

It Can't Be Rational Choice All the Way Down: Comprehensive Hobbesianism and the Origins of the Moral Order*

Gerald Gaus

1 OVERCOMING THE FATAL LACUNA IN "HOBBSIAN" POLITICAL THEORY

James Buchanan grounded his constitutional political economy project on what is commonly understood as Hobbesian contractarianism.¹ On this familiar view, the Hobbesian models individual agents as essentially self-interested, strategic, maximizers: the aim is to analyze the political order that emerges from the choices and actions of these Hobbesian agents. Because the aim is to demonstrate how certain legal and political orders can arise out of the interactions of self-interested agents, the assumed baseline condition is a normless interaction in a "state of nature." The model then shows that these agents would be caught in mixed motive games such as Prisoner's Dilemmas, where noncooperation is the dominant strategy, or at best Stag Hunt/Assurance games, with a strong tendency for players to play the risk-dominant equilibrium.² In either case, an agreement to end the war and accept a regime of rights benefits each agent, and so a large set of possible "social contracts" are on the Paretian frontier. Buchanan (1975: 1-35; 2001, 16: 15-27), of course, argues that agents will not accept the orthodox "Hobbesian" contract instituting a Leviathan with unlimited authority, but, rather, a constitutional regime. Specifying the rules — especially the basic procedural rules — of that regime is critical to the constitutional political economy project.

This familiar Hobbesian project is of special interest to economists who turn their attention to political philosophy, as it is common to equate Hobbesian self-interested agents with the basic model agents of rational choice theory.³ Insofar as economists are interested in the application of rational choice models, "Hobbesian contractarianism" is thus tempting, as it certainly was for Buchanan's political economy project. It would be a satisfying and elegant analysis of a sophisticated rule-based political order that it can

* A version of this essay was delivered to the symposium on "Tensions in the Political Economy Project of James M. Buchanan," organized by the F. A. Hayek Program for the Advanced Study of Philosophy, Politics, and Economics at the Mercatus Center at George Mason University. My thanks to the participants and an anonymous reviewer for valuable comments and suggestions.

emerge from the choices of rational, self-interested, individuals. And insofar as one is, like Buchanan (Buchanan and Tullock, 1965: 307-22), committed to a normative individualism that accords foundational value to respect for individual choices, understanding what political institutions would emerge from these choices helps us get insight into long-standing normative questions in political theory (Thrasher and Gaus, forthcoming).

Straightforward, basic, Hobbesian contractarianism, though, is undermined by what we might call its “fatal lacuna.” It supposes simply two relevant social states: Hobbesian agents acting (i) in a (literally) “unruly” condition (ii) in a rule-based political order, where (i) helps us see why we end up in particular type of (ii). This, though, supposes that there is no social and moral order that is “ruly” before the state. But there is: outside of, and in an important sense prior to, political and legal order are moral orders, including the general moral order of large-scale civil society. Overlooking the informal moral order underlying the legal-constitutional order, T. H. Green insisted, leads to the false conclusion that the state, not the preexisting moral order, is the fountainhead of individual rights (Green, 1889: 89). Moral orders are not simply the reflection or creation of the political, but independent sources of moral rights and claims. Now if individuals possess moral rights independently of the state, then the set of Paretian political institutions (“social contracts”) cannot be defined by the unruly baseline of the state of nature, but must refer to the more ruly baseline of the informal moral order. Moreover, the informal moral order does not simply constrain political and legal options, it creates them. As Montesquieu (1748: 292-315) stressed, the effectiveness of any constitution depends on the habits, manners and morals of the population. Some societies’ informal morals, Montesquieu held, simply do not support free constitutions; others do. The fatal lacuna of the straightforward Hobbesian social contract is that it is blind to this critical role of the underlying moral order, depicting effective social order as simply the artefact of a politico-legal regime.

