in routines but over time those routines become less fit and more severe change is necessary. If the value of the firm were locked in non-specific capital then this might easily be sold. But the firm’s value may be locked in routines, some of which may still be valuable, but others of which are harmful. At this point management in the joint stock company is too constrained by the firm’s design. Shareholders too would rather leave
than try to impose beneficial change. The inalienability of residual control and ownership, which when combined with the firm’s constitutional structure, imposes passivity in ordinary circumstances, ensures that ownership and control can be (relatively) easily recombined.

It is entirely possible that the firm’s routines no longer carry value in which case the firm cannot survive. It is possible too that only some portion of the firm remains valuable: the sales division may be successful but the manufacturing division is obsolete. The firm might still be taken-over by a competitor who wishes to extract the valuable portions of the company. In some cases the firm may file for Chapter 11 bankruptcy. While this allows the firm to be reformed it also involves a complex bargaining process, which takes place too late when debt-holders have already some of their assets. By contrast, takeover allows the new owner to eliminate much conflict within the firm.

While the firm may frequently resemble the stakeholder corporation though its routines, stakeholder control rights cannot be rapidly or easily recombined. Knowledge problems make voting an inefficient mechanism for governance. Employees and customers who find themselves dissatisfied with management should prefer to leave rather than attempt to renegotiate the truce embodied in routines. Shareholders too should sell if they dislike management but shareholders can sell their capital interest in the firm easily while simultaneously surrendering their voting rights. If sufficient shareholders feel similarly then the firm may be devalued and taken-over by a party who is capable of building a significant interest in the firm to overcome the free rider problem.

No such mechanism exists in the stakeholder firm. Even if the firm has no physical capital it must still develop routines. There is initial risk that forces labor or consumers to act as
capitalists. If the firm is successful then it will increase productivity though learning but expansion requires that the firm hire new employees or find new consumers. If the shareholder firm wishes to find new capital it can sell new shares at a higher price based on the firm’s success. Employment rights and consumption rights are not usually bought and sold. Where the interest of members potentially diverged (as must be the case with most employee-owned enterprises) existing member may want control over who may receive voting rights or under what circumstances a member may be dismissed.

VI. Conclusion

The modern firm is typically owner by stockholders. In the largest firms in the United States and other common law countries these owners are usually so diffuse as to affect no control at all. Rather than attribute this feature to a dereliction by the law it can be better explained by a morphogenetic view of the firm. Routines within the firm broker an ongoing truce while controlling information costs within the firm. It is unsurprising then that the routine-based firm should at times seem to be so unresponsive.

Yet firms must sometimes be controlled from the top. This hierarchical form of control is so anathematic to the routine-based firm that it requires, and to preserve routines must require, a significant outside force before the firm can be recast. The joint-stock corporation is flexible enough to allow this significant recasting but flexible enough as to protect the firm’s routines in most circumstance. The organizational form of the firm preserves and grows knowledge. As

---

134 It used to be known for public offices, such as a military commission, to be traded but I know of no modern example of this practice in a commercial enterprise. Before the recent wave of demutalizations, stock exchanges would sell seats.
such understanding the morphogenic processes which guide the behavior of the firm is a necessary component of understanding growth.

A morphogenic theory of the firm may also be important in understanding phylogenic processes for transmitting knowledge over time. The ecologically rational routine-based firm predicts the existence of cascades in firm failure. If each firm embodies knowledge of its surrounding environment then it may become less efficient over time as its environment changes. When the environment changes sufficiently to cause one firm to fail this modifies the environment over all firms and may cause additional failures leading to a cascade. Further research may indicate whether bubbles and recessions can be explained by such cascades.
References
References


Bainbridge, S.M. (2012), Corporate Governance after the Financial Crisis, New York, NY: Oxford University Press


175


4. Crisis and Federalism

In 2007-09, a financial crisis shook markets. The crisis precipitated not only a severe recession, the effects of which are still felt, but a legislative and regulatory response in the Dodd-Frank Wall Street Reform and Consumer Protection Act. The new law strengthened federal regulators while weakening state regulators in key areas—namely corporate governance law which shall be the primary focus of this essay. The law’s authors argue the act was a necessary response market failure, while academic critics such as Bainbridge (2010, 2012) and Pearce and Broughel (2013), claim the regulator’s new powers fail to protect investors. I present a more serious constraint: key provisions of the legislation shift the process of legal innovation from discovery by courts under market discipline to a design process controlled by experts under political constraints. I argue that this will generate less efficient law, increase opportunities for rent-seeking, and may increase externalities. Moreover, it undermines the ability of the law to act as a commitment mechanism.

Market discipline in law comes from competition between jurisdictions. Proponents of state competition argue that competition generates a race to the top, with states competing to provide the most efficient laws (Romano, 1985, 1987; Bainbridge, 2010, 2012; c.f. Hayek, 1939; Tiebout, 1956). Critics argue that it creates a race to the bottom: boundedly-rational investors

\(^{135}\) Public Law 111-203
\(^{136}\) Under the pre-emption doctrine a federal law that conflicts with a state law will pre-empt (i.e. invalidate) the state law. Although a matter of constitutional law, the doctrine is not found directly in the U.S. Constitution but has been inferred by the Court from the Supremacy Clause (Article VI, clause 2). The exact inference and meaning of the doctrine has varied over time and between Courts. See Greve (2012) for a general summary.
cannot properly monitor managers without the protection of the law. When that law is chosen by corporations, management will choose the venue that suits their own best interests over those of investors. They argue states will ignore externalities including the macroeconomic effects of corporate misbehavior. Empirical evidence is scant, frequently conflicting, and fails to resolve the debate (c.f. Bhagat and Romano, 2002; Romano, 2004). I argue this failure results from both critics and proponents misconceiving the nature of state competition. Rather than choose between static rules, investors must commit to a changing set of rules, under the dual pressures of rent-seeking and innovation.

When investors choose between states, they seek secure property rights, but also flexibility to respond to changing circumstances. The law seeks to avert crises but must respond effectively when crisis strikes, by adjusting the law to take new information into account with creating uncertainty. To permit beneficial change, but restrain rent-seeking, states must rely on a mixture of semi-flexible institutional constraints including constitutions and independent judiciaries. Interstate competition may also serve as an external constraint but if states are to make a credible commitment to such a restraint they must empower a federal government which must somehow itself be restrained (Weingast, 1995).

In the modern European Union and the pre-constitutional united States, federalism is the alternative to multiple independent states. In those cases, federalists fear(ed) a combination of parochial state-interest and inadequate central control would hinder internal free trade and leave states free to export externalities. In Europe, a weak form of federalism has allowed the governments of Greece and other southern European countries to impose macroeconomic externalities on the rest of Europe. Yet federalism is also the alternative to a strong central
government which reduces states or provinces to mere “corporations,” taking their power at the
discretion of the national government in the same manner as cities do from their state
governments (Greve, 2012). The latter alternative weakens the competitive strand of
federalism just as the absence of a federal government.

State choice is thus embedded in federalism but the form of federalism is important (Rodden,
2006; Greve, 2012). A strong central government can overwhelm states but a weak government
can fail or worse, empower states to export costs (Ibid.): a problem most recently illustrated by
the European fiscal crisis. The institutions of federalism must protect beneficial competition and
strengthen the state’s commitment to the protection of property rights; restraining the states
without creating an unrestrained federal government. In the United States this balance has
historically been preserved in the realm of commerce by dividing federal jurisdiction into two
functions. First, the dormant commerce clause provides enforcement of state competition with
judge-made law governing relations between the states (Greve, 2012). Second, the affirmative
commerce clause permits political lawmakers to directly govern commercial relations through
regulatory action (Ibid.).

Dodd-Frank, and particularly the corporate governance provisions of Title IX, seeks to shift the
traditional balance toward the latter form of law. While regulation in response to a financial
crisis is far from unprecedented – crisis has served as the spur for financial regulation in England
and the United States since the 16th century (Banner, 1998) – the Act differs from previous

---

137 It was Hamilton who referenced “corporations” at the constitutional convention, where he argued
strongly that states should be relegated to the same importance as a city government, though Greve
(2012) reports that Madison supported the same perspective. Equally pertinent to this investigation,
Hamilton’s decision to use the corporation as his example draws on the traditional practice of
corporations being created by individual acts of the legislature which could be amended equally easily
(Greve, 2012).
legislation in its scope. Although enacted in response to a crisis in banking and financial markets, the law expands federal oversight into areas far outside of banking regulation, including into corporate governance. Its predecessor in this respect was Sarbanes-Oxley, passed in response to a series of corporate frauds, and can be criticized on many of the same grounds as the more recent Dodd-Frank legislation (Bainbridge, 2010, 2012; Romano, 2004).

I. The Financial Crisis of 2007-09

The most recent crisis began with a housing boom fuelled by low interest rates, a consumption boom, and new financial innovations (Duca et al., 2010). The latter included mortgage products to consumers (Sanders, 2008), and securitized debt products. Schwarcz (1994) describes how securitization, then a relatively new financial innovation, allowed banks to reduce the cost of raising funds by spreading risk. Loans to low income/high risk individuals (subprime lending) could be bundled together and sold as investment grade securities. These securities broadened availability of credit but also reduced transparency (Reinhart and Rogoff, 2008). Moreover, existing regulations required that funds representing ordinary investors could only purchase securities that were rated as “investment grade” by professional credit rating agencies. This, in turn, enforced a practice of agencies being paid directly by security issuers and may have contributed to their inaccurately rating risky securities as investment grade.

Federal Reserve policy and financial innovation combined with government mandates, leading to increased subprime mortgage lending. Furthermore, key sub-prime lenders, the Government Sponsored Enterprises (GSEs) Fannie Mae and Freddie Mac, were perceived by investors as having implicit government-backing (Duca et al., 2010). The GSEs were consequently able to borrow at low premiums that failed to reflect the default risk than the underlying mortgages,
further fueling the credit boom. Demyanyk and Hemert (2011) show that the quantity of loans made by financial institutions grew dramatically while the quality of loans declined and risk premiums fell. In late 2007, uncertainty about the value of collateralized securities led to the freezing of inter-bank lending (Bordo, 2008). At the same time, weaknesses in subprime lending, which were initially disguised by rising house prices, were exposed when the housing market slowed.

While the boom and subsequent crash were manifested in the housing market through novel financial innovations, both Duca et al. (2010) and Reinhart and Rogoff (2008) find that the boom itself followed a familiar pattern of innovations that are not fully understood by investors or regulators, combined with easily available credit, leads to overpricing of assets which eventually corrects through financial markets crash. Taylor (2009) argues that the cheap credit, not financial innovation, was the main cause of the boom and crash. Indeed, securitization preceded the dot-com boom of the 1990s (Schwartz, 1994) and Sanders (2008) finds that many of the consumer-oriented mortgage products had been available long before the boom but only found popularity when the cost of credit fell.

While cheap credit (either from innovation or easy money) may have sparked the initial boom but that over time, the later stages may be caused in part by market over-reaction observed in the behavioral finance literature (De Bondt and Thaler, 1985) and predicted by heterogeneous agent-based models (Arthur et al., 1997). Piazzessi and Schneider (2009) find that momentum traders increasingly became the driving force behind increased demand for housing. Using a survey method found that early in the boom up to 85.2% of households believed it was a good time to buy a house and up 72% of those cited favorable credit as the reason. Later in the
boom, fewer households believed it was a good time to buy a home. However, the number of momentum traders doubled from 10% to 20%. Consequently, toward the later stages of the boom, the use of simple heuristic rules may have driven house prices away from fundamental values.

While the role of financial innovation, and the regulatory response to new instruments, played a questionable role in the boom, the regulatory response clearly influenced the severity of the subsequent crash. To protect investors from bank failures, the United States and other countries require that banks maintain a minimum ratio of equity to assets (capital asset ratio). Through a series of modifications from the 1990s onwards, including the approval of the international Basel II accord, these rules sought to create a market mechanism to discipline banks. The rules required that banks use the market price of assets for this accounting and did not set a fixed capital asset ratio but varied such that banks holding riskier assets must also hold greater equity.

These rules were designed to provide certainty, and did so during economic stability. If a single bank made a poor investment the reduction in asset value would be recorded against the bank’s equity forcing the bank to either sell assets or return to markets to raise new equity. However, when crisis struck the rules were too rigid and exacerbated instability during the crisis Kashyap et al. (2008). When general asset prices fall, all banks are forced to recapitalize regardless of whether they were well run. With markets quickly becoming illiquid it becomes difficult to accurately price assets. When banks face losses they will choose to dispose of assets quickly rather than try to recapitalize, resulting in fire sales which further reduce the value of assets on

---

138 Also known as short memory traders, momentum traders are investors who expected the price of housing to rise further based on recent experience.
other banks’ balance sheets. Thus, banks become unable to make “good” loans, exacerbating the downturn.

The Basel II requirements rely on statistical methodologies known as “value at risk” (VaR) which failed to account for endogenous risk (Daníelsson et al., 2001). By calculating risk from times of stability, these methods are inaccurate predictors of risk in a crisis. Furthermore the requirements are pro-cyclical meaning that banks are forced to increase their capital asset ratios during a downturn, exacerbating the problems caused by declining balance sheets. Greenlaw et al. (2008) find the average VaR numbers for several major investment banks rose 34% between August 2007 and February 2008.

The rules at the time of the crisis were designed to respond to problems that existed under previous, stable conditions. When the circumstances changed, inflexible rules exacerbated the problem and locked-in bad practices. Not only were banks forced to impose market externalities but market participants were unable to alter bad practices. While this practice of relying on ratings agencies paid by the security issuer may have needed no regulatory requirement to survive the bull market, the financial crisis provided new information that would ordinarily spur innovation. Instead of responding immediately to this information, investors and securities issuers were forced to wait for regulatory action to address the problem.¹³⁹

The boom and subsequent crash illustrate the importance of financial innovation. New ideas can generate real improvements but rather than instant adjustment suggested by the rational expectations literature (Muth, 1965; Lucas, 1972) it takes time for markets to discover the correct price of new instruments (c.f. Hayek, 1937, 1945). Historical surveys find a consistent

pattern whereby financial innovations are initially wrongly valued, leading to a later market correction (Duma et al., 2010; Reinhart and Rogoff, 2008). Markets, and regulators, must discover the correct set of values, and rules, through a process of discovery, in which crises play an important role in generating information. Yet while markets correct, sometimes through a severe crisis, rules can fail to adapt to changing conditions, or when they do, can undermine property rights. I argue in the next section that specific parts of the Dodd-Frank act will exacerbate this problem.

