

John Thrasher

Ordering Anarchy*

Abstract:

Ordered social life requires rules of conduct that help generate and preserve peaceful and cooperative interactions among individuals. The problem is that these social rules impose costs. They prohibit us from doing some things we might see as important and they require us to do other things that we might otherwise not do. The question for the contractarian is whether the costs of these social rules can be rationally justified. I argue that traditional contract theories have tended to underestimate the importance of evaluating the cost of enforcement and compliance in the contract procedure. In addition, the social contract has been understood narrowly as a method of justifying specifically moral or political rules. I defend a broader version of contractarianism as a justificatory model that can be used to evaluate any set of social rules or institutions that impose costs on agents. In so doing, I argue that contractarianism is a general method of evaluating and justifying the rules that order the structure of social life.

Keywords: Contractarianism, David Gauthier, John Rawls, Compliance, Conventions, Folk Theorem, James Buchanan.

“Men want freedom from constraints, while at the same time they recognize the necessity of order. This paradox of being governed becomes more intense as the politicized share in life increases, as the state takes on more power over personal affairs [. . .] ‘Ordered anarchy’ remains the objective, but ‘ordered’ by whom? Neither the state nor the savage is noble, and this reality must be squarely faced.”
(Buchanan 2000[1975], xv–xvii)

* I want to thank audiences at the University of Arizona and the University of Utah for their helpful comments on earlier versions of this paper and specifically for comments from Gerald Gaus, Keith Hankins, Shaun Nichols, David Schmitz, Jonah Schupbach, Danny Shahar, Cynthia Stark, Jacob Stegenga, Chad van Schoelandt, and Steve Wall. I also want to thank Bernd Lahno for his helpful suggestions.

1. Introduction

Contractarianism is typically divided into moral and political contractarianism. The first is concerned with justifying moral principles or rules and the second for authorizing political authority or political principles of justice. Understood in this way, contractarianism is, as Robert Sugden has described it, a “broad church” functioning as a means to justify a broad array of social rules or political institutions (Sugden 2013). As broad as this interpretation of contractarianism is, however, I want to argue that this understanding of contractarianism is still too narrow and misses the real power of the contractarian approach. I will defend a version of contractarianism here that is a general approach to justifying social rules and their related institutional forms.¹

The real problem that contractarianism is uniquely situated to solve is the problem of reconciling individual reasons for action and the constraints of social rules. The basic normative question for individuals is ‘what should I do?’ This question has two basic answers, one that comes from individual reasons and one that comes from social rules. Individuals have reasons for action that come from their own assessment of what will best allow them to achieve their values, goals, or ends whatever those may be. The basis for these reasons need not be self-interested and the general problem of contractarianism is to determine whether a given social rule or set of rules can be justified to individuals with a diverse set of individual values and goals. Individual reasons tend to be teleological and outcome based and have the general form ‘do x in order to achieve outcome y’ where the outcome is preferred because of the value, goal, end that it is related to. Reasons from social rules tend to have the form ‘x is prohibited/allowed/required’ regardless of whether doing the action in question has any relation to one’s values, goals, or ends. The question is whether, when these two sources of reasons conflict, there is some way to reconcile that conflict so as to give effective normative guidance to individual actors.

Contractarianism is an attempt to show that these two sources of reasons for action (individual and social) are reconciled, insofar as they can be, by justifying the social rules on the basis of how they help individuals achieve their individual aims. Put another way, it is an attempt, as Rawls argued, to settle the question of justification “by working out a problem of deliberation” (1999, 16). If individuals, from their own points of view and considering their rational values and aims would or could agree to various social rules or institutions, those institutions or rules would be rationally justifiable. Contractarianism understood this way is not a method of authorizing general political or moral authority but, instead, a way of evaluating the rational justification of social rules generally.

Using this broad understanding of contractarianism, I will argue that the contractarian method of justification should play an essential role in evaluating social institutions, even those that arise from a process of social evolution

¹ Throughout I will use the term ‘institution’ in a very general way as some set of social rules that constrain social action. Here I am loosely following the general definition of institutions used by both Douglass North (1990, 3) and Andrew Schotter (2008).

that do not rely on the state. We can accept that social institutions tend to arise from an evolutionary, path-dependent process of historical development while still imposing a rational test on those evolved institutions. More importantly, a general contractarian test can be used to determine the attractiveness of attempting to move from any given institutional equilibrium to another and providing reasons for doing so. The social contract can be used as method of ‘ordering anarchy’ of linking the rational and the historical. As such, it has a natural affinity with the kind of contract theory that Sugden (2013) advocates and the form of conventional social organization that Anthony De Jasay (1998; 2013) has advocated.

2. The Cost of Rules

All social rules, however beneficial they may ultimately be, have a cost. That cost has two aspects a) the opportunity cost of forgoing alternatives because of the social rules in question and b) the cost of enforcing compliance with those social rules. The first we can call the *opportunity cost* of rules and the second we can call the *enforcement costs*. The opportunity cost of a rule is easy to see. If I am hungry and there is a loaf of bread in a shop window that would alleviate my hunger, it might seem natural to take the loaf of bread. If, however, there is a social rule that one must not steal and, assuming that this loaf of bread belongs to another person or to the shop, the rule will forbid me from taking the easiest means to achieve my ends. This type of cost is often more subtle than this example might suggest, however. Imagine a person deeply devoted to alleviating the suffering of others through giving her time and money to charity organizations. In a society that taxes income for whatever reason, the social rules will tell this person to give a certain amount of her income to the state rather than give that same amount to charity. In this case, the social rule is telling the person to forgo the pursuit of what she sees as an important value—alleviating suffering—in order to follow the social rule.