Section 2 defends the core of the fatal lacuna thesis — Montesquieu’s claim that an effective constitutional order depends on an informal social and moral order. Given this, a plausible Hobbesian political theory cannot jump from the unruly condition to a lawful

ruly one, since the success of the latter depends on informal moral rules, and getting them right. However, in section 3 we see that, despite what may be the popular reading, Buchanan's Hobbesian analysis does not manifest the fatal lacuna: he proposes what we might call "Comprehensive Hobbesianism." The legal, political, *and* moral orders all can be justified by Hobbesian analyses: as we go deeper and deeper into the basis of order, we encounter deeper and deeper Hobbesian arguments. It is, as it were, Hobbesian rational choice all the way down. Section 4 sketches Buchanan's Hobbesian model of the origins of the moral order. I argue that it pretty clearly fails as a rational reconstruction of the origin of the large-scale moral order needed to support a large-scale political and economic order, though it is an insightful model of stabilization of a preexisting informal order. As is so often the case with innovative thinkers, even they do not achieve their ultimate goals (in this case, reconstructing norm origins) they enlighten us about the overall phenomenon (norm stabilization). Section 5 — again taking a cue from Buchanan — generalizes this problem to Comprehensive Hobbesian analyses of moral order in general. A serious tension in his political economy project is revealed: Buchanan's analysis justifies large-scale politico-legal orders, but is unable to show how these large-scale moral orders can come about. I conclude in section 6 by evaluating the upshot of my analysis for the overall project of constitutional political economy.

2 THE IMPORTANCE OF THE MORAL ORDER

2.1 *Moral Norms*

Normative guidance — motivation to comply with social rules despite the fact that cheating would be to our material advantage — goes far back in human evolutionary history, and is probably one of the defining features of our species (Kitcher, 2011; Boehm, 2012; Gaus, 2015). This is captured by Bicchieri's (2006: 52-5; 2016, chap. 1) formalization of a social norm, in which it is characterized as an individualized sensitivity (see §4.4 below) to a shared social rule (one that is actually followed in one's group or culture) that leads one to forgo material gains.⁴ Critical to her analysis is that norm followers are moved both by their empirical expectations as to what others will actually do (will others follow the rule?) and their beliefs about what others in their network normatively expect

of them (will others hold me responsible for failing to conform?). In an impressive series of experiments she has measured these expectations, and shown that they are good predictors of behavior in ultimatum and dictator games (e.g., Bicchieri and Chavez, 2009; Bicchieri and Xiao, 2009).⁵ She has subsequently (2016) shown that her model allows measurement of norms in field settings in both small and large-scale societies.

It should be stressed that on Bicchieri's account, compliance with social norms cannot be reduced to either (i) personal approval of the behavior required by the norm or (ii) a person's desire to escape punishment. As regards (i), she has provided compelling evidence that attitude change often fails to translate into action change — partly because a person is moved not simply by his own attitudes but, critically, by his beliefs about what others with whom he interacts normatively expect from him. Bicchieri (2016: chap. 1) draws our attention to a UNICEF study on violence toward children, which reported both high rates of caregiver disapproval of punishment *and* of caregiver punishment. Similar findings have been reported concerning prison guards (Bicchieri, 2006: 180). And, as she reports, in some African countries there appears to be a similar pattern concerning female genital cutting — high disapproval rates combined with high participation rates (Bicchieri, 2016: chap. 1). Concerning (ii), while of course informal social rules are typically enforced by punishment (Gaus, 2011: 103-122), we also know that when punishment fails to correspond to what people believe are legitimate normative expectations, punishment easily turns into “anti-social” counter-punishment. As Samuel Bowles and Herbert Gintis (2011: 26) stress, effective punishment depends on beliefs about its legitimacy: unless those to be punished and their friends and allies are convinced that the rule being enforced is legitimate, a punishing action taken as a means to protect social cooperation can lead to weakening it. Experimental evidence (e.g., Hopfensitz and Reuben, 2009) confirms that attempts to punish readily evoke counter-punishment when the offender does not experience guilt, which is associated with moral norm violation. Thus the internalization of moral norms is critical: in most cases the acceptance of a moral norm motivates individual compliance.⁶

Most economists recognize that the spontaneous order of the market presupposes legal institutions (Buchanan, 2001, 16: 27). Far fewer follow Hayek (1988: 6): “To

understand our civilisation one must appreciate that the extended order resulted not from human design or intention but spontaneously: it arose from the unintentional conforming to certain traditional and largely *moral* practices, many of which men tend to dislike, whose significance they usually fail to understand, whose validity they cannot prove, and which have nonetheless fairly rapidly spread by means of an evolutionary selection....” However, in the last twenty years there has been a growing consensus that market orders rest on the evolved informal moral order, and in particular of moral norms and rules (Schwab and Ostrom, 2008; Friedman, 2008; Richerson and Boyd, 2008; Rose, 2011; Platteau, 2000: 325ff; Gaus, 2015).