II. The Dodd-Frank Regulatory Response

The causes of the crash were complex: market error was compounded by inflexible regulation that had yet to be tested against a serious crisis. Yet, while the financial crisis has been held up as an example of under-regulation (Cheffins, 2009), there is scope too for a greater regulatory flexibility to prevent or mitigate future crises. The regulatory response, initiated by the Dodd-Frank Act of 2010, only expanded and centralized regulation. While both regulators and markets committed serious errors the Act shifted power from markets to regulators (Peirce and Broughel, 2010). No regulatory powers were eliminated and only one regulator, the Office of Thrift Supervision, was abolished and its powers redistributed to other regulators (Ibid.).

Bainbridge (2012) argues that the law was not a response to the problems that caused the crisis but rather an opportunistic attempt to expand the power of the federal regulators at the expense of states. Bainbridge focuses on the decisions to expand federal jurisdiction over corporate governance under Title IX of the Dodd-Frank, even though corporate governance was not a proximate cause of the crisis (Cheffin, 2009). The law followed a precedent first set by

140 Public Law 111-203
Sarbanes-Oxley, shifting away from disclosure rules which had typified modern federal securities regulation, and into corporate governance which had previously been left to the states (Romano, 2004; Bainbridge, 2012).

This paper is interested primarily in the corporate governance aspects covered in Title IX but other aspects of the Act will also influence relations between investors and corporations, and particularly the manner in which corporations choose to finance their activities. An examination of these other titles also offers illumination into the problems of mitigating externalities under a centralized regulator. In particular, the creation of “systemically important” designation into the regulatory scheme, effectively institutionalizing the concept of “too big to fail” (Pierce and Broughel, 2012). As later sections will describe in more detail, a designation such as this may undermine the ability of the law to commit to defined property and contract rights in the future (c.f. Rodden, 2006). The remainder of this section will therefore summarize Dodd-Frank in its entirety, drawing heavily on the work of Peirce and Broughel (2012), before returning to the look at the corporate governance aspects in greater detail.

Title I of the Act established two new regulatory agencies. First, the Office of Financial Research will have power to collect and analyze financial data and disseminate its findings. Second, the Financial Stability Oversight Council (FSOC) will have power over several regulators to manage systemic risks. Significantly, this new regulator will have responsibility over non-bank financial firms, amounting to a broad and somewhat amorphous new area of responsibility in financial regulation. FSOC will have the power to designate firms as “systemically important” which, in turn, will empower the Federal Reserve to impose additional regulation on those firms including enhanced capital requirements and leverage limits.
Peirce and Broughel note that any designation of systemic importance provides an implicit guarantee of risk from the federal government in the event of the crisis. These additional regulatory measures are intended to limit the ability of firms that are already “too big to fail” from externalizing the cost of excessive risk taking onto the federal government. However, authors argue that dynamic uncertainty prevents regulators from accurately identifying systemic risk until after the fact, implying that the additional powers will do little to help regulators avoid a financial crisis but will increase the incentive of firms under the designation to take additional risk that raises the likelihood of a crisis.

Title II of the Act creates a power to be known as “Orderly Liquidation Authority” within the Federal Deposit Insurance Corporation (FDIC). This authority allows FDIC to override standard bankruptcy procedures in favor of an alternative resolution process. Pierce and Broughel note that this process is less transparent than standard rules for bankruptcy and eliminates a standard rule of law relying on judicial oversight, with a broad discretionary authority in the hands of FDIC.

Title III of the Act is the only provision to abolish a regulatory authority: the Office of Thrift Supervision. However, the Title reassigns those powers to the Comptroller of the Currency and the Federal Reserve rather than permitting deregulation. Furthermore, the Title creates a new Office of Minority and Women Inclusion at financial regulatory agencies. The provision’s sponsor provides no evidence that the lack of female or minority inclusion was responsible for the financial crisis or that the Office will reduce a systemic risk. Rather, the decision to include this provision is consistent with Bainbridge’s (2012) argument with respect to the

corporate governance provisions of the Act, that crisis provided an opportune moment to enact
previously advocated measures. The Title also expands federal deposit insurance to cover
amounts up to $250,000 and expands the total base covered by deposit insurance. As with
previous provisions this increases the ability of financial firms to externalize the cost of risk and
may increase systemic instability (Pierce and Broughel, 2012).

Title IV of the Act expands the Securities and Exchange Commission’s (SEC) authority over hedge
funds by requiring advisors to hedge funds and other private funds to register with the SEC. It
also prevents individuals from counting the value of their home toward the minimum net-
wealth threshold required to invest in a hedge fund. This provision makes it harder for
individuals to meet the requirement to take personal responsibility for risks to their investment
under a less onerous regulatory regime. The Title also raises the threshold for an advisor to
become registered with the SEC while leaving intact the alternative state regulatory regime.
This may, paradoxically, have the effect of transferring responsibility for registering advisors
from the federal government to states (Peirce and Broughel, 2012).

As with other sections of Dodd-Frank, the hedge fund registration rule was not based on any
evidence of hedge fund failure during the crisis. Shadab (2009) notes that hedge funds
outperformed more heavily regulated funds, in part because of their ability to invest in products
regulators had identified as being riskier. Rather, the SEC proposed the rule in 2004 but it was
struck down by the court for exceeding the Commission’s statutory authority at the time (Pierce
and Broughel, 2012).142

---

142 The rule was struck down in Goldstein v. SEC, 451 F.3d 873, 877 (D.C. Cir. 2006).
Title V of the Act creates a new agency to be known as the Federal Insurance Office (FIO) within the U.S. Treasury. The new Office will have responsibilities to monitor the insurance industry including the power to monitor systemically important insurers. FIO can recommend an insurance company be designated “systemically important” by FSOC. Peirce and Broughel argue that the institutional structure governing the two agencies increases the likelihood that this recommendation will lead to a designation by FSOC with the same incumbent problems of increasing systemic risk described above.

Title VI of Dodd-Frank, known as the “Bank and Savings Association Holding Company and Depository Institution Regulatory Improvements Act,” gives the Federal Reserve regulatory power over a range of new financial entities. Peirce and Broughel note that Act assumes that regulators have sufficient knowledge to manage systemic risk before problems become apparent, even though errors by the Federal Reserve contributed to the financial crisis. The authors further assert that this award of power increases institutional incentives to alter rules of property and contract in the event of a new crisis.

Title VI also implements the Volker Rule, named for former Federal Reserve Chair, Paul Volker. The rule prohibits banks from engaging in proprietary trading, with the intention of preventing institutions from receiving an implicit subsidy from the government insurance of deposits for high risk activities which benefit shareholders. However, Pierce and Broughel note that the law assumes regulators can easily distinguish between proprietary trading which increases risk and hedging and market-making activities which increase diversification and reduce systemic risk. The rule potentially leads to rigidities which can make it harder to adjust to local knowledge (c.f.
Hayek, 1937, 1945). It also transfers discretionary power into the hands of regulators which may reduce rigidity somewhat but at the cost of less clear property rights.

Title VII of the Act awards SEC and the Commodities and Futures Trading Commission (CFTC) the power to regulate over-the-counter derivatives. These instruments are used by. While these instruments are important to farmers, utility companies, and other firms, as a means of managing risk, they are not as widely traded as equities or futures. By mandating the use of clearing houses, designed for more liquid markets, the Act raises the cost of engaging on risk management activity. Moreover, the CFTC recognizes that the clearing houses themselves may be a new source of systemic risk.  

Title VIII requires the Financial Stability Oversight Council to designate certain activities as systemically important, and empowers the Federal Reserve, SEC, and CFTC, to impose additional regulatory requirements on those activities. This power extends to any “company engaged in activities that are financial in nature or incidental to a financial activity,” thus expanding the authority of financial markets regulators well beyond the scope of financial corporations. Moreover, the Title expands the concept of “too big to fail” beyond the financial realm and creates a moral hazard through the provision of ex-post insurance to clearing houses (Pierce and Broughel, 2012)


144 Ibid, at p.95, fn. 9, note: “Dodd-Frank Wall Street Reform and Consumer Protection Act, § 803( 5 ). The scope of activities that are financial in nature or incidental thereto is as broad as the Fed defines it under the Bank Holding Company Act, U.S. Code 12 (1956), § 1843( k ). There is no de minimis threshold below which a company is beyond the reach of the FSOC.”
Title IX includes the corporate governance provisions of Dodd-Frank which are summarized in greater detail at the end of this section. In addition, the Title creates several new offices within SEC, expands the Commission’s enforcement powers, and amends regulations governing credit rating agencies. However, while the rigidities imposed by the regulators were a factor in the crisis, the Title only addresses the specific problems that may be unique to the crisis and increases structural rigidities by expanding SEC authority over ratings agencies. Furthermore, both Peirce and Broughel (2012) and Bainbridge (2010, 2012) note that many of the provisions of this Title have no bearing on the crisis but are proposals which had failed to win legislative approval prior to the crisis.

Title X of the Act creates the Consumer Financial Protection Bureau (CFPB). Zywicki (2012a) notes that as with many other provisions of Dodd-Frank, the provision was proposed prior to the financial crisis, by then-Professor (currently Senator) Elizabeth Warren, who was to be the agency’s first head, and was adopted when the financial crisis provided the opportunity. The new agency also has unusual independence from the legislative and executive branches of government (Ibid.). CFPB draws its income from the Federal Reserve but is permitted to requisition whatever budget the agency head requests. Moreover, the agency head cannot be removed without cause by either Congress or the President. Zywicki argues this peculiar level of insulation, while protecting the Bureau from political forces, also insulates it from accountability and opens the door to rent-seeking and regulatory capture (c.f. Tullock, 1967; Stigler, 1971; Peltzman, 1976). The Title also imposes restrictions on issuers of debit cards to charge fees to vendors. Again, the provision is unrelated to the systemic risk factors associated with the
financial crisis, but is a direct price control favored by retailers who stand to gain from lower fees (Zywicki, 2011).

Title XI seeks to increase oversight into the Federal Reserve’s financial crisis assistance programs. The Title amends the emergency lending authority to mandate broad-based eligibility, prohibits support for insolvent firms, and requires Treasury approval of Federal Reserve crisis lending. It also requires Congressional approval of emergency guarantee programs and requires delayed publication of the certain Federal Reserve activities to increase accountability. The Title directs the Government Audit Office to audit the Federal Reserve’s financial crisis assistance programs but does not require ongoing audit of the Federal Reserve. Peirce and Broughel (2012) argue that while these increases in accountability are positive they do not adequately overcome the problems created by other sections of the Act in giving substantial regulatory authority to an organization that is historically less accountable than other regulatory agencies.

Title XII authorizes the Treasury to develop programs, including subsidies, to increase loans to low-income individuals. Peirce and Broughel note that programs to subsidize credit are likely to lead to its underpricing, which was a substantial factor in the financial crisis. The proposed subsidy for credit is likely to increase loans that contribute to systemic risk, and increase government exposure to the cost of making said loans. The Title also introduces restrictions on Payday Lending, the practice of private lenders advancing wages though short, high interest loans. Zywicki (2009), in a review of a previous incarnation of the proposed restrictions on
Payday Lending,\textsuperscript{145} that such loans typically filled a demand at a lower cost than alternatives rather than existing as a consequence of bounded rationality.

Title XIII, known as the “Pay It Back Act,” contains a number of provisions that largely serve to reduce the budget deficit but do not address the causes of the financial crisis. Peirce and Broughel note that the Title is primarily aimed at reducing the temporary emergency programs enacted during the financial crisis, although it does not eliminate those programs. The Title also lacks measures to substantially reduce the federal government’s role in mortgage markets or exposure to risk in the same.

Title XIV also addresses the mortgage market, by establishing standards for lending and reducing flexibility in the types of mortgage product lenders can offer. Pierce and Broughel note that these standards do not alter the substantial government involvement in the mortgage market which played a significant role in underpricing credit. More importantly, however, regulatory requirements prior to the crisis created substantial market rigidities that contributed to the financial crisis and the subsequent failure to adjust to changing circumstances. Regulators risk creating similar unforeseen problems by imposing new rigidities that limit the possible mortgage product, mandating the long-term 30-year loan, and effectively proscribing market innovation.

Title XV, the last Title of Dodd-Frank, is aimed at eliminating the availability of conflict minerals in the United States. The Title creates disclosure requirements for the use of conflict minerals and additionally requires companies involved in the commercial development of natural resources to disclose any payments to foreign governments. These disclosure requirements are

\textsuperscript{145} After the proposed legislation, H.R. 1214, known as the Payday Loan Reform Act of 2009, was not released from committee, key provisions were incorporated into the Dodd-Frank Act.
to be enforced by SEC. Pierce and Broughel note that the Title does not increase protection for
investors but raises monitoring costs to the same while potentially disadvantaging U.S.
companies engaged in resource extraction.

Much of Dodd-Frank therefore appears to be tenuously linked to the problems associated with
the financial crisis, and in many cases motivated by rent-seeking by outside groups or
entrenched bureaucracies. Higgs (1987) observes that crisis is frequently used in U.S. history as
a means of advancing unrelated political objectives. Banner (1998) finds a similar trend in
securities law in both the U.S. and Great Britain. Romano (2004) finds evidence that Sarbanes-
Oxley, the most recent significant amendment to the federal law of corporations followed a
similar pattern. Bainbridge (2010) finds that this is also the case in a more detailed analysis of
the individual corporate governance provisions in Title IX of Dodd-Frank.