Another example may help to show the multi-faceted and ubiquitous nature of this cost. In the United States, the Food and Drug Administration has many rules prohibiting individuals from obtaining potentially life saving or suffering alleviating drugs and treatments that have not been approved by the FDA, even for terminal illnesses. To follow these rules is to forgo the benefits that one might be able to receive or to help other receive in order to comply with the social, in this case political, rules. The costs of complying with these rules will in many cases be very onerous indeed. If someone is given the opportunity to save or prolong their life by using a potentially helpful though unapproved drug, what could possibly be their reason for following the rule rather than their own reason? This is the general problem that confronts any social rule that constrains the actions of individuals.

The second type of cost that social rules impose is the cost of enforcement. This is the direct cost of setting up and maintaining a mechanism for imposing sanctions on non-compliers. Formal mechanisms of monitoring and sanctioning require the establishments of institutions and agents that are specifically authorized to engage in sanctioning and monitoring. These institutions, be they public political mechanisms as in the case of police and courts, or private as in the case of internal monitoring within a firm or private security, require funding for their establishment and maintenance. The direct cost in these types of institutions will be in the form of user fees or, in many cases, taxation. The direct cost then will tend to be monetary to those who fall under the aegis of these formal mechanisms of enforcement and monitoring. In the case of informal monitoring and enforcement the cost will tend to be imposed on the members of the institution themselves. For instance, a community garden or park may limit access or restrict use in various ways and rely on the users of that resource to monitor and sanction non-compliance via access or use rules. These costs will be direct if they require the members of a club or group to enforce or monitor the rules directly in the case of informal enforcement and the direct monetary and resource cost of establishing and maintaining formal enforcement and monitoring mechanisms.

Regardless of the type of cost that various social rules impose, either in terms of enforcement or the opportunity cost of following the rules, individuals under a regime of rules will reasonably ask whether those costs are worth paying. In its most general form, this is the question of the justification of those rules. Without directly introducing questions of authority, coercion, or moral foundations then, it is possible to raise a justificatory question with regards to any set of social rules that impose costs. It is this question of whether the costs of the rules are worth paying to the individuals who fall under those rules that leads naturally to a contractarian answer. Contractarian justification is a test of whether a given set of individuals would see the costs of a set of rules as worth paying. As such, it is a general justificatory test that is not limited to a particular moral or political domain.

Contractarianism, at its heart, is a method of justifying social rules be they moral, political, legal, or conventional. These rules require justification insofar as they claim to make authoritative demands on how one should act. If our social rules merely restated our rational or prudential reasons for acting, they would be otiose. Social rules do not, however, merely serve to remind us what action is the best means for achieving our goals, interests, or values. Instead, they often constrain or direct the pursuit of our ends and values in ways that often force us to give up on some of the means or ends that we otherwise would have pursued. There is a potential (and often actual) conflict between the demands of our social rules and what our individual rationality may tell us is the most attractive option.

When we live in society we are confronted with two sources of, at least *prima facie*, authoritative reasons for action: social rules and individual rationality. When these reasons for action conflict—as they often will—we need a rational basis for acting on one set of reasons over another. Put another way, we need

a rational justification for our social rules to reconcile the demands of our social rules with our individual rationality. Without a rational justification for the demands of our social rules, they can only be seen as alien impositions on the rational pursuit of our values and ends. In relation to moral rules, David Gauthier described the potential disconnect between individual rationality and social rules as a “foundational crisis”, arguing that the traditional methods of justifying our social rules are no longer compelling (1991, 15). Our social rules often ask us to give up on the pursuit of the goals and values that we hold most dear. Why should we pay the price that the rules demand without an assurance that the rules themselves are rational, that is, are ultimately necessary for us to live together peacefully under terms of mutual benefit? Contractarianism attempts to reconcile our social rules and rationality directly by showing that rational individuals would agree to (at least some) social rules because those rules are mutually beneficial. We can do better overall by living under certain social rules and institutions than we could do without them.

3. Endorsement and Compliance

There are a number of complications hidden in the above claim that contractarianism aims to justify social rules to those whom they are meant to apply. Surely not everyone can be expected to come to endorse every social rule, even from a purely principled point of view. Communists cannot be expected to endorse many property rules, for instance. At the extreme, the committed anarchist will likely not see any justification for political rules and the committed moral nihilist or psychopath will not see any principled justification for moral rules. Does the failure to justify social rules to the anarchist or the nihilist sink the contractarian project?