2.2 *Law and Moral Norms*

There is, I think, a tendency to suppose that while informal social rules are important to small-scale societies, in large-scale societies they are supplanted by formal legal institutions. And this common view captures an important truth: in large-scale economic orders formal law sometimes competes with, and supplants, smaller informal networks with overlapping competencies, say, about contractual performance. In Taiwan, for example, numerous small norm networks provided the basis for contractual enforcement in small- and medium-sized firms, thwarting the development of western formal legal contractual relationships (Platteau, 2000: 285-7). As formal legal institutions expand their effective jurisdictions, they tend to supplant these informal competing networks. This should not be denied.

Nevertheless, this common view misses the fundamental role of moral norms in upholding legal systems and rendering them effective (Platteau, 2000: 290ff.). We can distinguish two ways in which the informal morality is critical to the effectiveness of law: (i) at the level of a specific law, *L*, its effectiveness often depends on its relation to specific informal moral norm, *N* and (ii) at the general level of legality as such, whether there is a norm upholding obedience to law (Tyler, 1990).

Re (i): as William J. Stuntz (2000: 1871) points out, “The mass of the population avoids seriously bad behavior not because they know it can be found in the codes, but because they know the behavior is thought to be seriously bad (and only secondarily because seriously bad behavior can often get you thrown in jail). For the most part,

criminal law regulates actors in the legal system, while popular norms — morals — regulate the conduct of the citizenry.” Platteau (2000: 290) concludes: “[c]learly, the fact that laws provide external validation of underlying social norms appears to be an important factor of effective implementation of the former.” Although, as Montesquieu (1748: 304-5) stressed, there are mutual influences between moral norms and legal regulation, it is generally more accurate to see the informal norms shaping the law than *vice versa* (Robinson, 2000). Laws that run counter to the moral norms of the populace are not only apt to be ineffective but, according to Stuntz, self-defeating: laws against normatively-approved behavior often strengthen, not weaken, the norm in the population. Thus laws that depart from the basic moral norms of a society mostly likely will be ignored, often engendering contempt for the law. Gerry Mackie (2017) has recently noted that there are hundreds of critical cases around the world in which practices — among them female genital cutting, caste discrimination, child marriage — have been widely criminalized yet continue to be practiced, as they are supported by local informal norms. Mackie, following Iris Marion Young (2011) concludes:

Criminalization is an appropriate response to a criminal injustice, a deviation from accepted norms, its harmful consequences intended, knowingly committed by identifiable individuals, whose wrongdoing should be punished. It is not an appropriate response to a structural injustice, in compliance with accepted norms, its harmful consequences unintended byproducts, and caused by everyone and no one. The proper remedy for a harmful social norm is organized social change, not fault, blame, punishment.

“[W]hen manners and customs are to be changed,” Montesquieu (1748: 289) observed, “it ought not to be done by laws; this would have too much the air of tyranny: it would be better to change them by introducing other manners and other customs.”

Re (ii): Tyler’s panel study of compliance in Chicago (1990: 59) showed that a person’s moral convictions about fairness and justice are important determinants of whether she will comply with laws that conform or conflict with them. But Tyler also found that whether citizens’ accepted the legitimacy of politico-legal procedures has a strong impact on whether people comply with the law. In this vein Mackie (2017) argues that the data indicates that an effective legal system requires a generalized norm supporting legal obedience. Recent political philosophy has tended to be dismissive of a general obedience to the law, or a general recognition of its moral authority, but without

some such support in the informal moral order, legal regulation is apt to be ineffective (Gaus, 2011: 449ff).