Bainbridge identifies six key corporate governance provisions in Title IX. First, section 953 of the
Act requires each listed firm to contain a statement explaining the relationship between
executive compensation and the firm’s financial performance. Moreover, firms must disclose
total compensation to the Chief Executive Officer (CEO), compensation to the median employee,
and the ratio between the two. While the former requirement may improve the ability of
investors to monitor CEO’s, the latter offers no clear benefit to investors.

Under section 972 the Act also requires the SEC to adopt a rule requiring firms to divulge
whether the CEO and Chairman are the same individual and the firm’s reason form making that
decision. Bainbridge (2012) notes that the provision is intended to increase incentives for firms
to separate these positions and also that there is no clear empirical evidence to suggest such a
move would be universally beneficial.
The new disclosure rule in section 972 came partially at the behest of the California Public Employee's Retirement System (CalPERS). While CalPERS notionally represents the interests of investors, the institutional design of the System has typically caused it to favor the interests of unionized workers who comprise its membership, and also of union interests in general at the expense of investors (Romano, 2004). CalPERS also played a role in lobbying for section 971, establishing SEC's authority to promulgate rules for proxy access. Again, these provisions were not aimed at addressing any specific cause of the financial crisis but SEC was already preparing to enact such rules prior to the financial crisis. The section therefore clearly asserted SEC's authority to do so and provided protection against a potential legal challenge.

Section 952 establishes a number of provisions governing compensation committees for listed firms. In particular, Bainbridge notes a requirement that each member of a compensation committee be independent. Bainbridge argues that mandate, as with the requirements for separating the Chair and CEO roles, imposes excessive rigidity on the firm's governance and prevents firms and investors from taking advantage of local knowledge of particular situations.

Section 954 establishes rules that require repayment of incentive-based compensation if the firm issues a restatement of its financial report. Again, Bainbridge argues the provision attempts to impose an excessively rigid regime that may require repayment of incentives by individuals with no control over the firm's financial reporting while not covering individuals such as proprietary traders who may be able to significantly adversely affect the firm's financial position. As with the previous provisions, the legislature suffers from unable to effectively write universal ex ante rules which provide for effective governance for all firms.
Finally, section 951 requires that shareholders be periodically permitted an advisory vote on executive compensation packages. Again Bainbridge argues that the provision was included at the behest of outside interests including the AFL-CIO and the Consumer Federation of America, whose goals may substantially differ from those of investors whose interests the SEC and securities law is intended to serve. Furthermore, Bainbridge surveys a substantial empirical literature and finds no clear evidence to support the claim that executive compensation is characterized by a severe principal-agent problem.

III. The Nature and Challenge of Corporate Governance

The goal of corporate governance is to better align the behavior of managers with the interests of residual claimants (shareholders). Corporate law seeks to improve the ability of corporate charters to achieve this goal while also supporting the interests of other claimants against the firm. This includes the holders of issued debt but also other claimants such as suppliers and individuals with outstanding tort claims. Debt-holders must be paid a fixed dividend regardless of the firm’s performance while shareholders possess residual claims over the firm’s wealth. Creditor claims must be settled ahead of shareholder claims. However, while shareholders are able to vote on aspects of the firm’s decision-making, creditors have no such power unless the firm defaults. This priority of payment and the rule of limited liability means that shareholders receive the full upside of any risk but are able to defer the downside to debt-holder. This leaves shareholders a substantial incentive to take excessive risk (for a more in depth discussion see Corporations and Crisis in this dissertation).

Despite the power to vote on corporate governance, shareholders typically behave as debt-holders, selling their stake if they are dissatisfied with management (Ibid; c.f. Bainbridge, 2102).
Joint-stock corporations securitize ownership rights, usually leaving no party with a controlling stake. As with debt, this allows for specialization in risk-bearing. Efficient markets hypothesis suggests that markets will perfectly price this risk, even though no individual has a complete incentive to monitor the risks taken by management (Samuelson, 1965; Fama, 1965a,b; 1970). A growing experimental and empirical literature, however, suggests that markets can err in assessing probability (Tversky and Kahneman, 1983; DeBondt and Thaler, 1985). More significantly, the underlying claim that probabilities can be assigned to one-off events is by no means uncontroversial among probability theorists (Gigerenzer, 1991). In an alternative formulation, Hayek (1937, 1945) argues that the role of markets rather is not to perfectly predict values but to discover prices through a process of trial and error.

Hayek’s theory deviates from neoclassical price theory by assuming that investors are bounded rationality of investors. Simon (1972) identifies three categories of cognitive limitations. First, information is costly to acquire. Second, investors have limited processing power to understand and interpret information. Third, even absent the first two constraints there is radical uncertainty over future outcomes. In the case of organizations (corporations or governments), Williamson (1975) argues that language adds a further form of bounded rationality: even if the agents who comprise an organization possess all the requisite knowledge to make an informed decision the information must be accurately conveyed to the decision-maker. Behavioralists add a further category: the use of heuristics which lead to biases such as myopia (see e.g., Thaler et al., 1997). The last form of bounded rationality is controversial with scholars such as Gigerenzer and Goldstein (1999, 2002, 1996) argue that these heuristics are a procedurally rational response to the former limitations (c.f. Simon, 1986; Cosmides and Toobey, 1994).
Hayek (1952) also (somewhat presciently) predicted that individuals would use heuristics to overcome bounded rationality. His discovery theory of prices is consistent with the observed deviations by markets from random walk predicted by neoclassical price theory (Mandelbrot, 1963). More recently, computer-based modeling has successfully generated many of the aspects of modern markets by assuming that investors rely on simple heuristics rather than a fundamental approach (Arthur et al., 1997; Arthur, 2006). These simulations artificially generate a “true” price, allowing the experimenter to identify inaccurate pricing but ignoring the allocative function of markets (that is, that the true price is not a known variable).

The mispricing of assets, when it occurs, affects all securitized holdings. Stockholdings, which include an inalienable control right over the company, are affected by additional constraints stemming from bounded rationality. If shareholders maximize the benefits of specialization in risk-bearing then none will have an encompassing interest in the firm’s profitability and will only invest in acquiring (and processing) information up to the extent of their own stake. More importantly, specialists are often hired for knowledge that the owners do not possess (Fama and Jensen, 1983a,b). When the specialists’ output is then combined unpredictable, or unmeasurable, market factors, it may become impossible to fully monitor behavior.

The owners cannot determine what portion of firm performance should be attributed to managerial behavior and what portion to external market conditions. When managers’ compensation reflects only a portion of the wealth they create they may prefer to expend company resources on perquisites such as corporate jets, or empire building, against shareholders’ best interests. Moreover, because it is inefficient to concentrate risk in managers’ hands, compensation plans rely on stock options to generate incentives while protecting
managers against downside risk. This leaves managers with an incentive to take on excessive risk to inflate the firm’s performance.

The problem of monitoring is complicated further because shareholders themselves have heterogeneous interests. The rules governing corporations must maximize the value of shareholdings, but also resolve conflicts between shareholders in the interests of economic efficiency. Managers and employees often hold stock, as do business customers. A stock-holding manager who votes for an anti-takeover amendment to the corporate charter gains job security but only bears the reduction in the firm’s stock value to the extent of his own holding. A stock-holding union which opposes efficiency-increasing lay-offs achieves the same. Outside shareholders may also have ideological preferences at odds with profit maximization, such as a desire that the firm produce its output domestically at a higher cost or that it pay foreign workers above the market wage.

Individuals with large holdings have a greater encompassing interest in maximizing the value of the company (c.f. Olson, 1982, 2000). When the legal protection for shareholders is weaker but still present, as is the case in much of Western Europe, concentrated ownership often provides an alternative form of protection for shareholders (Schleifer and Vishny, 1997). Nevertheless, a large shareholder must still be restrained from simply draining the company’s assets (Ibid.). Much of the early development of Anglo-American securities law sought to cure this problem and it remains unresolved in much of the developing world (Ibid.). If the majority shareholder is unable to commit to non-predation in future periods, then it will become difficult or impossible to raise capital in the first place. A single shareholder need not be concerned with theft against himself, and this is often the preferred response to legal regimes where shareholder protection
is close to non-existent (Ibid.). Yet a single shareholder cannot take advantage of specialization in risk bearing or the benefits of spreading risk over a portfolio.

Debt, when properly enforced can still provide access to capital but debt is hard where stock is flexible. When future returns are uncertain debt lacks the flexibility provided by stock. Yet while debt is unforgiving it also gives lenders no power to intervene until the firm defaults. The owners cannot be restrained and have an incentive to take excessive risk. Thus when risk is significant the inability to protect small shareholders can become a fundamental impediment not only to the ability of the individual firm to raise capital but to the existence of a functional capital market.

Firms are constrained in their ability to raise capital by two factors: conflict between stakeholders and bounded rationality (Williamson, 1975). Before a firm can issue stock it wants to commit to shareholders that it will repay a portion of the profits. Once the firm has raised capital the management, or primary owner if there is one, has every incentive to renege on the earlier deal and expropriate the profit. Much of the firm’s value is tied up in quasi-rents: in other words the firm rarely has the same value in liquidation as in operation (cf. Klein et al., 1978). The same uncertainty that forces the firm to turn to the soft constraint of stock over the hard constraint of debt also makes it impossible to write a complete contract. Instead the firm must specify the decision rule that will be used to resolve conflict.

Eisenberg (1976) argues that the firm’s charter forms a constitutional agreement. As such the same process of pre-constitutional bargaining described by Buchanan and Tullock (1962) also describes the forces which form the firm’s articles of incorporation. The authors’ conditions of explicit bargaining in a vacuum, and the Wicksellian requirement of universal consent, may even
better describe the firm than the nation-state. Investors do not participate in the bargaining stage until the end where they must either accept or reject the bargain. If they reject the bargain then they do not involve themselves further with the firm, but if they accept then they must participate, however indirectly, in the continued contract.

This consent is given by accepting stock at a market price, yet as the behavioral evidence reviewed earlier demonstrates, investors may misprice risk. Investors may fail to correctly predict the effect of rules in an uncertain future, but investors may still use history as a factor in determining which rules are likely to be efficient. When investors use heuristic approaches they will select rules which have been value-increasing in the past. Hayek (1973) argues that for legal rules, this evolutionary selection mechanism alone can discover efficient legal rules. However, evolutionary selection remains constrained by a path dependency and a tendency to settle into local optima.\footnote{Path-dependency, initially developed by Arthur (1989), describes the limitations imposed by past conditions which are no longer relevant. For example (Liebowitz and Margolis, 1995) found that the QWERTY keyboard was initially developed to slow down typists who jammed typewriters. At the invention of the electronic keyboard there was already significant human capital invested learning in the keyboard design causing QWERTY to be favored over potentially more efficient competitors.}

Unlike debt-holders, shareholders may vote on some issues, but voting too is constrained by bounded rationality and by conflict between shareholder interests. The firm may set a low requirement to effect change, such as giving all shareholders access to the proxy machinery. This can strengthen the power of small shareholders against management but may reduce the barriers to management malfeasance. Access to the proxy machinery also gives all shareholders the ability to impose significant costs on the business. Much of the shareholder activism of the
1990s was driven by outside interests such as the California public pension fund (CALPERS) under supervision of public unions (Romano, Bainbridge).

At the other extreme, requiring a supermajority allows small shareholders to effectively restrain special interests or large shareholders (if the supermajority requirement sufficiently exceeds the large shareholder’s stake) but can impede shareholder control over management. Under a supermajority requirement holdouts can block beneficial changes which occur frequently (though not always) through the takeover mechanism. In order for a value-increasing takeover to proceed the buyer must be able to capture enough of the benefits of reforming the firm to justify the risk and expense. If a minority can hold out they can potentially capture a far greater portion of the available rents at the expense of the buyer and other shareholders. While access to the proxy machinery creates a commons, supermajority requirements which hand the exclusionary rights to block change to minority shareholders create the opposite tragedy of the anti-commons and leave power in the hands of management.

While shareholders are willing to pay a high premium to vote, they rarely exercise that power. Investors will usually vote with management, and against activist shareholders, or sell their shares. Paradoxically, shareholders will approve of a measure qua voter-owner but disapprove the measure in markets through lower prices. Though the response through markets price is not above reproach, there is reason to take it as the shareholder’s true preference. Boundedly rational shareholders, and particularly smaller investors, can free-ride off market information generated by larger and more informed voters when they make their decisions to buy or sell but when they vote the cost of transmitting this information becomes higher. Efficient market hypothesis, and the event studies which draw on that principle, suggest market prices must be
correct. While behavioral studies do not support that conjecture, at least in its strong form, market prices may still better represent shareholders’ interests than voting. Even though prices may not adjust immediately, they may do so over time, through the Hayekian discovery mechanism. This implies that prices associated with market-tested rules can incorporate knowledge not readily accessible to all voters.

Prices, however, only restrict management’s ability to raise new capital: if all shareholders can do is sell their shares then there is no ex ante control on management malfeasance. Easterbrook and Fischel (1991) argue that the periodic need for additional capital provides adequate market discipline. However, the need to return to markets only constrains managers to the extent the firm needs new capital. If the firm sits on an excess of capital then markets alone cannot constrain management. Shareholders would prefer an ex ante commitment against such behavior but this may be hard to extract.

Concentrated ownership allows shareholders to maintain control but when large shareholders are constrained, as is necessary to protect smaller investors, managers have free reign over the company. The only recourse is the law. Yet the law too defers to managers too. Under the business decision rule only outright theft by managers can be prosecuted (Easterbrook and Fischel, 1991). Should management choose to engage in empire building or buying inefficient perquisites such as corporate jets, then the law remains silent (Ibid.).