The problem is a fundamental one. If social rules must be justified to each individual from their various points of view because they impose costs, a diversity of viewpoints might reduce the potential core of agreement to a null set. This has led many contractarians to restrict the diversity of their contractual agents by specifying narrowly the utility functions or preferences of their agents. Rawls, for instance, reduces the diversity of his contractual agents by arguing that all individuals in the original position are “equally rational” and “similarly situated” behind the veil of ignorance so that “we can view the agreement in the original position from the standpoint of one person” (1999, §24, 120). Gauthier does not go quite so far in reducing the diversity of his contractual agents, both in terms of their strategic rationality and what bargains they are able to accept (1986, 61; 143; Thrasher 2014). John Harsanyi and others introduce a strong symmetry condition into their agreement procedures that restrict the diversity of rational options (Harsanyi 1982, 46–49; 1977, 190). As I have argued elsewhere, reducing the diversity of possible agreements in the contract procedure may be essential to generating a unique answer to what ra-

tional contractors would agree to, but insofar as doing so introduces constraints into the contractual deliberation that are not themselves requirements of rationality, they undermine the entire contractarian approach to justification insofar as that justification relies on showing that the costs of social rules are rationally worth paying (Thrasher 2014).

To rationally evaluate the attractiveness of social rules, individuals must be able to evaluate the relative cost of those rules versus an alternative and this is impossible without taking into account both the expected cost and the expected benefit of those rules. Using a mundane example as an analogy, imagine you are selecting a bottle of wine for dinner. You are having something that pairs well with a strong red and your three choices are an excellent vintage French Bordeaux, a good Australian Shiraz, and a decent California Cabernet. Your ranking, excluding price or budget is:

Bordeaux
Shiraz
Cabernet

Table 1: Wine Ranking (Excluding Price or Budget)

This ranking reflects the overall attractiveness of these options without taking into account their cost, as well as your ability to pay and your willingness to pay. Given these assumptions, the rational choice is to select the Bordeaux. Now imagine we introduce a budget (ability to pay) constraint that rules out the Bordeaux, but leaves the Shiraz and the Cabernet. Now it looks like the rational choice is for the Shiraz (assuming contraction consistency). Imagine further that we now introduce the relative costs of each bottle and your willingness to pay. The Shiraz is more expensive than the Cabernet, say, and if you choose the Shiraz you will have to economize elsewhere, for instance, by eliminating a dessert or glass of port after the meal. This may make the Cabernet the most attractive option once cost has been included into the decision calculus, even though it was the least attractive choice when cost was excluded.

The moral of this little oenophilic example is that evaluating the cost as well as the attractiveness of an option can make a huge difference in which option it is most rational to select. Though the selection of social rules and the selection of wine do not share many features, they share this very important one: cost matters. We need to know the relative trade-offs we will have to make when selecting one wine or one rule over another to make a truly rational choice. Because of this, the evaluation of compliance costs must be incorporated into the contractual choice procedure. This changes, somewhat, both the idea of what the social contract is doing and how it relates to practice.

4. Modeling Justification

The original interpretation of the contractarian project is as a deliberative procedure for showing why rational individuals are justified in endorsing certain social rules or principles. The problem of justifying social rules—of showing that individuals have reason to endorse those rules—is, as Rawls rightly notes, transformed into a problem of deliberation or, more clearly, as a problem of rational choice (1999, §4, 16). This is what Rawls meant when he wrote that the “theory of justice is a part, perhaps the most significant part, of the theory of rational choice (1999, §3, 15)”. It would be more precise to say that the problem of justification becomes a problem of rational choice in contractarianism, an insight that Gauthier developed more fully. We are still only concerned with the justification of endorsement; the justification of compliance is taken to be an additional issue that is worked out differently in Rawls and in Gauthier.²

The problem with this approach is, as I suggested in the last section, that there is good reason to think that rational choice will be impossible, or at least very error prone, if costs are not weighed in with benefits at the outset. This raises the question of the appropriate way of introducing the costs of compliance and enforcement into the choice situation. This is done by thinking of the social contract as a model of justification that takes into account both the costs and benefits of a given set of rules into account. To see how this would work we need to understand two things: 1) what does it mean for the social contract to be a model of justification and 2) how do we incorporate costs into that model? I will look at each question in turn.

As I have already suggested, instead of thinking of the social contract as a form of hypothetical consent or as a promise, we should see it as a model of what reasons we have for endorsing and complying with certain social rules. The goal of contractarianism is not, on this view, to show that each and every person would consent or promise to endorse and comply with a given set of social rules. Instead, it is a normative model for what we could rationally agree to and, hence, what we have reason to endorse and comply with. Given the general problem of justifying the costs of social rules to a potentially diverse set of individuals, it is helpful to think of contractarian justification as a model; specifically a model of the reasons that rational individuals would have for endorsing and complying with particular social rules. Models are indirect representations

² Rawls addresses this problem in the context of the stability of the well-ordered society governed by appropriately chosen principles of justice. In *Theory*, he argues that there will be congruence between the principles of right chosen in the contract and the principles of good embodied in the moral psychology of individuals who live in that society. In *Political Liberalism*, this account changes substantially and he gives up on this congruence argument (see, e.g., Weithman 2010). The case of Gauthier is substantially more complicated and relates to assumptions he makes about both the “translucency” of individuals and the benefits of what he calls “constrained maximization” as well as the possibility of individuals living under a justified contractarian moral system to become socialized to that system. Buchanan’s contractarian theory, though in some ways less developed than the one’s presented in Rawls and Gauthier’s work, does take compliance into account at the outset.