2.3 *The Fatal Lacuna*

We are now in a position to appreciate the fatal lacuna in the standard Hobbesian social contract. The typical analysis begins with essentially self-interested individuals (call them “egoists”) in a normless environment, and then asks to what formal institutions they would agree in order to improve their deeply suboptimal relations. In its most plausible versions, the answer to this question should, firstly, help us evaluate our current institutions to see if they satisfy the test of mutual benefit and, secondly, point the way to reforming them. However, if the individuals already endorse and live by an informal moral order (call them norm-followers) (i) their problems will not be the same problems as those of egoists; (ii) the solutions viable for egoists and norm-followers will be different; and (iii) norm-followers will evaluate proposals differently, as the proposals must correspond to their already-endorsed moral norms. This is not to say the norm-followers will be entirely uninterested in the Hobbesian conclusions: in some cases they may approve of institutions of mutual benefit, and use the Hobbesian analysis as a guide to reforms. But often norm-followers will dismiss the problems of egoists and, importantly, will conclude the institutions fit for egoists would ill-suit them. Most obviously, egoists will stress an instrumentalist approach to compliance, seeking to show how legal punishment regimes can channel self-interested individuals into compliance, and so may erect extensive enforcement procedures that would be counterproductive for norm-followers. At the same time the Hobbesian theorist is apt to be over-optimistic about the range of institutional schemes that could produce effective social cooperation, as he will not take the prior moral norms as constraints on solutions (and on the viability of legal regulations). The Hobbesian thus will overlook both resources (regarding compliance) and constraints (about the fit with preexisting norms), rendering his proposals of limited value.

3 BUCHANAN'S RECOGNITION OF THE MORAL ORDER

Most readers of Buchanan's better known works, I would venture, would see the previous analysis as applying to his Hobbesian project. "The Hobbesian jungle," he wrote in 1977, "is something to be avoided and something that people with rational self-interest will seek to avoid through general agreement on law along with the requisite enforcement institutions.... The contractarian seeks '*ordered anarchy*'; that is, a situation described as one which offers maximal freedom for individuals within a minimal set of *formalized rules* and constraints on behavior" (2001, 16: 26, emphasis added). And indeed, he sometimes identifies "rules" with "law" (2001, 16: 23), suggesting that all order is formal order.

That, though, is not his considered view. "Life in society, as we know it, would probably be intolerable if formal rules should be required for each and every area where interpersonal conflict might arise" (Buchanan, 1975: 118). "A society," he tells us, "is held together by some combination of moral community and moral order. Its cohesion is reduced by the extent to which moral anarchy exists among its members" (2001, 17: 188). A moral community then, "exists among a set of persons to the extent that individual members of the group identify with the collective unit." In contrast:

A *moral order* exists when participants in social institutions treat each other as moral reciprocals, but do so without any sense of shared loyalties to a group or community. Each person treats other persons with moral indifference, but at the same time respects their equal freedoms with his own. Mutual respect, which is an alternative way of stating the relationship here, does not require moral community in any sense of personal identification with a collective or community.... In a moral order, it is possible for a person to deal with other persons who are not members of his community if both persons have agreed, explicitly or implicitly, to abide by the behavior precepts required for reciprocal trust and confidence.

The emergence of the abstract rules of behavior describing moral order had the effect of expanding dramatically the range of possible interpersonal dealings. Once rules embodying reciprocal trust came to be established, it was no longer necessary that both parties to a contract identify themselves with the same moral community of shared values and loyalties (2001, 17: 189).

True moral anarchy is thus the absence of moral community *and* moral order. Individuals treat each other simply strategically: "each person treats other persons exclusively as a means to further his own ends and advantage" (2001, 17: 190). With these richer concepts in play, Buchanan now is able to offer a moralized understanding

of the idea of “ordered anarchy” as “universal adherence to the rules of mutual respect among persons” — i.e., moral order (2001, 17: 191).

Buchanan manifestly views the moral order as necessary to a functioning and free politico-legal order. “For several decades, however, our moral order has been in the process of erosion. Larger and larger numbers of persons seem to have become moral anarchists; they seem to be losing mutual respect for one another along with any feeling of obligation to abide by generalizable codes of conduct” (2001, 17: 197). His “moral order” (an idea equivalent to Hayek’s “Great Society” (2001, 17: 441-2)) is a far-flung moral network extending far beyond particularistic, moral communities:

Indeed, I should argue that it is through the development of the rules of such a “moral order” that humans have successfully confronted the challenge of civilization. The moral order that allows humankind to supersede the effective limits of its expressly revealed moral community is best described as the order of *law*, in which persons abide by abstract rules of law that lay down rights of separate persons and that provide bases for mutual respect, tolerance and, most importantly, for trade and exchange that, in turn, greatly enhances the level of personal well-being for all participants (2001, 17: 208).