---

147 For this reason firms often combine debt with equity to provide finance. The debt component provides a hard requirement while leaving the firm a usually adequate, but not excessive, equity cushion.
IV. The Nature and Role of Corporate Law

If corporate law defers largely to management then it seems to serve little purpose. Critics allege that under state competition the law serves a more nefarious purpose: managers choose the law that suits their own interests over those of shareholders. This section argues that corporate law serves a dual purpose of creating credible commitments by stock issuers and providing knowledge not readily available to market participants. The section also explores some of the limitations of the law and addresses means of addressing those limitations, including federalism.

The law of corporations must first enforce contracts between buyers and sellers of stock. Parties will be reluctant to enter into contracts for intangible property unless they are confident that such contracts create secure property rights. Both investors and firms have an interest in secure property rights during ex ante bargaining. It is only after the initial bargain that one party has an incentive to defect. Secure property rights thus enhance the ability of stock-issuers to commit. Much of the initial development of corporate law dealt with matters of breach of contract by corporate officers or outright fraud by the same, and was resolved by applying the law of property to securities (Banner, 1998; Shleifer and Vishny, 1997).

In both England and the United States the early development of corporate law was largely conducted by judges, who followed common law precedent and adapted rules to fit the new circumstances presented by the emergence of securities (Banner, 1998). These laws evolved over time but under the constraint of stare decisis, the common law principle that judges must respect the precedent established in previous cases. However, the development of the law of corporations continued to develop far beyond the simple law of contract. This transformation
occurred not only under common law courts but under state and federal legislation. Unlike the property cases that preceded securities law, the creation of the joint-stock corporation requires legislative action. Early corporations required a specific charter which could be revoked (Banner, 1998). Only later did states establish general incorporation rules, which lowered the costs, and particularly the rent-seeking costs, of charting a corporation (Ibid).

While the federal government has enacted specific corporate charters it has never established general incorporation rules, leaving states free to establish their own rules (Ibid). Federal courts have, however, established the right of corporations to incorporate in any state and thereafter to do business in every state. This development permitted New Jersey to establish the first rules for foreign incorporation, eliminating the need for the corporation to be primarily located in the state of incorporation and thereafter creating a market in corporate law. This state market has increasingly been intruded upon by federal laws which supersede state incorporation laws (c.f. Bainbridge, 2012). Though the U.S. constitution defines separate spheres of authority to the state and federal governments, federal courts have applied the doctrine of pre-emption to allow states to regulate within broad limits unless the federal government pre-empts state law with its own rules (Greve, 2012).

The expansion of corporate law beyond the realm of contract may be explained through the bounded rationality of shareholders. These shareholders may prefer external constraints on their own behavior to constrain their tendency to vote against their own interests or to renege \textit{ex post} on their contract with debt-holders. Shareholders would prefer to make such agreements in advance but corporate charters, which govern relations between shareholders and management, are necessarily incomplete contracts. Uncertainty makes it impossible to
write terms for every contingency, and a combination of transaction costs and bounded rationality may prevent parties to a contract from writing down all contingencies that may be predicted (Williamson, 1975). Parties may rely on the law to act in unanticipated circumstances.

The law must do more than protect property rights: it must also determine the bundle of rights to be associated with each party to the transaction. Occasionally the law outright restricts an arrangement of property rights, such as the ability adopt a poison pill amendment, or to separate voting rights and residual claimancy once stock has been issued. More frequently the law assigns defaults which parties may contract around. These defaults can have significant consequences when transaction costs are high (Coase, 1937, 1960) or when they cover circumstances not anticipated by either party.

Off-the-shelf bankruptcy procedures deal with an unintended, if not altogether unpredictable, contingency in corporate governance. La Porta et al. (1996) find these procedures have a significant impact on the ability of firms to raise capital. Although parties are free to contract around these procedures, Hart (2000) notes that most firms choose to use the default because relations would require either frequent, and costly, renegotiation, or the existence of a complete *ex ante* contingent contract. Miller and Stiglitz (1999) also note that while external enforcement theoretically allows parties to make such contracts, these are costly to write precisely because parties cannot predict all future contingencies. Furthermore, La Porta et al. (1996) find that countries with strong investor protections (meaning pro-shareholder or anti-management rules) have more developed capital markets. King and Levine (1993a) find that

![Image](https://via.placeholder.com/150)

148 Although there is significant variance in the law, U.S. firms are generally permitted to separate ownership and voting claims in the initial offering.
there are also significant macroeconomic externalities to protection of shareholder property rights in the form of higher GDP growth and greater capital accumulation.

Shareholders are willing to pay a premium, at least tacitly by responding to market information, for rules that increase value even when those rules may act as a constraint on shareholder choice. An investor might benefit from a law that prevents the company adopting a poison pill, even with shareholder approval, because the shareholder knows that as a voter he is boundedly rational. However, relying upon the law merely shifts the problem of commitment to a higher level.

If shareholders are not confident in their ability to restrain managerial misbehavior then it seems paradoxical that they should pass that responsibility to democratic institutions. Shareholders are no less boundedly rational when voting on the law than when voting on corporate rules. Moreover, while private action typically only involves the specific parties to the contract but a political bargain exists between all political actors. Dixit and Olson (2000) show that Coasean bargaining is highly sensitive to transaction costs, making political bargains with a larger a number of parties highly difficult. Nevertheless, several scholars have argued that political markets are efficient. Wittman (1989, 1997) argues that even if voters are boundedly rational that this does not create a bias. By the miracle of aggregation, even a large number of poorly informed voters will cancel out, leaving informed voters to determine the outcome of the election. Becker (1983) argues that competition between special interest groups can contribute to political efficiency as there are greater incentives for favorably affected groups to lobby for efficient policies than for unfavorably affected groups to lobby against.
By contrast, Olson (1982) argues that the impact of interest groups should be detrimental to efficiency. While Olson (1965) previously argued that it collective action problems would impede the formation of groups, his later work notes that the likelihood of forming an interest group is dependent on group-size and time. Due to free-riding, individual members of large but diffuse groups are more likely to be uniformed than small concentrated groups. The larger the group, the harder it is to prevent free-riding by members but, because organizing interests is time-consuming – the likelihood of forming a group increases over time. Contrary to Becker (1983), a similar-sized but longer-established interest group might be more able to lobby effectively for an inefficient policy over a newer group.

Shareholders’ interests are even more diffuse within the political entity than in the corporate, leaving managers even greater discretion. Concentrated outside interests, such as labor unions, may also intervene to the detriment of shareholders. Moreover, while shareholders choosing between firms can free-ride off market prices, such price signals are entirely absent when a single rule is imposed on all firms.

V. **Tiebout Competition in the Market for Corporate Governance**

If political markets cannot discipline government, competition between governments may offer a superior alternative. Tiebout (1956) suggests jurisdictional competition as a market in the provision of public goods which, as Tullock (1971) notes, includes efficient law. Horizontal (Tiebout) competition operates when individuals can move freely across jurisdictions and choose the bundle of goods and services they prefer. He applies the same “as if” assumptions to jurisdictions as competitive markets, including: voter-consumers who are perfectly mobile and perfectly informed, perfect competition between a very large number of jurisdictions, and
local governments that seek to maximize revenue.\textsuperscript{149} Tiebout never advanced the claim that markets in public goods should serve as a mechanism for discovering good law: instead he argues that markets allow heterogeneous voters to self-select into homogenous communities with efficient levels of public-good provision. Tiebout competition also provides the same market discipline to local governments as they to firms. Yet discovery may the more important function of markets in law although both critics and opponents of horizontal competition recognize serious constraints (Romano, 1985, 1987, 2002; Bebchuk, 1992).

Markets in public goods are “lumpy:” there are too few competitors. While the number of states – fifty – may be sufficient in some product markets, the law is a considerably more complex product. Consumers must purchase the entire menu of law on offer and there are far fewer states than combinations of rules, or corporations. Furthermore, there is no free entry into the market. New states cannot (usually) enter the market and while existing states may experiment with alternative combinations, they make the law uncertain by doing so.

Tiebout assumes jurisdictions seek to maximize revenue in the same manner as firms maximize profit. Subsequent public finance literature has sought to evaluate horizontal competition under other assumptions. Mieszkowski (1972; c.f. Mieszkowski and Zodrow, 1989), taking a benevolent government as a starting point, finds horizontal competition detrimental. The author argues that property taxes will increase the cost of capital resulting in higher gross of tax property prices and lower wages and land rents. Local governments will under-provide public goods to avoid driving mobile capital out of the jurisdiction. Miezkowski and similar works reach

\textsuperscript{149} Many legal critics of Tiebout competition ignore the “as if” component of the positive economics, most eloquently defended by Friedman (1953). The relevance of the assumptions is nonetheless worth further consideration.
their conclusions by assuming that capital is mobile and consumes no public goods, but population is fixed and the sole recipient of public goods (see Wilson 1999 for a review). Furthermore, governments are constrained in their ability to levy a head tax.

When governments are both benevolent and unconstrained by information, as a significant portion of the public finance literature assumes (Ibid.), it is unsurprising that the constraints of competition would fail to yield improvements in voter utility. An alternative literature stays closer to the Tiebout / Niskanen (1971) assumption of a revenue maximizing leviathan. Buchanan and Brennan (1980) modify the assumption to incorporate rent-seeking, defining the government utility ($U_G$) as

$$U_G = sT$$

Where $T$ is tax revenue and $s$ is the ratio of tax revenue to extractable rents. However, the value of $s$ is fixed in the pre-constitutional stage, allowing the authors to consider the question of what level of rents should be allowable, but leaving a model of horizontal competition that is not substantively different from a leviathan model. Caplan (2001) proposes an alternative assumption where government utility is defined as

$$U_G = R + \theta G$$

Where $G$ is government spending, $T$ is taxation net of spending (i.e. extractable rents) and $\theta$ is a constant ($0 < \theta < 1$) indicating the ratio at which governments are willing to trade-off rents for spending. In other words, the function assumes governments gain some utility from expanding the size of the bureaucracy (as in a leviathan model) but would prefer to retain revenues as rents (as in a rent maximization model). In addition to assuming Tiebout competition, Caplan
also assumes electoral competition, but includes party preferences in the voters’ utility function. It is this assumption which admits rents into the model which will always accrue to the advantaged party up to the point at which the median voter is indifferent between two parties. The model predicts this will occur regardless of horizontal competition.

Caplan’s model assumes jurisdictions rely on a property tax and that when an individual moves to a jurisdiction, he fully invests in immobile property. In equilibrium the return on capital must be equal across jurisdictions meaning that any variation in rent-seeking must be fully reflected in property prices. Immobile property is associated with a bundle of rights. Rent-seeking activity reassigns those rights to government leaving the owner of the property with a diminished and less valuable bundle. The owner can only resell the rights currently associated with the property and once those rights are expropriated by government, cannot be sold by the owner of the property.

It is from these assumptions Caplan obtains the result that horizontal competition would be ineffective. However, the assumption that capital is entirely immobile is too restrictive. Indeed, Caplan’s model builds on an earlier work by Epple and Zelenitz (1981) which allows a combination of mobile and immobile capital. As with Caplan, the authors assume a mobile population and immobile capital (land which is tied to a particular jurisdiction). However, the housing supply is held to be variable and dependent on the availability of land and the price of housing. The model assumes all governments maximize the same profit function:

\[ \pi = tP_H(t, G)H_S[P_H(t, G)] - N(t, G)G \]

\[ ^{150} \text{i.e. total rent extraction exactly equals the median voter’s preference for the advantaged party.} \]
Where \( t \) is the level of a flat property tax, \( G \) is government spending, \( H_S \) is supply of housing, \( P_H \) is the price of housing, and \( N \) is population. Although the model omits key variables relevant to this study – namely it is unable to account for commitment mechanisms – in an adapted form it may serve as a useful starting point. The adjusted government utility function becomes:

\[
\pi = tP_K(t)K_S[P_K(t)]
\]

The price of housing is here replace by \( P_K \) denoting the price of capital (i.e. the trading value for stock in the jurisdiction), \( K_S \) is the supply of capital (i.e. the value of stock issued in the jurisdiction excluding government effects), and \( t \) now represents pure rent-seeking but remains a constant portion of capital. The latter term assumes that the optimal set of rules is known and that governments intentionally deviate from those rules only to extract rents.

As with most models in this literature, Epple and Zelenitz make expenditures a function of population but taxation is a function of capital. That is there are no head taxes and no public goods accrue to capital. Consequently, population is a pure cost and capital is a source of revenue. These assumptions, if perhaps unrealistic in some circumstances, are well suited to the competition in incorporation. While physical capital may consume public goods, incorporation consumes only the law whose minimal cost may be safely ignored. Both government expenditures, and population which now imposes no costs and provides no revenue, may also be ignored. This creates no problem as governments are competing for capital, not people. In Epple and Zelenitz original model housing supply is a variable given by the function

\[
H^j_S = Lh^j_S(t^j_H)
\]
Where $H^j_S$ is housing supply in jurisdiction $j$, $Lh^j_S$ is total supply of housing given as a function of price in $j \left( P^j_H \right)$. There are no physical boundaries to constrain how many firms may be incorporated in the same jurisdiction. However, the responsiveness of the supply of stock issuances in $j$ may be rewritten as

$$\frac{\partial K^j_S}{\partial P^j_K} = \theta \frac{P^j_K}{K^j_S}; \quad \theta = \theta(C_R)$$

where elasticity of supply of stock issuances ($\theta$) is a function of the cost of reincorporation. The investors income constraint is given by $l = P_K(1 + t) + B$, where $B$ is a numeraire good representing consumption. Following Epple and Zelenitz, individual choose their demand for stock ($K^j_d$) to maximize utility taking rent-seeking as a given. This gives the indirect utility function

$$V\left[ P^j_K(1 + t) \right] = \max_K \left[ K^j_d, l - P^j_K k^j_d(1 + t) \right]$$

Giving the individual demand function for stock issuances in $j$

$$k^j_d = k^j_d\left[ P^j_K(1 + t) \right]$$

In equilibrium there is no movement, thus marginal utility is equal across all jurisdictions.

$$\frac{\delta V}{\delta P^j_K} \left[ P^j_K(1 + t) \right] = \frac{\delta V}{\delta P^j_K} \left[ P^j_K(1 + t) \right]$$

Furthermore, each unit of investment should yield an equal return, giving constant utility across jurisdictions.
\[ V[P^i_K(1 + t^i)] = V[P^i_K(1 + t^j)] \]

Intuitively, individuals seeking to maximize their return will receive the same rate of return on their investments net of government influence. This may be further demonstrated by taking J as the number of jurisdictions and \( \eta \) as the elasticity of demand for stock. We can write the effect of a change in the level of rent-seeking on capital asset prices as

\[ \frac{\delta P^i_K}{\delta t^i} = \frac{P_K[\eta - (J - 1)(\theta - \eta)]}{j(1 + t)(\theta - \eta)} \]

With one jurisdiction this becomes

\[ \frac{\delta P^i_K}{\delta t^i} = \frac{P_K(\theta - 2\eta)}{(1 + t)(\theta - \eta)} \]

Under competition, an increase in rent-seeking will be exactly countered by a fall in \( P_K \), leaving gross-of-tax price \( [P_K/(1 + t)] \) unaffected. To demonstrate this, take the upper limit \((J \rightarrow \infty)\)

\[ \frac{\delta P^i_K}{\delta t^i} = -\frac{P_K}{(1 + t)} \]

Multiplying by \((1 + t)\) demonstrates that in perfect horizontal competition the effect of an increase in the tax rate leaves gross-of-tax price unaffected.