of some complex operation or phenomenon. Because of the complexity of targets, models will often involve some important simplifications. These simplifications or abstractions have a point that is related to the use of the model. For instance, the Volterra predator-prey model, the Price equation, and Hamilton's rule are incredibly simple and abstract models of very complex interactions in biological populations that, nevertheless, have proven to be useful for theorists and scientists (Weisberg 2007). Similarly, maps may involve significant simplifications or distortions and still be useful; indeed it is often because of these distortions that maps are useful. A walking map of the DC mall will be useless if I am attempting to navigate the DC Metro system and a metro map will be useless if I am trying walk from the Jefferson Memorial to National Art Gallery. Even though these two maps represent overlapping parts of the world each has a specific use.

The social contract, I am arguing, should be understood as a model of justificatory reasons. More specifically, we should think of the social contract as modeling what reasons rational individuals have to endorse and comply with certain social rules. The exact form of the model will vary depending on its use. For instance, a social contract model of reasons for endorsing and complying with a particular set of political rules in society A may be very different from a social contract model of coordination norms in society B. This is because the reasons for endorsing and complying with specific social rules, whatever other complexities may be modeled in the social contract, must take into account the cost of compliance. Of course there will be many other reasons for endorsing or complying with a set of social rules but those will vary radically depending on the particular rules in question. Regardless of the specifics of those rules, however, individuals will always consider the cost of those rules from their point of view. Contractarianism is, fundamentally, the project of showing that the cost of social rules is worth paying. Put another way, the social contract is a device for testing whether we are better off with our social rules than without them. Do our social rules aid rather than hinder cooperation? Do they allow us to live together peacefully or are they tools for oppression? Only if these questions can be answered can rational contractors be expected to agree to those rules.

The social contract understood as a model of justification will have three components. First, a model of the diverse set of potential contractors (the members of the group liable to the rule) more specifically a specification of their basis for making rational choices—a model of rationality. Second, it will have a deliberative model of choice, that is, how these contractors come to a rational agreement. In Gauthier's contract this is a bargaining model, in Rawls and Harsanyi it is a model of the idealized choice or one suitably situated chooser. Gaus uses a social choice aggregation model coupled with an evolutionary process (2011). There are many other possible agreement procedures. Third, the contract model will have a justificatory target: what is to be justified. In contractarian theories of justice this tends to be principles of justice, but it may also be moral principles, social rules, or even lower level institutional rules. In any case, the link between the reasons of the contractors and the reasons of actual people must be sufficiently strong for the justificatory output of the contract to have the right

kind of normative force for actual people in real institutions. They must see the reasons of the contractors as reflecting their reasons.

It is not my purpose here to give a complete contract theory so I will leave the question open about how exactly to model the contractors and the agreement procedure. My concern is to show that since the justification of the social rules is related to the cost those rules impose, we must also take into account the cost of complying with and enforcing those rules when the contractors evaluate them. Otherwise, we cannot be sure that the reasons of the contractors in the model will reflect the reasons of actual people outside the model.

As I have argued, endorsement of social rules and compliance with those same rules is conceptually distinct. There may be a reason for an individual to rationally endorse a social rule but there may also be a reason for that same individual to evade complying with that rule in a particular case. Hobbes's fool and Hume's sensible knave are classic cases where an individual might prefer there to be certain rules that everyone else obeys while seeing the benefit of not obeying those rules when there is a benefit to doing so. Gaus (2011, pt. II) has argued that this kind of non-compliance problem is a fatal one to the traditional project of rational choice contractarianism. The problem may be even worse than Gaus, Hobbes, and Hume thought, however. As David Rose has argued, there are often possibilities of opportunism that amount to non-compliance in grey areas where the agent has specialized information that those monitoring compliance will necessarily lack (2011, 34–36). In any case, any account of contractarian justification must be able to do two things: 1) show that rational individuals would have reason to endorse a set of social rules X and 2) show that rational individuals would have reason to comply with the rules in X in particular cases. That is, contractual justification must show that endorsement of the rules and compliance with the rules is worth the cost.

To fuse the rationality of compliance and endorsement in the contract we need to introduce a way to model the cost of the rules into the decision calculus of the agents in the contract. One basic way to do this is to model the rules in question as a bundle of goods, analogous to microeconomic theory. Individuals come to the choice or agreement procedure with a utility function that is a representation of their ranking of those goods as well as as 'budget' for how much they are able to pay for those goods. In the case of social rules this would be how much they are able to pay for the compliance and enforcement of those social rules. The ability to pay is a representation of what they would have to give up in order to select those particular rules and this budget represents the limit of what they are willing to give up for the benefit of the rule. Put another way, any rule that falls above their budget line represents a net cost rather than a net benefit. As such it would be irrational for them to select such a rule. In addition to their 'ability to pay' for a given rule there is also their 'willingness to pay' for a rule. This will typically be higher than their budget line (though it need not be) and will often reflect the relative value they place on the status quo rules versus a proposed alternative. Once we introduce willingness and ability to pay constraints into their basic utility functions the choice situation will likely change

in the same way that in the last section the relative attractiveness of Cabernet changed substantially once a budget and willingness to pay were introduced.

For the evaluation of compliance costs to be modeled properly and to be rationally evaluated, however, we need to understand the relationship between the proposed rules or institutions being evaluated in the contract model and the value attached to them by the contractors. Thinking of the rules as a bundle of goods is a simplification that hides an important complication in this model. Goods are, in economic theory, assumed to be directly consumed or used as intermediate products to achieve some end, that is, they are ranked on the basis of the outcomes they are expected to produce for the individuals involved. This is true of social rules, but in a more complex way. In addition, part of the cost of the social rules as well as their expected benefits depends on the particular enforcement mechanism associated with them.

5. Outcomes, Rules, and Enforcement

The traditional model of rationality used in contractarian theory assumes that individuals choose rules or principles because they see them as the most effective expected prospect to their higher ranked outcome states. In many, if not most cases, choices are made under risk or uncertainty. In cases of risk, individuals know or can roughly assess the probability that their possible actions will lead to particular outcomes. The classic example is the risk of making a particular gamble, say in a game of roulette. For example, the expected value of betting one dollar on red in an American roulette table is $-.053$ given the payoff odds. There are important questions of what it means to *know* the probability of a given prospect in different interpretations of probability, but the idea is that in situations of risk actions are gambles with an expected prospect to a particular outcome or set of outcomes. Under uncertainty the situation is different, it is impossible to accurately know what the probability of a prospect leading to a particular outcome is with very much warrant. According to Frank Knight (1921), uncertainty characterizes many of the economic situations we find ourselves in, especially entrepreneurial situations.

The relevance of this discussion to the social contract is that, as James Buchanan and Geoffrey Brennan (2000[1985], 33–36) have pointed out, the selection of social rules by contractors occurs under conditions of uncertainty rather than risk. This is the case for two reasons. First, social rules typically cover a set of actions that is temporally extended, often indefinitely. One may have a strong sense that he will benefit from rule *X* this month or this year, but he may be less certain that he will benefit after that. Political constitutional rules often work this way. A particular supermajority rule, for instance the filibuster rule in the American senate, may seem beneficial to the party in the minority, but once they are in the majority they will realize the rule can be used against them. As such, the rule must be evaluated for how it will be expected to work in the long-term

and not merely in cases at hand. Second, social rules will often cover a variety of different cases and it will be impossible for individuals to predict with any degree of accuracy the exact effect the rules will have on them. They will only be able to predict the general tendencies of such rules. These two features of social rules that lead to uncertainty in choice are what Buchanan and Brennan call the “permanence” and “generality” of the rules (2000[1985], 34). Of course, permanence is not usually actual permanence, but rather uncertainty about the temporal extension of the rule. The same is true of generality.

This doesn't make rational choice impossible over rules but it does put the contractors under a “veil of uncertainty” (Buchanan and Brennan 2000[1985], 35). Contractors will be uncertain about how the rules will affect them directly, although they will reasonably be able to evaluate the general tendency of rules to benefit or hurt them. As Buchanan and Brennan note, this uncertainty will tend to make agreement more rather than less likely the more general the rules tend to be. Individuals choosers, whatever their particular differences, will be driven to take a general point of view when they evaluate the rules in question.

All of this will be familiar to those acquainted with Buchanan style constitutional political economy. The additional question posed by the introduction of costs into the contractual procedure here is how individuals are meant to evaluate the compliance costs of the rules and the costs of enforcement mechanisms given that there will be significant uncertainty regarding the costs as well as the benefits of various rules.

Given this uncertainty, we can think of the contractual model as having two basic justificatory functions, evaluative and authorizing. The evaluative function acts to show individuals what rules it is rational to endorse and comply with and the authorizing function is to authorize the appropriate enforcement mechanism to stabilize the rules that are chosen. All stable social institutions require some form of enforcement to reduce the incentive of non-compliance. Some conventional institutions can rely on endogenous enforcement, while other institutions must rely on exogenous enforcement to achieve stability. Contractarian theory has tended to focus on the authorization of exogenous enforcement of contracts and social rules, but all enforcement, insofar as it imposes costs on non-compliers, requires authorization and justification. In a society understood as a cooperative venture for mutual advantage the proper justificatory basis for the social rules and their enforcement is mutual benefit. This is the kind of justification that contractarianism provides. The model of contractarian justification aims to show that a set of social rules or institution is a) mutually advantageous to those who are expected to comply with the social rules and b) is a stable institutional equilibrium. All institutions and social rules, whether they rely on exogenous enforcement or not, should meet these two contractarian standards.

The distinction between endogenous and exogenous enforcement is crucial and is key to the authorizing element of the contractarian model. I will discuss the distinction in more detail in the next section. The point here is that whichever form of enforcement mechanism is appropriate, since it imposes costs on those to whom it applies it will require rational justification. This is the cru-

cial role that contractarianism understood as a justificatory model plays in any social system governed by social rules. It is irrelevant whether those rules are 'grown' conventional rules or 'made' rules. In either case, insofar as they impose compliance and enforcement costs, they require rational justification and the only plausible way that they can be justified to a society of diverse individuals with different and often conflicting goals, values, and ends is through a contractarian justificatory process. The choice of an endogenous versus an exogenous enforcement mechanism for the enforcement of the rules we have reason to endorse is a crucial one, but it can only be decided along roughly contractarian lines.