Retaining the assumption of profit-maximizing government, we can obtain the impact of competition on government behavior by substituting elasticities into the government profit function. The function is the same for all governments so superscripts may be ignored to give

\[ \frac{\delta \pi}{\delta t} = \frac{P_K K_s}{(1 + t)} \left[ 1 - t\theta + \frac{t\theta(1 - \theta)}{j(\theta - \eta)} \right] \]
Assuming $\theta$ and $\eta$ are constants and solving for profit-maximizing level of $t$ yields

$$t = \frac{J(\theta - \eta)}{\theta [J(\theta - \eta) - 1 - \theta]}$$

The result demonstrates that the ability of governments to raise rents is constrained by relative elasticities of demand and supply, and by competition. When supply is perfectly elastic but demand is not, jurisdictions can exploit demand elasticity only in the absence of competition.

With no competition ($J=1$)

$$\lim_{\theta \to \infty} t = \frac{-1}{\eta + 1}$$

The jurisdiction will extract positive rents but are constrained by the elasticity of demand for stock issuances, as investors switch to consumers goods at higher levels of rent extraction.

Within Epple and Zelenitz model, each jurisdiction is effectively a Bertrand oligopolist, giving the result that for values of $J$ greater than one

$$\lim_{\theta \to \infty} t = 0$$

Thus when the supply of capital is perfectly elastic and there is competition between jurisdictions, rent-seeking is effectively eliminated. To achieve perfectly elastic supply in each jurisdiction requires that firms be able to reincorporate costlessly in response to any change in rent-seeking. Yet this assumption undermines very purpose of the law which is that firms be able make enforceable commitments. If the firm could reincorporate in another jurisdiction at no cost, or at least do so at no more cost than changing the corporate charter, then legal restriction would merely be another self-enforced restraint.
Alternatively, we may consider the case of perfectly elastic demand for stock issuances. This is more realistic given that investors can usually buy and sell stock at a reasonably minimal transaction cost. As the elasticity for demand for stock issuances increases toward perfectly elastic supply, $\eta \to \infty$ to yield

$$
\lim_{\eta \to \infty} t = \frac{1}{\theta}
$$

The price of stock is exogenous but the jurisdiction can still exploit rents from inelasticity in the supply of stock issuances. It is worth noting that without perfectly elastic supply, perfect competition yields the same result

$$
\lim_{f \to \infty} t = \frac{1}{\theta}
$$

Perfect competition is less realistic than perfect elasticity of demand for stock issuances within each jurisdiction. However, given that the former seems at least plausible and that there is some competition between jurisdictions, we may proceed with some confidence on the premise that inelasticity of supply of stock issuances drives the ability of jurisdictions to rent-seek. As noted above, a degree of inelasticity of supply is necessary to overcome the commitment problem.

both observe that when one party to a contract can chose the jurisdiction \textit{ex post} the result of horizontal competition will necessarily be inefficient (Greve, 2012; Zywicki, 2005). When an individual or corporation can where to pursue a tort, or file for bankruptcy \textit{ex post}, rather than contractually specifying in advance which jurisdiction will govern the dispute, they will only consider their own preferences for law. In bankruptcy law, jurisdictions will compete to provide
more lenient law (under the restrictions of rent-seeking described above). Directors would choose bankruptcy law that favored shareholders and their own contracts over debtor interests. In the provision of tort law, competition results in law that favors the plaintiffs who choose where to file over defendants.

This implies an optimal level of supply elasticity ($\theta^*$) such that $\min_{\theta} E(\theta) = t(\theta) + m(\theta)$. Where $E$ is total extraction, $t$ represents rent-seeking by government and is increasing in $\theta$ in equilibrium, and $m$ is agency cost and is decreasing in $\theta$. While a degree of inelasticity is necessary, Epple and Zelenitz rely on using land as an input to housing to provide that immobility. Eliminating land as an input allows for the possibility that supply too is elastic (i.e. that firms will incorporate in what jurisdiction yields the highest return). However, while incorporation is not a physical investment, corporations are still somewhat tied to the jurisdiction in which they are incorporated. The cost of legal fees and use of the proxy machinery to acquire shareholder consent required for reincorporation is non-trivial (Romano, 1985, 1987, 2002; Bainbridge, 2012). Costs may be higher if the firm needs to replace its existing corporate counsel.

This cost, like the others, is not imposed upon the directors but split between shareholders. Directors may nonetheless be heedful of these costs. Any reduction in the value of the managers’ shareholding, or lessening of the firm’s chances of survival, will factor into the directors’ decision. In that the total cost of reincorporating in another jurisdiction is higher than altering the corporation’s rules, they are less likely to attempt the latter. When directors are forced to adopt a bundle of inefficient rules they impede the firm’s ability to return to capital

---

151 Though $t$ may initially be decreasing in $\theta$, see anon.
markets, at detriment to their own interests. Moreover, if the firm must expend capital to fight a takeover battle it may need to return to capital markets to replenish the available equity. Finally, if the firm fights a takeover bid by moving the firm further from maximum efficiency it may increase the likelihood that other firms attempt a takeover in the future.

At least some of these measures can be achieved through *ex ante* contracting. The firm’s charter can set a high bar for gaining shareholder approval to relocate, raising the cost of relocating to both more efficient and less efficient jurisdictions. Investors can also impose asymmetric costs that raise the cost to directors of moving to a less efficient bundle of rules, including restricting the firm’s pool of free capital and altering the director’s compensation bundle. Thus investors may have the power to influence value of $\theta_i$ (elasticity of supply in a given jurisdiction) when $P^i_K$ is lower than in another jurisdiction. That is, a firm moving from $J_i$ to $J_j$ would incur additional costs only when $P^i_K < P^j_K$. When the converse is true, the move would not restrict a firm’s access to capital and thus the additional measures would not raise the cost of reincorporating. While asymmetric costs do not raise the cost of moving to jurisdictions with less efficient rules, these measures are by no means costless. Rather, they are the same measures investors can apply to other agency costs which can also reduce efficiency by eliminating useful director discretion (such as the responsiveness to changing market conditions).

Asymmetric measures are also vulnerable to path dependency favoring the leader in incorporations (Romano 1985, 2002). Investors favor certainty in contracting, which is provided by an extensive code or set precedents. The leader in incorporations, currently Delaware, has the greatest experience in adjudicating conflicts between parties. Delaware can thus offer a
degree of legal certainty that other states cannot, even if those states have otherwise more efficient law. While the alternative set of rules might provide a long-run benefit, no individual firm has an incentive to switch to the alternative set of rules and bear the cost of developing precedent.

The optimal degree of capital mobility, and the mix of symmetric and asymmetric controls used to achieve this, will depend in part upon the ability of the jurisdiction to commit to providing future good law. La Porta et al. (1996) observe that investors substitute direct control measures for good law. When the firm initially incorporates and raises equity for the first time, investors are free to choose the most efficient bundle of rules. If states can commit to not increase rent-seeking then investors can choose to impose high costs on any relocation and rely more heavily on the law to provide checks on managerial indiscretion. If the states are unable to commit to future good law then investors wish to keep their capital relatively mobile as a guard against future rent-seeking. To overcome the efficiency-reducing nature of asymmetric controls and the danger of path-dependency, investors will desire a legal system that not only provides efficient rules in the present but also a commitment to providing efficient rules in the future.

VI. The Challenge Of Making Credible Commitments

Investors will seek secure property rights built upon a rule of law, defined by Hayek (1960) as the universality of law and the absence of *ex post facto* law making. Not only must the legal institutions of the state be powerful enough to protect investors but such power must somehow be restrained (Weingast, 1995). A study into the development of electrical infrastructure in 19th century America found consistent evidence that voters, given the power to do so, would seek to exploit immobile investments after the investment was made. Voters in already electrified
areas supported constitutional amendments permitting municipalities to impose price controls (Troesken, 1997). Voters still awaiting electrification opposed the amendment, knowing that price controls would allow them to exploit the generating plants once the initial immobile investment, and that this would discourage the initial investment. However, once that investment was made they had no incentive to continue opposing such amendments. Just as with the initial dilemma confronting investors and the issuers of capital, both parties would prefer to make an efficient enforceable contract at the outset but are unable to.

Several legal scholars (Romano, 1985, 1987, 2002; Bainbridge, 2012) have argued that rent-seeking is not only consistent with a competitive market in law but may be necessary to maintain it. If the net present value of future rents outweighs the benefits of a one-time defection then the state can commit to not defect.\(^{152}\) This requires that the state receive positive rents from incorporation and that these be non-trivial compared to the immediate benefits of rent-seeking. While most states and the federal government receive a relatively small portion of their income from incorporations, Delaware receives roughly a quarter of all state revenues through a franchise tax imposed on all firms incorporated in the state. Firms incorporated in their home state are frequently employers, and an interest group capable of swaying legislatures in their own right. In Delaware, out-of-state firms comprise the majority of incorporations, leaving each firm relatively little political power. Firms incorporated in the state are as likely to be the initiators of takeover as they are the target, eliminating the normal anti-takeover bias to protect domestic factions.

\(^{152}\) That is if \(\sum c_n / \beta^n > d\), where \(c_n\) is the value of non-defection in period \(n\), \(\beta\) is the discount rate, and \(d\) is the benefit of defection, it is beneficial not to defect in the current period.
The use rent-seeking as a commitment mechanism introduces an inherent inefficiency into horizontal competition and, more significantly, increases path dependency. If a state wishes to challenge Delaware for the incorporations market it cannot do so by offering a lower cost of incorporation as this eliminates the competitor state’s ability to make the same commitment. Moreover, the new state would not initially be dependent upon rents and would thus have more difficulty in making a credible commitment to future non-predation.

As with the model presented above, rent-seeking-as-commitment assumes that governments are profit-maximizing.\textsuperscript{153} By contrast, Olson (2000) demonstrates that while a single residual claimant (a dictator) may behave this way, the multiple competing interests that govern a democracy will not. While these competing interests can conceivably bargain to an efficient settlement, such a bargain is unlikely.\textsuperscript{154} Contra to Tiebout’s claim that once efficient rules are established voters should not wish to change those rules, boundedly rational individuals may change rules against their own interests just as within the firm. Acemoglu (2003) finds that the inherent unenforceability of political contracts further raises the cost of political bargaining as long term interests cannot contract with short-term interests to avoid inefficient rent-seeking.

States may seek to commit through constitutional measures. Constitutions, like corporate charters, must establish the terms by which the law is amended and can raise or lower the cost of changing the rules (Buchanan and Tullock, 1962). Raising the cost of changing rules is beneficial when rules are efficient, but efficient law can become inefficient over time as innovation and other factors alter the composition of the optimal bundle of rules. Enforcement of property rights allows stock issuers to commit to the bargain not only by permitting external

\textsuperscript{153} Or, at least, revenue-maximizing

\textsuperscript{154} See discussion above.
enforcement of a contract but by providing rules for unanticipated circumstances. If the law could never be changed in retrospect then it would be perennially limited by the foresight of its authors. The law may prevent beneficial innovation or worsen crises as in the most recent financial crisis.

If all members of the political entity have the ability to block change then someone not a party to the original agreement may hold-up change. Unions might insist that a beneficial shareholder protection can only be created if the agreement between the firm and its workforce is also amended. Yet the barriers to change can also be set too low. If only a majority of the political entity is required to change the rules then one party might successfully change the rules to the detriment of the other. The law must not only provide external enforcement of property rights; it must also credibly commit to provide for beneficial change.

Olson (2000) demonstrates that a sovereign (in this case, voters) can overcome this impasse by delegating power to an independent body. Delaware, which leads the competition for state charters, relies on a specialized Court of Chancery to handle all disputes relating to incorporation. The state also requires a two-thirds majority to change the articles of incorporation and informally, the state corporate bar, which derives most of its income from incorporations, has a veto on any changes (Romano, 1985, 2002). It is not the legislature but the Bar that acts as a stationary bandit, maximizing long-run rents by blocking rent-seeking measures by other groups while maintaining inefficient procedures that increase legal costs.
VII. Albion versus Leviathan: The Debate Between Code and Common Law

Historically, power has been delegated to common law courts, and while much of corporate law is codified, its enforcement and interpretation remains with the courts. By contrast, since the 1930s lawmaking at the federal level has switched significantly to expert-managed regulatory agencies (McKraw, 1984). Executive agencies, and some departments, are considered a part of the executive while independent agencies such as the Securities and Exchanges Commission (SEC) have commissioners serving fixed terms. These agencies typically require significant permanent staff and are themselves vulnerable to rent-seeking in the form of seeking an expanded budget or authority (Niskanen, 1971) but also to regulatory capture (Stigler, 1971; Peltzman, 1976). Congress exerts significant control over federal agencies through control over the budget (Weingast, 1984; Weingast and Moran, 1983; Weingast and Marshall, 1988). This practice has only been recently violated through Dodd-Frank, by the creation if the CFPB with its power to draw revenue directly from the Federal Reserve. Zywicki (2012a) argues this change in practice will exacerbate the problems innate to independent agencies, yet congressional control may introduce new problems still.

As with shareholders in firms, the congressional monitors and the voters who must monitor them, suffer the same incumbent problems of overseeing specialist regulators. Studies of the oversight process suggest that it is used largely to protect key constituents against adverse regulation or to protect against competition (Ibid.). The devolved regulatory process thus remains highly political but serves the purpose of reducing the costs to changing rules over a constrained purview.
Unlike regulators, federal judges may not have their pay reduced and cannot be dismissed without significant effort. State judges are more likely to serve shorter terms although they may be elected independently of the legislature. State laws are thus more prone to change, allowing for potentially greater innovation but also for less predictably in the law. By contrast with regulatory agencies, common-law judges are rarely experts in the subject matter they consider, yet the common law can, and does, achieve efficiency by drawing on the standard practices which usually govern parties (Cooter, 1996; Pritchard and Zywicki, 1998). By applying custom to disputes, the common law can access decentralized knowledge and this avoid the limitations of bounded rationality facing regulators who write a code.

Under the common law, judges decide cases first on the basis of the specific facts and then apply a principle. Their rulings apply only to the parties to the case before them. Nevertheless, judicial precedent is highly important. Under the English doctrine of *stare decisis* judges are expected to follow precedent set by higher courts established in previous similar controversies, but may take another position with justification. Thus novel circumstances can generate new precedent. Ultimately, it is the appeal process and actions taken by higher courts that will either ratify or deny a revision of precedent. In this sense, the common law process is dynamic and organic.

In common law systems, statutes are written precisely and cover one part of the law in specific detail. By contrast, the code law describes principles that must be applied broadly and not just to specific facts. In place of the detailed jurisprudence of common law, the judges or regulators who interpret code law must find ways to explain and apply the code in situations not considered by the original authors. While common law is continuously evolving, code law
changes when, and only when, the legislature or agency changes the code. Regulators can change the code within the confines of the statutory law but novel problems may require legislative action. Moreover, the process of changing the code makes no account of whether the change would alter existing relations or create new relations for novel circumstances. The easier it becomes to adapt under the code the harder it becomes to prevent rent-seeking and whereas common law judgments involve only to parties to a controversy, code law involves all members of the state.

The common law practice of deciding the law only after the facts of the case are known can avoid the limitations of bounded rationality but creates the prospect of uncertainty in the law. This uncertainty must be balanced against the inherent uncertainty created by the conflict between ongoing change and a static law. Hayek (1960) argued that no lawmaker can predict the all possible circumstances the law may be required to govern. Whereas code law starts with an inclusive set of principles and seek to apply these to new facts; common law jurisdictions start with the facts of the case and decide principles after. Thus while common law judges must only decide the principle on known facts, the writers of code law must decide the law before the facts are known.

Changing to the code is potentially more disruptive than setting a new precedent in response to new facts because the code must specify every possible circumstance. Although this leads to a code that is often more complex than the common law, it also leads regulators to write prescriptive standards which reduce the freedom of individuals to bargain in the shadow of the law. If regulators could change the code freely this may create excessive legal uncertainty. Consequently, the statutorily-defined process for changing the regulatory code can take several
years and requires judicial oversight (Daugiradas, 2005, Pierce, 1995). Agencies frequently prefer to use interpret existing rules (Mantel, 2009), or when regulatory change is needed before any private action can occur, the agency may choose inaction, effectively proscribing innovation. Along these lines, David and Brierley (1985) note that while common law systems are open to creating new rules for new facts, civil law systems are closed until revised because every situation is already accounted for within the principles.

The relative strength of the common law in resolving disputes is the ability of the law to resolve disputes in the manner the parties would have chosen to do so had they been able to write an ex ante contract. In a crisis, the code-based regulatory system moves slowly. The executive has the power to waive rules in a state of emergency but only on an ad hoc basis. When laws change slowly, as in the financial crisis, inflexible rules such as the Basel II requirements can worsen the crisis, but when requirements are relaxed the record of Dodd-Frank and Sarbanes-Oxley suggests that the result will be rent-seeking rather than beneficial rules. Statute law takes precedence in both types of jurisdiction but only the common law can change without legislative action. This makes the law more flexible but also allows for greater restraint upon the legislative body.

Posner (2007) argues that judge-made law is less likely to lead to rent-seeking. When litigation does seek to change the law, common law judges consider the cases of only those parties with standing. In code law, lawgivers consider the interests of all parties, regardless of standing, including tightly organized special interest groups. Rubin (1977), Priest (1977), and Goodman (1978) argue that the litigation will produce relatively efficient law if inefficient laws are litigated.
more frequently. By contrast, Tullock (2005) argues that litigation is at best a public good no different to lobbying with concentrated interests more willing to spend resources to change the law. He concludes that the common law may be more costly than code law because the majority of resources are consumed by the adversarial process while in the code law the greater proportion of resources are dedicated to the inquisitorial process. However, the costs of lobbying are mitigated by legal certainty. When the law is settled parties can usually still bargain in the shadow of the law (Coase, 1960; c.f. Rubin 1982).

While both Posner and Hayek argue in favor of the common law, they do so under very different theories (Zywicki and Sanders, 2008). Posner assumes that a wise judge can determine the most efficient law whereas Hayek argues that the law embodies information not available to any individual (Ibid.). As well as yielding greater efficiency, Hayek (1973) predicts the common law should afford greater protection for contract and property rights. He distinguishes between code law judges, and regulators, who are a part of the political bureaucracy, and common law judges who are considered independent. Judges in common law countries are therefore potentially more willing to protect property rights and the right of contract from the government. Olson (2000) also predicts that securing property and contract rights against the government is necessary to generating investment, noting that countries with stable contracting are most likely to have developed financial industries.

La Porta et al. (1996) find that common law countries have better legal rules for the protection of investors than code law countries. The authors’ results support their argument that although firms can opt out of these rules there are transactions costs associated with doing so. La Porta et al. (1997) looks at both the quality of investor protection arising from legal rules and the
enforcement of contracts, finding that poorer investor protection results in less developed financial markets. Furthermore, the authors find that countries under French civil law provide the poorest investor protection and have the least developed capital markets. King and Levine (1993a) find evidence that countries with more developed financial systems have greater rates of GDP growth, physical capital accumulation, and more efficient allocation of capital. Mahoney (2001) finds that common law countries experience faster growth in gross domestic product, but argues common and code law countries frequently arrive at the same set of rules.

Mahoney attributes faster growth to Hayek’s theory that the common law places stricter limits on the government thus providing for greater protection of property and contract rights. However, states can, over time, erode the common law through statutory law or by empowering regulators to produce a code. Judges themselves may deviate from the common law. While Landes and Posner (1975) rely upon the presumption that judges are motivated by non-pecuniary incentives to produce good law there is no reason to believe judges are exempt from agency problems or even that the judges conception of social good equates to efficient law. In the case of Delaware, it is the close links between judges and the Bar that provides concrete incentives for judges to keep the law competitive. These incentives are, in turn, provided by the rigors of horizontal competition.

VIII. Issues of Federalism

Weingast (1995) argues that federalism – meaning the combination of horizontal competition with the existence of a federal government – must itself be considered a mechanism by which jurisdictions commit to secure property rights and the rule of law. States should favor, at least from an efficiency perspective, market-preserving policies including a commitment to property
rights. Yet special interests within states can undermine market institutions. By overcoming
domestic faction, federalism benefits Delaware, which attracts more out-of-state incorporations
than any other state, but it also benefits states such as California that cannot commit as easily to
secure property rights and simultaneously the promotion of innovation. As noted above,
Delaware’s commitment to secure property rights requires that rents (the franchise tax and
revenue to the Delaware bar) from incorporation exceed the immediate benefits of rent-
seeking. That is only possible because of Delaware’s extreme small size (as measured by
government revenue) compared to the overall market in incorporations. While other small
states can compete with Delaware, no larger state, and certainly no unitary state, could achieve
the ratio.

While, Weingast’s (1995) federalism assumes that capital is mobile, the above analysis has
demonstrated that competition combined with internal commitments mechanisms. However, if
states can renege upon commitments to the common law and property rights in the absence of
competition, then they may renege as easily on the commitment to open trade with other
states. Weingast overcomes this limitation by assuming that federal government also serves as
the enforcer of political bargains between states (Ibid.).

Economists typically follow Oates’ (1972) concept of federalism as fiscal federalism, and hold
this to be separate from horizontal competition. Oates argued that in the absence of
externalities, devolving decision-making power to lower bodies will increase the responsiveness
of government by reducing the size of the bargaining group. This leaves only the question of
where power should reside to maximize efficiency, and ignores the relations between the levels

\[156\] But see Breton and Scott (1978) and Breton (1996) for a notable exception.
of government. Hayek’s (1939) earlier argument for federalism, which favored the creation of a European federalism around the American model, is more consistent with the federalist concept currently adopted by political scientists and the legal profession.\textsuperscript{157}

The constitutional convention exceeded a mere compact between the states and instead provided an independently powerful enforcer. The constitution prohibits states from raising barriers to trade against other states either directly, in the import-export clause, or indirectly, in the commerce clause. The danger addressed by federalism is not just that states should renege on their compact with other states, but that they should renege on the political bargains between their own citizens. By enforcing the free movement of capital and people, the federal government raises barriers to rent-seeking within states. Greve (2012) argues that American federalism achieves this through the dormant (but actively enforced) commerce clause. Corporations are permitted to incorporate in any of the fifty states, but states are prohibited from excluding corporations domiciled in other states or from imposing regulations on the same that would impede interstate commerce.

Whereas Oates merely argues for where power should reside within government, based on efficiency criteria, and whereas Tiebout argues only for the competitive advantages of jurisdictional competition, the complete case for federalism goes beyond both arguments. Horizontal competition and vertical competition (between levels of government) are both necessary components of effective federalism. However, the federal government does not compete with states in the same manner as states compete with one another.

\footnote{\textsuperscript{157} Hayek’s work on the subject preceded the creation of the European Union by some years. It is less clear that Hayek favored the current form of European federalism.}
Horizontal competition enforces property rights because firms can incorporate in Delaware, which can credibly commit to secure property rights, instead of California which cannot. If special interests cannot seek rents in Delaware they may turn to the federal government, which does not face competition. The history of Dodd-Frank, Sarbanes-Oxley, and previous federal legislation, shows that this has indeed been the case (Romano, 2004; Bainbridge, 2010). Special interests, unable to lobby Delaware, instead turned to the federal government. Thus the problem of commitment, first experienced by the firm, and then by the state, is now experienced by the highest level of government. Having created a federal government that is strong enough to protect free markets, that government must now be restrained such that it cannot erode them (Weingast, 1995).

This problem has been given little attention in economics which has concerned itself primarily with the problem of assignment of responsibility. By contrast, political science has concerned itself primarily with where residual power is allocated. In Oates’ taxonomy, when a central government empowers local government to make choices, but retains the discretion to determine the allocation of decision-making, this becomes a federal system. In the political scientist’s lexicon the system remains centralized. To become truly federalist, the national and state governments must have separate competencies. These competencies cannot merely be stated in “parchment barriers” but must be internally enforced though intuitional arrangements.

As with states, constitutional constraints can raise the cost of altering rules. The effectiveness of federalism depends upon the existence varying costs of change at different levels of government. Whereas, the economic federalism literature treats responsiveness as an exogenous variable, determined by distance from the electorate, this logic deserves to be stood
on its head. The level of responsiveness at each level of government is chosen in response to the different responsibilities of each level. Rather than federal government delegating down to more responsive governments, states delegate authority up to governments more capable of commitment.

Varying levels of responsiveness permit federalism to function. Making the federal government less responsive raises the cost of rent-seeking at the federal level. States, with greater responsibilities for governing the complex relations between individuals, cannot survive the same constitutional constraints as are imposed upon the federal government and instead rely on external commitment mechanisms to secure property rights. The federal government, by leaving the majority of lawmaking to the states, can afford to be less responsive.

The more rigorous restraints on the federal government take several forms. Federal judges are appointed for life while state judges are usually appointed for shorter terms. This makes state judges more responsive to immediate concerns but gives federal judges greater independence from the political process (Landes and Posner, 1975). Until the 17th amendment brought about the direct election of senators, the federal government was held in check by the states themselves, relying on the tendency of local factions to resist the expansive tendencies of the national government (Zywicki, 1997). Finally, the federal constitution is considerably harder to change than most state constitutions and requires the consent of three-quarters of states.

While the federal government has fewer responsibilities for managing the complex relations between individuals, it must still manage the relations between the states. The federal constitution not only defines the separate competencies of each of the branches of government but also lays out the relationship between the federal and state governments and between the
states themselves. This, in turn, allows the courts to enforce the compact between the states with fewer opportunities for rent-seeking than through executive or legislative enforcement.

Just as an *ex ante* contract can be diverted by its original purpose by a combination of uncertainty and conflict of interest, so the advancement of technology, particularly the new communications technology of railroads which transformed the US economy in the 19th century, presented new challenges and new conflicts within the federal structure. Greve (2012) notes that for much of United States history, states have sought to impede the free movement of capital in ways not originally anticipated by the constitution’s architects. The courts in turn have found new precedents governing relations between the states, including the finding that firms incorporated in any state may conduct business in every other state. Just as the common law allows the judiciary to impose rules that would have been included in *ex ante* contracts had the parties possessed perfect foresight, so it allows a somewhat active judiciary to permit beneficial constitutional change while avoiding the hold-up and rent-seeking problems associated with formal amendment (Zywicki, 1997).