6. Making and Buying Enforcement

The discussion so far has been leading to the crucial question at the heart of the debate between the non-contractarian conventionalist and the contractarian, namely whether there is any role for a contractarian justificatory procedure in a system of conventionally evolved social rules. I have been arguing that there is, that any system of social rules whether it is an evolved set of norms or a made set of rules requires rational justification insofar as it imposes costs on non-compliers and requires enforcement. This is not a moral requirement. We do not owe each other a general justification of our actions. If we did, this would make us, as David Gauthier (2003) has argued, into a society of moral debtors. Instead, the requirement for justification is a rational and practical one. Social rules are not simple, direct outgrowths of our individual goals and values. Acting in accordance with social rules must, to be rationally justified, benefit us on the whole but the rules themselves are not simple rules of thumb or generalizations of individual best practices. Rules constrain our ends and means and compliance with these constraints has a cost. It is this cost that requires justification.

This is, however, to speak in the abstract. In particular cases compliance costs are applied through enforcement, i.e. monitoring and sanctioning. To live under a system of social rules is to be liable to various sanctions for non-compliance. In some cases the monitoring and sanctioning is done internally and personally. In the case of many moral and religious rules, for instance, it is the agent who monitors and sanctions his or herself directly in the form of guilt and regret. When rules have been fully internalized, this is a powerful form of enforcement. Many people are haunted by forms of religious guilt years after they have stopped believing the doctrine to which the guilt was originally attached. As powerful as this enforcement mechanism is, however, it has two main problems. First, it is private and idiosyncratic. Guilt relies on the beliefs and affects associated with those beliefs that have been internalized over time. What rule breaking causes guilt in particular people varies based on their upbringing and the rules system that they have internalized. Especially in a

society composed of people with very different individual belief systems, there will be substantial variance in what kinds of rule breaking triggers guilt. Both the rules and the experience of guilt is typically private and not suitable for enforcing compliance on a broad array of social rules, many of which may not have been deeply internalized. Guilt is different from shame in this way, which is a more public social enforcement mechanism.³ Associated with this is a larger problem that guilt can become unmoored from the rules with which it was initially associated and can become 'free-floating'.⁴ This is true of other emotions like anxiety and love, but is particularly important in the case of guilt. One may feel guilty for something they believed they have done that is wrong or for things that seemingly have no significance. More to the point for our purposes, they may feel extreme guilt at breaking 'taboos' or rules that cannot be justified on reflection.⁵

Second, guilt comes in variable strengths. Some may find the guilt associated with breaking even seemingly trivial rules unbearable, while others may feel very little guilt when breaking very important rules. The judge and jury that dole out the penalty of guilt are fickle and operate on principles that are largely hidden. While a sense of guilt is essential to creating a stable system of rules when monitoring will always be imperfect and enforcement uneven, it cannot do the job on its own. Something more is needed.

The flaws with the personal, private enforcement that guilt generates can and must be supplemented with a public, uniform system of monitoring and sanctioning for non-compliance. This does not, however, create a necessary justification for a political or state based mechanism of enforcement. Contractarianism is a theory of how social rules can be rationally justified on the basis of the mutual advantage of individuals who live under those rules. As I have emphasized though, an important component of whether it is rational to endorse a particular set of rules is the compliance costs associated with those rules. The type of enforcement will matter to the contractors and some forms of enforcement will be more effective and, in the sense I have been discussing it, cheaper than others. Some enforcement mechanisms can be implemented by the members of social institutions, others will require the development of a mechanism that is solely or largely responsible for enforcement. We can distinguish these two types of enforcement, respectively, as endogenous and exogenous enforcement mechanisms.

Endogenous enforcement mechanisms rely on those who live under a particular social rule or institution to monitor and sanction non-compliers directly. Human society arose out of Pleistocene hunter and gatherer bands and in these small societies all of the members of the group do the monitoring and sanctioning (Boehm 1999). There is no king or chief responsible for adjudicating disputes

³ On this difference between shame and guilt and its importance in society see: Dodds (2004, chap. 2); Williams (2008, 63–66, *passim*).

⁴ For a general discussion of 'free-floating' emotions and their significance see Gaus (1990, 59–60).