Although the delegation of authority to the courts may provide similar benefits in arbitrating relations between the states, as between individuals, the federal government is not subject to competition. Moreover, the federal Supreme Court does not rely upon the spontaneous generation of rules Hayek argues is vital to common law courts, but follows a constructivist approach to rulemaking (Zywicki and Sanders, 2008; c.f. Hayek, 1973; Smith, 2003). The combination of the lack of a discovery process to access efficient rules and a competitive process to overcome rent-seeking, can allow inefficient rules to survive.
Olson (1982) argues that inefficient rules will not only survive but will accumulate over time. In a stable society, special interest groups, once formed, do not dissipate naturally and so their number grows over time, adding to rent-seeking and encumbering the process of creating new beneficial rules. Regulations lead to yet more regulations, creating and closing loopholes, in a process that “increases the complexity of regulation, the role of government, and the complexity of understandings, and changes the direction of social evolution” (Ibid. at p74). In American corporate law this increase is manifested in a stable, and perhaps irreversible, power shift from state common law courts to the more easily captured federal regulators. Most recently, the Dodd-Frank Act expanded federal regulator’s authority over corporate governance in a detailed 848 page bill. Yet, while Olson predicts the outcome of Dodd-Frank, his theory ignores the role of crisis as a catalyst for change (c.f. Higgs, 1987). This most recent legislative action, in direct response to the 2009 financial crisis, follows a consistent trend in Anglo-American securities law of crisis directly precipitating a regulatory response (Banner, 1998).

IX. The Impact of Crises

The evidence reviewed by Higgs (1987) and Banner (1998) presents a strong empirical case that crisis lowers barriers to changing rules.\(^{158}\) A temporary reduction of the cost of changing rules is not necessarily detrimental. Crises can also provide important information, demonstrating which market practices, and which rules, are efficient. The demand for law changes while rigid

---

\(^{158}\) The theoretical case is less clear. Both federal and state governments have facility for overriding barriers to change during an emergency. It is also possible that crisis facilitates bargaining between groups either by weakening the power of Olsonian groups or by creating incentives that compel groups to cooperate. Both Banner and Higgs, however, focus on how public perception of crisis alters the demand for law. This is consistent with the problems described above of expert monitoring of regulators where boundedly-rational voters may not be able to fully assess the quality of law but instead rely on more easily accessible metrics for success (i.e. avoiding crisis).
rules, such as the Basel II standards, can become harmful by preventing markets from responding to the new circumstances. As shown above, the law can improve welfare by responding to changing circumstances by developing precedents that reflect conditions parties would have preferred if they had perfect foresight. The common law makes specific provision for crisis under the doctrine of *Force Majeure*, which provides for a relaxation of property and contract rules. Efficient corporate bankruptcy law, which is designed to compel parties to renegotiate their property rights in a firm facing liquidation, is associated with higher growth (Hart, 2000; Claessens, 2002).

Relaxing rules in a crisis is thus potentially beneficial, but Zywicki (2012b) argues that weakening property rights can create uncertainty, discourage new investment, and impede recovery. Relaxation of rules can create a moral hazard problem if individuals expect rules to be relaxed in future crises and prevents the law serving its function as an external commitment mechanism. Making it easier to change rules also lowers the barriers to rent-seeking and can lead to the regulatory ossification described by Olson. Extensive studies by Bainbridge (2010) and Romano (2004) of the legislative history of recent expansions of federal corporate law (Dodd-Frank and Sarbanes-Oxley respectively) support Zywicki’s contention: instead of responding to new information, the law imposed new rent-seeking measures that had been proposed, but were unable to garner support, prior to the crisis. However, these studies look only at legal change at the federal level. While the pressures of Olsonian decline affect both state and federal government, crises impacts horizontal and vertical competition differently.

---

30 Williston, Contracts §§ 77:31

239
Investors can also use newly generated information to evaluate market practices and rules under horizontal competition. Whereas, immobile capital previously constrained investors’ ability to seek out new rules, crises have the effect of destroying existing capital and weakening path dependency. In a study of the effect of natural disaster on jurisdictional competition for population, Beeson and Troesken (2006) found that jurisdictions that were attracting new people prior to crisis quickly replenished lost population while jurisdictions that were losing population did not. Similarly, when jurisdictions competing for capital, temporarily increased capital mobility will benefit while jurisdictions with efficient rules while jurisdictions with formerly efficient rules which are able to exploit path dependency will fail to attract new capital.

While crisis may open opportunities for increased rent-seeking at the state level, it also brings constraints. States will rely on similar mechanisms, such as using common law courts to provide flexibility while committing to efficient rules in the post-crisis period. When states fail to do this, capital will shift to alternative jurisdictions with efficient rules. Investors will leave a state with inefficient law, as was the case when Delaware overtook New Jersey (Banner, 1998). But rather than a guarantee that the prevailing law shall be efficient, or even that the state with the most efficient law should immediately lead the market for incorporations, the process of competition is uneven and not spontaneous.

Yet if horizontal competition falls short of Tiebout’s ideal, in its absence rent-seeking effects will dominate at the federal level. Rent-seeking will tend to shift the creation of law from common law courts, which will provide greater restraint on rent-seekers, to the code-based regulations which can be more easily influenced by the political process. Moreover, because the relaxation of rules created by crisis also affects the constitutional rules governing the relationship between
the states and the federal government, rent-seeking will also tend to shift power from the states to the federal government.

While regulators can and do take action to avoid future crises, they cannot avert crises altogether. Critics of horizontal competition argue that state-based rules will increase opportunities for moral hazard by failing to account for externalities (c.f., Bebchuk, 1992). Although both stock-holders and the issuers of capital can choose the venue for determining disputes, when a firm faces crisis that requires a relaxation of the rules, corporate law must also arbitrate other claims against the firm. Debtors may account for the firm’s state of incorporation but customers and suppliers who may have claims against the firm are less likely to do so, and those with tort claims against the firm have no control over the firm’s chosen jurisdiction.\textsuperscript{160}

For these reasons, there is no choice in \textit{ex post} bankruptcy law. The federal government sets default rules which parties can only avoid by explicit contract. However, bankruptcy law is only applied once the firm finds itself in crisis. Creditors cannot compel bankruptcy unless the firm defaults. The owners of the firm, who are immune to the full cost of risk-taking under limited liability, would prefer the firm to fully account for the benefits of risky ventures but not the costs which fall on creditors. Moreover, excessive risk-taking can have macroeconomic consequences well beyond the reach of the firm as was demonstrated in the financial crisis. There is no conflict between shareholders and no incentive to seek rules that would constrain the moral hazard.

\textsuperscript{160} This assumes the state of incorporation also arbitrates the firm in crisis.
When there are externalities, as are created by excessive risk-taking, this implies that the higher level of government should intervene (Oates, 1971; Bebchuk, 1992). The federal government can take all parties’ interests into account and thus theoretically eliminate these externalities. However, this encompassing interest may also leave the federal government open to greater rent-seeking. By relaxing rules, *ex post* crisis mitigation may increase the likelihood of future crises (Zywicki, 2012b).

In a study of federal systems in developed nations, Rodden (2006) finds that when federal governments that are unable to commit to non-intervention in a crisis, local governments are more likely to take risks that increases the likelihood of a crisis and subsequent bail-out. Rodden demonstrates this problem in several federalist South American countries, and in Germany. Most recently, Eurozone countries experienced a similar dilemma when the macroeconomic effects of a collapse of southern European nations forced northern European nations to intervene, and allowed the former to continue to run deficits with impunity.

If lawmakers cannot commit to not relax contract and property rights this may increase risk-taking by firms. In the most recent financial crisis the federal government bailed-out banks and major corporations which were deemed “too big to fail,” and altered bankruptcy law *ex post* for General Motors and Chrysler. The concept of “too big to fail” has now been incorporated into federal law through the designation of “systemically important” entities, and the FDIC’s power to override bankruptcy embodies a clear intent to bypass property rights in future crises.

States engaging in relaxation of rules are disciplined by horizontal competition. They may change rules that become inefficient or impede necessary innovation but, as demonstrated above, if a state is unable to commit to efficient rules then investors will remove capital to other
states. However, the federal government is not subject to the same discipline and can be more easily captured by special interests.

The claim that federal law should reduce the problems of externalities ignores the problem of commitment. State governments may continue to pay heed only to the interests of the parties to the original contract while the federal government response is inherently more susceptible to politicization due the encompassing nature of its jurisdiction. In other words, while Delaware can ignore the consequences of a failure of General Motors on the economy of Detroit, and continue to provide efficient rules that lead to future investment, the federal government is subject to rent-seeking by lobbies representing unions and the state of Michigan, and is more susceptible to waive rules as it indeed did. The ability of states to disregard macroeconomic and other externalities increases their ability to commit to non-intervention. Thus, the existence of externalities does not imply that the federal government should intervene.

X. Conclusion

The most recent financial crisis renewed a public debate on the respective role of the states and the federal government in determining corporate law. The political debate has been determined in favor replacing market-discipline and the common law at the state level with a political-markets and a code at the federal level. The academic debate remains unresolved with regards to federalism but casts clear doubt on the specific legislative response in Dodd-Frank. The evidence reviewed here demonstrates that much of the new law is likely to be counter-productive. Indeed the Act is not so much a response to the new information provided by crisis so much as an embodiment of existing ideas, frequently driven by rent-seeking, passed under the veil of crisis.
Yet while scholars are able to point to the flaws in the new law, there is no coherent theory of why such failure should be endemic to federal lawmaking but not state lawmaking. This failure stems in part from a misconception of the nature and purpose of federal institutions that is corrected here. The existing literature on horizontal competition focuses entirely on the production of efficient law and ignores the role of commitment.

Horizontal competition as it is presented here is not efficient in the sense of efficient markets hypothesis. Rather it is a process for discovering favorable rules. Crises that disrupt path dependency also provide knowledge about the effectiveness of the rules, just as they provide knowledge about the effectiveness of market institutions. By treating error as a priori evidence of the failure of markets or state law, federal lawmakers interrupt this discovery process, potentially stagnating future innovation and opening the door to greater rent-seeking.

The same literature which treats horizontal competition as efficient ignores commitment as a reason for leaving lawmaking to states. Yet commitment is the very purpose of corporate law and should be the focus of its study. The ability to commit depends upon viable institutions for the protection of property rights and commitment to provide future good law. Horizontal competition alone is inadequate due to the immobility of capital, but it is the existence of horizontal competition that permits such institutions to prevail against rent-seeking and the existence of those institutions that is necessary for success in horizontal competition.

The literature on federalism should take account of the importance of horizontal competition, and the role of the federal government in increasing the ability of states to commit. The existing literature which assumes the control of externalities to be the primary reason for the federal government ignores the federal government’s purpose as an enforcer of competition.
Moreover it ignores the difficulty, generated by the internalization of externalities, which face the federal government when it exceeds this role, to make credible commitments to maintain property rights during a crisis. Viewed from this perspective it is clear that the most recent legislative action is likely to be inefficient and increase the risk of subsequent crises.
References
References


Shadab, H.B., “Hedge Funds and the Financial Crisis,” Mercatus On Policy, Mercatus Center, No.34


5. Conclusion

I. The Science of Choice-Restriction

Economics is primarily a science of choice. Indeed, Becker (Coase, 1988) describes the economics as a discipline distinguished not by subject matter but by approach. This approach is not without its critics, including those from within the discipline. Buchanan (1964a,b), echoed by Williamson (2002), has argued that economics, as a discipline, went in the wrong direction by becoming a science of choice. Instead, he argues, economics should be a science of contract. Simon’s (1972, 1986) critique of the rational choice approach went as far to propose an alternative standard for rationality. Yet despite the individual successes of these economist, they may nevertheless sympathize with Coase (1988), himself a critic of “the preoccupation of economists with the logic of choice,” when he described his own work as “much cited and little used.” Mainstream economics remains a science of choice (Becker, 1996; Sugden, 1991; Vanberg, 2004).

More recent critics of rational choice have focused on the unreality of the assumptions; trivializing the economist’s defense that these assumptions were never intended to be realistic (Freidman, 1953). While behavioral scholars share with earlier critics the belief that rules are necessary to help individuals structure choice, they adopt a constructivist approach to rule-creation and so ignore Vernon Smith’s admonition that the task of the behavioral economist should be to explain the “manner in which institutions serve as social tools that reinforce, even
induce, individual rationality.” That is, they ignore or minimize those institutions that arise through unguided action and evolutionary processes (c.f., Hayek 1973).

In this dissertation I have presented an argument that both design and evolution can provide beneficial rules. These rules restrict choice, and in this sense, are closer to a science of contract, but these rules also exceed mere contract. As Hayek (1937, 1945, 1973) proposes, rules can also embody knowledge that guides individuals to make better decisions. Thus rules serve a dual purpose, suggested by Williamson (1975) of both ensuring commitment and overcoming bounded rationality. However, if rules are to be credited with overcoming bounded rationality then it is necessary to understand not only how institutions may induce individuals to make optimal decisions, but how individuals should select those institutions under the same cognitive limitations they intend to overcome. To answer that question, this dissertation presents an economic science of choice over choice-restriction.

II. Better than Rational

The modern behavioral critiques of rational choice build off Kahneman’s (2003; Kahneman and Tversky, 1973) work into biases and heuristics. These biases can be observed in a laboratory, and are confirmed by a limited body of empirical work, largely focused on financial markets (see e.g. De Bondt and Thaler, 1985, 1987, 1995). Differing from earlier work on bounded rationality that identifies the limits of human cognition, constraint on information, and uncertainty over the future (Simon, 1972, 1986), Kahneman (2003) argues that individuals deviate from rationality in predictable ways when they use automatic processing instead of deductive processing. These critics argue that “choice architects” can improve welfare by structuring
choices to direct individuals to make the decisions those individuals would have made had they used deductive processing (Sunstein and Thaler, 2003).

Critics of rational choice in the behavioral school have sought to replace *homo economicus* with *homo sapiens* but they have left in place the remainder of the economic model. They assume that individuals still exist in an atomistic world and that the neoclassical result is indeed optimal. On this basis Rebonato (2012) argues that the behavioralist critique depends upon neoclassical constructs. The behavioralists assume, in the words of the neoclassical economist John Eatwell, that “if the world is not like the model, so much the worse for the world,” and try to force human beings to behave more like rational man.

By contrast, Simon (1986) argues that heuristics may be a procedurally rational response to limited cognitive capacity, imperfect information, and uncertainty. Later scholars have found demonstrated that many problems are non-computable (Lewis, 1985, 1982): they can only be solved by heuristics. Sen (1977), Goldstein and Gigerenzer (1999), and Ostrom and Gardner (1993, Ostrom et al., 1992) argue that individual rationality may not even be the optimal individual behavior but that instead individuals may hope to be better than rational, by adopting rules that also embody commitments or tacit knowledge.

To produce a model that explains how such rules can be developed in the absence of design, the first essay builds off Cosmides and Tobey’s (1992, 1994, 2005) finding that individuals are equipped with specialized heuristics for social exchange. I demonstrate that this propensity to social exchange can form the basis of a market in beneficial rules. Individuals pick bundles of rules by selecting between groups, creating an evolutionary mechanism by which bounded rational individuals can choose rules that better align individual behavior with optimal
outcomes. I show that these rules are more likely to serve individual preferences than those that are designed and imposed by rational choice architects due to the same limitations of bounded rationality. Moreover, social exchange incorporates a mechanism for innovation and the replacement of rules that become suboptimal.

Yet there are limits to social exchange. Market exchange between strangers is still desirable. Indeed, as the second essay demonstrates, the ability of large numbers of strangers to cooperate in production is an important aspect of growth and innovation. It is in the formal exchange setting of financial markets that critics of rational choice find the greatest evidence for bounded rationality. In the absence of social exchange there is good reason to believe that this is true, yet even in formal exchange, informal routines and practices play an important role in providing knowledge and the means for commitment. The second and third essays, examine these formal relations with the aim of explaining how hierarchical organizations can overcome bounded rationality. I begin by taking a perspective from within the nexus of contracts between shareholders, management, and other stakeholder that comprise the firm and the then from the external perspective of the legal institutions that seek to uphold those contracts.

III. The Firm as Routine

Firms offer a unique perspective into bounded rationality within an organization. Firms can be more easily created and destroyed than more permanent government institutions. Firms also permit more variation in organizational form. Yet one form has dominated: the market-financed joint-stock corporation. Importantly, this form does not rely on centralizing information or decision-making and, while voting maintains a premium, the form relies little on active shareholder monitoring.
As with the first essay, I assume that organizations are limited by the same constraints of bounded rationality as well as that identified by Williamson (1975): limits to communication within the organization. Under bounded rationality, the joint-stock corporation firm with diffuse and disinterested shareholders may be an optimal structure of ownership. Although the extreme example of such a firm, the Berle- Means (1932) corporation, is unrepresentative of the majority of firms even in the United States, this belies the importance of passive and impersonal ownership as a source of finance.

Joint-stock corporations, including the Berle- Means firm, have been criticized as an ownership structure for the conflict the structure creates between constituent groups. Individual shareholders have too small an interest to properly monitor management and workers. Moreover, because of uncertainty and specialist knowledge that cannot easily be transmitted, it is impossible to write complete contingent contracts. Consequently, there is ongoing conflict between managers and shareholders. Furthermore, there is inherent conflict between groups of shareholders and potentially between the firm and its employees and customers. Critics have proposed a number of alternatives, including the debt-financed management ownership, the bank-financed models of Germany and Japan, and stakeholder-ownership models which award voting rights to employees or customers.

Yet while these models have found niches, they have failed to achieve the same level of success as the Anglo-American model of the joint-stock corporation. This can be explained by how that model efficiently stores and processes knowledge, and how the model allows for innovation and growth. Understanding how firms acquire new knowledge is key to understanding the underlying processes which drive growth, and ultimately to understanding why firms are
organized as they are. Any theory of knowledge is, by definition, also a theory of bounded rationality. Both individuals within the firm, and the firm itself, are bounded rational and must develop approaches to managing knowledge.

As Simon (1972) has noted, organizations viewed at the macro level behave as individuals while individuals viewed at the micro level behave as organizations. Just as individuals rely on heuristics, firms rely on a set of established routines to not only cope with repeated tasks but to allow for innovation and learning. In their seminal work on the routine-based firm, Nelson and Winter (1982) propose that optimal routines can selected by an evolutionary process. Unlike the neurological literature described in the first essay, their model does not include a mode of heuristic learning but instead relies on firms beginning with a finite set of rules and eliminating inefficient routines. Thus, no new knowledge is added to the firm and eventually the firm must cease to grow. The second essay demonstrates an alternative view of firm growth which combines Nelson and Winter’s insights with those from the organizational learning literature.

In the routine-based firm, information is not collected and processed hierarchically but is distributed throughout the firm as tacit knowledge. Routines both embody knowledge from experience and reduce conflict. Routines also permit from gradual change as individuals learn better processes for performing activities, allowing for greater efficiency than if processes were determined centrally. Because this knowledge is difficult to access centrally it is also difficult for managers or outside shareholders to effectively monitor behavior. However, by reducing the ability of the firm to deviate from normal behavior, routines also reduce shirking and predation within the firm. Consequently the firm’s organization is path dependent and difficult to alter.
To avoid conflict and maintain efficient routines, shareholders, or any other party with voting rights, should typically take a passive approach regardless of ownership structure and, in the case of shareholders, sell their shares when they disagree with management rather than actively managing the firm. Moreover, shareholders may prefer to limit opportunities for shareholder activism. However, loosely-monitored firms are not immune to malfeasance, or inefficient routines, in the long run. The internal organization of the firm impacts the optimal structure of ownership to ensure growth and provide for innovation.

Alternative ownership structures can be structured to avoid residual claimants from actively managing the firm but do not provide an easy mechanism for individuals with residual claims to exit the firm. If employees or customers are the owners of the firm then exit is more costly while the relations between residual claimants become more complex. This added complexity also makes it more difficult to raise finance, or to alter path dependent routines when these become inefficient. The nature of shareholder ownership rights ensures that residual claims are matched to residual voting rights, and that these may be re-concentrated when active control over the firm is beneficial.

These findings are at odds with current legal trends to increase the power of activist shareholders. The value of voting rights is not that the shareholder should exercise those rights but that those rights remain attached to residual claims and may be recombined by a buyer who wishes to take an encompassing interest in the firm. Moreover, “democratizing” control of the firm has negative implications for innovation and growth. Unless rights can be separated and combined, as in the joint-stock corporation, it will be alternately both too easy and too hard to effect change with implications for rent-seeking and innovation in rules. This is the dilemma
that faces governments who seek to regulate financial markets. Governments cannot combine residual claims to create an encompassing interest and so must rely on alternative mechanisms to achieve this goal, albeit imperfectly. The third essay discusses those mechanisms.

IV. Commitment in Government

Conflict between shareholders and management, and the impossibility of writing complete contingent contracts, provide theoretical justification for the law to intervene in shareholder agreements. Indeed the law does intervene. Yet the law faces the same constraints of bounded rationality and the impossibility of writing complete contingent contracts as do shareholders. Those problems are worsened by the inherent unenforceability of political contract.

Furthermore, by expanding the involved parties to include the entirety of the body politic, interests not only become more diffuse, but groups who not a party to the original contract can either promote inefficient rules or block beneficial change to serve their own ends.

The most recent expansion of corporate governance law, in Title IX of the Dodd-Frank Act, imposes new restrictions on relations between shareholders and voters, many of which expand opportunities for shareholder activism (such as non-binding votes on director pay), or serve specific outside interests (such as disclosing if the firm may have purchased conflict minerals).

Moreover, the law moves authority from state governments to the federal government. This move is likely to be counter-productive when voters are faced with the constraints of bounded rationality they face as consumers and shareholders: limited cognitive ability, limited information, and uncertainty.
Again, rules which embed knowledge and are reduce conflict can overcome the problems of bounded rationality but designed rules face the same problems of bounded rationality they are intended to overcome. Jurisdictional competition does provide an alternative source of discipline in the law but scholars of horizontal competition have focused on competition under similar assumptions as perfect competition in goods markets. Yet these models assume perfect rationality. In the alternative model of jurisdictional competition proposed in the third essay, where rules are selected through evolutionary competition, perfect constructivist rationality is not needed. When individuals are modeled as boundedly rational as both voters and consumers, jurisdictional competition outperforms voting as a means of discovering optimal laws.

Jurisdictions compete not only to provide rules that are currently optimal, but to provide good rules in unpredictable circumstances. Indeed, the primary benefit of law that extends beyond the enforcement of property rights is to provide a framework for resolution of contract disputes in areas that could not be identified \textit{ex ante} and contracted for. Moreover, because jurisdictional competition must impose a commitment on both parties to accept the jurisdiction’s law in future disputes, it is impossible for competition to be effective unless there is a significant cost to relocation.

As with the heuristics discussed in the first essay, and the routines discussed in the second essay, choice-restricting institutions have an optimal plasticity. Insufficiently rigid rules create a calculation problem described by Cosmides (1989) whereby “plasticity is death.” This is compounded by the problem of conflict identified by Williamson (1975), which exists in government in the form of rent-seeking. However, excessively rigid rules lead to inefficiency
under conditions of radical uncertainty. This is resolved by using higher-level rules as a source of local routines. In cognitive science this is the case with heuristics that determine how to solve a novel problem while in corporations it is the case with “double-loop” learning.

In government, constitutions form the meta-rules that underlie decision-making. These meta-rules establish institutions specifically designed to change rules at the least possible disruption to property and contract rights. States, and particularly the state of Delaware, still rely heavily on common law courts which can create new precedent for novel situations that reflects the ex ante intent of the parties to the contract. By contrast, the federal government has shifted toward the use of rules designed by expert regulators which are more likely to be vulnerable to both the calculation problem and the problem of conflict inherent in the political process.

A substantial federalism literature supports the shift of rule-making from the state to the federal level, although not necessarily the shift from common law to regulatory code, to minimize externalities beyond that states under horizontal competition have no incentive to control. However, in the model proposed in the third essay, the key role of federalism is to serve as an additional level of meta-rules that facilitate commitment to providing future optimal rules.

States cannot easily commit to providing securing optimal institutions for providing law in the absence of jurisdictional competition. Only in states such as Delaware, where out-of-state incorporations are a significant portion of the state’s revenue, can the state credibly commit to preserve institutions such as common law courts and constitutional constraints on political interference. Larger states may prefer to commit but are unable. The federal government has no advantage in providing law but has an advantage in enforcing jurisdictional competition.
Conventional models of federalism seek to identify the optimal distribution of responsibilities between states and the federal government on the principle that the latter has encompassing interest but is less responsive. The third essay argues that the federal government is less responsive because its role as an enforcer of commitments demands less flexibility while states control lower level rules and therefore must be more responsive. Under a commitment view of federalism, the federal government’s internalization of macroeconomic externalities may argue against elevating control to a higher level. The ability of states to ignore these externalities combined with the constraints of competition allows states to credibly commit to maintain property and contract rights when such externalities do arise. In turn this reduces the moral hazard to corporations to engage in activities would lead the federal government to enact a bail-out or relax rules.

By reversing the view of federalism, to one based on appropriate assignments to one based on federalism, the logic of controlling ex post externalities is similarly reversed to leave the responsibility in the hands of states who can credible commit to not change rules except to respond to novel circumstances and only where it is beneficial to the initial parties to the contract. The alternative view of federalism and horizontal competition presented in the third essay does not predict that states will always provide efficient law or that they are immune to rent-seeking. Rather it suggests that markets in law will enable the discovery of good rules and of the institutions best suited to ensure future good laws. Horizontal competition is not, however, any panacea. The law, like individuals, will be flawed and constrained by the limitations of bounded rationality.
V. Conclusion

The goal of this dissertation is not only to identify the limitations of rationality, and how choice-restricting institutions may overcome them, but to determine which choice-restricting institutions are most likely to be effective. The first essay looks at how individuals are constrained by bounded rationality in the atomistic world of economic models, before adding in the institutional constraints which serve to align behavior with preferences. It shows that individuals can arrive at such institutions through social exchange as well as through design. While necessarily imperfect, social exchange can be more efficient than design when the constraints of bounded rationality on designers, and those who must monitor them, are incorporated into the model.

Governments are just as constrained by the limits of bounded rationality as are the individuals who comprise the political body. Indeed governments face additional problems when the decision-making group is expanded beyond the initial parties to the contract and there is no residual claimant. To overcome these problems, governments will rely on rules as commitment mechanisms and to embody knowledge to overcome those limitations just as individuals do through heuristics and social exchange, and organizations do though routines. However, the location of rules, whether determined by the federal government or state governments, or by government of private action, will determine how likely those rules are to strike an appropriate balance between delivering the benefits of stored knowledge and providing for commitment, and ensuring sufficient flexibility.

Whereas individuals in social exchange and corporations rely heavily on failure to guide evolutionary competition, governments rely on voting. Yet private organizations in evolutionary
competition have evolved to use voting in a limited capacity, usually leaving routines to manage problems. Governments can use on horizontal competition to allow for a degree of evolutionary selection of rules but in the absence of free entry and exit government cannot achieve the same degree of responsiveness.

Complex problems are better solved at lower levels where experimentation is possible. Governments, moving progressively through local, state, and federal, are more capable of attaining higher levels of commitment through more rigid rules. However, when governments attempt to solve more complex problems they are no longer able to offer the same levels of commitment but still lack the mechanisms to execute beneficial change as effectively as uncoordinated social exchange and commercial activity.
References
References


Curriculum Vitae

Mark Adams received a Bachelor of Arts in Economics from the University of Liverpool in 2004 and a Master of Arts in Economics in 2009.