⁵ Kurt Baier (1954) makes much of the important difference between 'taboos' and justifiable social rules.

or imposing sanctions, the group as a whole is responsible for this task. In more modern societies, Elinor Ostrom, Cristina Bicchieri, and many other recent theorists have shown that many social rules similarly rely on group level monitoring and sanctioning for their stability (see, e.g., Bicchieri and Chavez 2010; Bicchieri and Xiao 2009; Bicchieri 2002; Ostrom and Gardner 2000; Ostrom 1990; Schwab and Ostrom 2008). Ostrom and her collaborators have, in particular, described how informal, non-political rules can arise to manage what she calls “common pool resources”, public goods that are rivalrous but where exclusion is costly (Ostrom 1990). The tendency with these types of resources is towards overconsumption and overuse. She has shown that in many of these cases, informal rules can arise to manage use of the resource. In general, the so-called folk theorem in game theory shows that in many cases of repeated interactions between individuals the threat of the withdrawal of cooperation in cases of non-compliance can be a powerful sanction to stabilize cooperative norms and rules.⁶

As in most things, though, context matters and there are limits to the circumstances under which endogenous enforcement can work. Monitoring costs must be relatively low and information about non-compliance must tend to be reasonably accurate and easy to acquire. Without this, rules stabilized by endogenous enforcement can easily break down (Vanderschraaf 2007; 2010). The question becomes, in these types of cases, what the most effective enforcement mechanism will be. This question can be put, in the language of the theory of the firm, of whether to ‘make’ or ‘buy’ enforcement (see, e.g., Coase 1937; Williamson 1981). Following Coase and Williamson, the answer to this question will hinge on the relative transactions costs of each option. We can think of making enforcement as endogenous enforcement where members of the group do monitoring and enforcement and each member of the group has a roughly equal ability and need to monitor and sanction. Buying enforcement, in this context, is establishing an exogenous agent responsible for monitoring and sanctioning enforcement. The transactions costs are the monitoring and enforcement costs as well as the cost associated with ineffective monitoring and uneven enforcement.

In some cases it will make sense to organize interaction under the aegis of an external enforcement mechanism to reduce monitoring and enforcement costs in the same way that it makes sense in the market to move from direct bilateral contractual exchange to exchange within the firm. The contractarian test of mutual benefit and cost can be used to evaluate the rationality of organizing an institution with either form of enforcement mechanism. The choice, though, will be one of relative costs and benefits to the individuals involved and will be specific to the particular institutional and environmental setting. It is unlikely that there will be one form of monitoring and enforcement that will be appropriate for all institutions in a given society. This point leads us to a change in emphasis in contractarian thought.

⁶ For limits to the application of the folk-theorem as a model for the development of human cooperation see Bowles and Gintis (2011, chap. 5). Kim Sterelny (2012) doubts that modeling the development of cooperation as arising out of Pleistocene bands is appropriate at all, however.

7. Ordering Anarchy

In the previous sections, I have argued that contractarianism is, at its heart, a method of providing rational justification for endorsing and complying with social rules. I argued that traditional contractarian theories put too much emphasis on the justification of the social rules themselves and too little emphasis on thinking about the role that enforcement plays in stabilizing those rules. Specifically, the cost and the form of enforcement are underemphasized. I have argued that the cost of compliance should be included in the initial contractual model and that mechanisms for stabilizing compliance need to be justified in that model as well. Individuals may choose to ‘make’ or ‘buy’ their enforcement based on which method is most effective at securing compliance to rationally justified social rules at the lowest cost to the individuals involved.

The question naturally arises, at this point, of what form a contractarian society would take. I think the answer to this question is largely open. Contractarianism is a justificatory theory based on several premises. One is that social rules need to be justified to individuals since they are the ones who must rationally agree to endorse and comply with the social rules. Another is that rational individuals would only agree to rules that they see as making them better rather than worse off. This leads to an assumption that any set of rules that can be rationally agreed to by a set of individuals will be, as a condition of agreement, mutually beneficial and Pareto superior to any alternative set of rules. The broad vision of a contractarian society is one based on the underlying idea that individuals have their own lives to lead and that, suitably organized, social life will allow them to pursue their collective and individual goals more effectively. Beyond that, though, there is very little the contractarian can say about the specific form a contractarian society would take, without being much more specific about the people that would make up that society.

One implication of this is that a contractarian society need not be either a statist one (relying primarily on exogenous enforcement) or an anarchist one (relying primarily on endogenous enforcement). It may be either or something else entirely. Considering the diversity of individuals in modern societies and differences in problems that social rules are erected to solve, it seems most likely that it will be a mix of both. What will determine the precise nature of this mix of endogenous and exogenous enforcement will partly have to do with the institutional history of the society in question. The contractarian test can be used to evaluate the relative attractiveness of one set of social rules and enforcement over another, but these tests will tend to be comparative and binary. Is institution *X* better (along contractarian lines) than institutions *Y*? Pareto improvements will be made at each node but the process as a whole will be path-dependent and there is no reason to assume that two similar societies, even starting from the same social rules, would converge on the same set of social rules at a time in the future. A contractarian society will be shaped by history as well as reason.

James Buchanan argued that “anarchy is an ideal for ideal men; passionate men must be reasonable (Buchanan 2000[1975], xv)”. Being reasonable, however, does not mean that all order must be imposed from on high. Our social rules shape our engagement with one another. The goal is to create and maintain the rules that allow us to order the anarchy of the human and natural world around us for our benefit. They are tools that we can use to better or circumstances or to worsen them. The idea of the social contract is a framework for ordering anarchy for mutual benefit and peace. It takes humans as they are and helps them evaluate institutions as they might be. As such it is a general tool for evaluating all forms of social rules, however they are enforced.

References

- Baier, K. (1954), “The Point of View of Morality”, *Australasian Journal of Philosophy* 32(2), 104–35.
- Bicchieri, C. (2002), “Covenants without Swords: Group Identity, Norms and Communication in Social Dilemmas”, *Rationality and Society* 14(2), 192–228.
- Bicchieri, C. and A. Chavez (2010), “Behaving as Expected: Public Information and Fairness Norms”, *Journal of Behavioral Decision Making* 23(2), 161–78.
- and E. Xiao (2009), “Do the Right Thing: But Only If Others Do So”, *Journal of Behavioral Decision Making* 22(2), 191–208.
- Boehm, C. (1999), *Hierarchy in the Forest: The Evolution of Egalitarian Behavior*, Cambridge: Harvard University Press.
- Bowles, S. and H. Gintis (2011), *A Cooperative Species: Human Reciprocity and Its Evolution*, Princeton: Princeton University Press.
- Buchanan, J. (2000[1975]), *The Limits of Liberty: Between Anarchy and Leviathan*, Indianapolis: Liberty Fund.
- and G. Brennan (2000[1985]), *The Reason of Rules*, Indianapolis: Liberty Fund.
- Coase, R. (1937), “The Nature of the Firm”, *Economica* 4(16), 386–405.
- Dodds, E. R. (2004), *The Greeks and the Irrational*, Berkeley: University of California Press.
- Gaus, G. (1990), *Value and Justification: The Foundations of Liberal Theory*, Cambridge: Cambridge University Press.
- (2011), *The Order of Public Reason: A Theory of Freedom and Morality in a Diverse and Bounded World*, Cambridge: Cambridge University Press.
- Gauthier, D. (1986), *Morals by Agreement*, Oxford: Clarendon Press.
- (1991), “Why Contractarianism?”, in: Vallentyne, P. (ed.), *Contractarianism and Rational Choice: Essays on David Gauthier's Morals By Agreement*, Cambridge: Cambridge University Press, 15–30.
- (2003), “Are We Moral Debtors?”, *Philosophy and Phenomenological Research* 66(1), 162–68.
- Harsanyi, J. (1977) *Rational Behavior and Bargaining Equilibrium in Games and Social Situations*, Cambridge: Cambridge University Press.

- (1982), “Morality and the Theory of Rational Behavior”, in: Sen, A. and B. Williams (eds.), *Utilitarianism and Beyond*, Cambridge: Cambridge University Press, 39–62.
- De Jasay, A. (1998), *The State*, Indianapolis: Liberty Fund.
- (2013), “Conduct and Contract”, *Rationality, Morals and Markets* 4 (Special Topic: Can the Social Contract Be Signed by an Invisible Hand? G. Brennan and B. Lahno eds.), 53–60.
- Knight, F. (1921), *Risk, Uncertainty, and Profit*, Boston: Houghton Mifflin.
- North, D. (1990), *Institutions, Institutional Change and Economic Performance*, Cambridge: Cambridge University Press.
- Ostrom, E. (1990), *Governing the Commons: The Evolution of Institutions for Collective Action*, Cambridge: Cambridge University Press.
- and R. Gardner (2000), “Coping with Asymmetries in the Commons: Self-Governing Irrigation Systems Can Work”, in: McGinnis, M. (ed.), *Polycentric Games and Institutions: Readings from the Workshop in Political Theory and Policy Analysis*, Ann Arbor: University of Michigan Press, 399–420.
- Rawls, J. (1999), *A Theory of Justice*, Cambridge: Belknap Press.
- Rose, D. (2011), *The Moral Foundation of Economic Behavior*, New York: Oxford University Press.
- Schotter, A. (2008), *The Economic Theory of Social Institutions*, Cambridge: Cambridge University Press.
- Schwab, D. and E. Ostrom (2008), “The Vital Role of Norms and Rules in Maintaining Open Public and Private Economies”, in: Zak, P. (ed.), *Moral Markets: The Critical Role of Values in the Economy*, Princeton: Princeton University Press, 204–27.
- Sterelny, K. (2012), “From Fitness to Utility”, in: Binmore, K. and S. Okasha (eds.), *Evolution and Rationality*, Cambridge: Cambridge University Press, 246–73.
- Sugden, R. (2013), “Contractarianism as a Broad Church”, *Rationality, Morals and Markets*, 4 (Special Topic: Can the Social Contract Be Signed by an Invisible Hand? G. Brennan and B. Lahno eds.), 61–66.
- Thrasher, J. (2014), “Uniqueness and Symmetry in Bargaining Theories of Justice”, *Philosophical Studies* 167(3), 683–99.
- Vanderschraaf, P. (2007), “Covenants and Reputations”, *Synthese* 157(2), 167–95.
- (2010), “The Invisible Foole”, *Philosophical Studies* 147(1), 37–58.
- Weisberg, M. (2007), “Who Is a Modeler?”, *British Journal for the Philosophy of Science* 58(2), 207–33.
- Weithman, P. (2010), *Why Political Liberalism?: On John Rawls’s Political Turn*, Oxford: Oxford University Press.
- Williams, B. (2008), *Shame and Necessity*, Berkeley: University of California Press.
- Williamson, O. (1981), “The Economics of Organization: The Transaction Cost Approach”, *American Journal of Sociology* 87(3), 548–77.