4 BUCHANAN’S PREACHING MODEL OF THE ORIGINS OF NORMS

4.1 Internalizing Ethical Externalities

One might be tempted to suppose that, given his recognition of the fundamental role of the moral order in supporting an extensive politico-legal order, Buchanan might, after all, not view economic rationality (qua effective promotion of self-interest) as foundational to modeling social life in general, but only to justifying the political and economic orders. While perhaps the analysis of constitutional rules is grounded on self-interested maximizers, not so the very foundation — the moral order. This, though, would be an error. Although Buchanan endorses the foundational role of the moral order, he nevertheless seeks to give an economic analysis of the origins of the moral order itself. For Buchanan it is Hobbesian rational choice analyses all the way down.⁷

As with many analyses of social norms and moral rules (Bicchieri, 2006: 26-27; Gaus, 2011: 53ff), Buchanan begins with the (all-too) familiar Prisoner’s Dilemma, in Figure 1.

		Betty	
		C_b	D_b
Alf	C_a	3	4
	D_a	1	2*

FIGURE 1: ECONOMICALLY- (AND ETHICALLY-) RELEVANT EXTERNALITIES (ORDINAL UTILITY, 4 = BEST)

Of course, as Buchanan points out (and which, on many views Hobbes did in the sixteenth century), although $\{D_a, D_b\}$ is the sole equilibrium, Alf and Betty could benefit from a contract that would enforce $\{C_a, C_b\}$. Buchanan observes that when either Alf or Betty choose D they are imposing externalities on each other. “The existence of *externality* is necessary for the initiation of any effort aimed at constraining behavior” (2001, 17: 216). In this case the economist can point out that a $\{C_a, C_b\}$ contract, despite the constraints it entails, is a Pareto improvement over the $\{D_a, D_b\}$ noncooperative equilibrium, so this is an externality that is economically relevant, and with which economics readily deals. Now consider Figure 2.⁸

		Betty	
		X_b	Y_b
Alf	X_a	2	3*
	Y_a	1	4

FIGURE 2: (POTENTIALLY) ETHICALLY-RELEVANT EXTERNALITIES

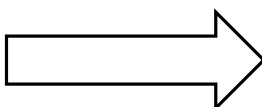
The sole Nash equilibrium in this game is $\{X_a, Y_b\}$. Once the equilibrium has been achieved, no Pareto improvements are possible. But still, when Alf plays X_a he imposes an externality on Betty: if he had played Y_a Betty would have been better off.⁹ A shift on the part of Alf from X_a to Y_a will have significant “spillover benefits on B[etty]” (2001, 17: 219). This externality, Buchanan argues, is economically irrelevant (since there are no mutual gains to be had) but is potentially ethically relevant. Betty may hold that Alf is defecting from the best rule of “everyone play Y ,” and the costs he imposes on her by defecting to X -ing are ethically important. There may be an ethical case for, say, $\{Y_a, Y_b\}$,

which eliminates the externalities that Betty imposes on Alf. Thus while for economics the externalities modeled in Figures 1 and 2 are categorically different, this is not so for ethics.

Suppose, then, Betty seeks to eliminate the externality that Alf imposes on her in Figure 2: she seeks to get him to also play γ . Of course, given the payoffs in Figure 2 this would be irrational for Alf: γ_a is not his best response to her γ_b . As with many accounts of norms, Betty must induce a change in Alf's payoffs, so that he is playing a different game, as in Figure 3.

		Betty	
		X_b	Y_b
Alf	X_a	4 2	3* 3*
	Y_a	3 1	2 4

Game 1



		Betty	
		X_b	Y_b
Alf	X_a	4 2	3 3
	Y_a	3 1	4* 4*

Game 2

FIGURE 3: BETTY MORALIZES ALF, AND EACH (SORT OF) GAINS IN THE PROCESS

In this transformation, Betty induces a preference change in Alf that eliminates the externality he imposed on her; in Game 2 he comes to see γ -ing when she γ s as the best thing to do. As Buchanan notes, we cannot say that the move from Game 1 to Game 2 is Paretian: since the preferences are different, they are strictly noncomparable games. Yet it is also true that in Game 2 both parties have an equilibrium outcome $\{\gamma_a, \gamma_b\}$ that each prefers to the $\{X_a, Y_b\}$ outcome that was in equilibrium in Game 1. So, at least after the fact, both parties may judge themselves to be in a preferred state (2001, 17: 225).

On Buchanan's model, then, Betty has an incentive to invest in changing Alf's preferences; we can think of this as her efforts to induce Alf to internalize the "act on γ " rule. She has an incentive to preach to Alf that γ is the thing to do.

To be productive, investments in the promulgation of moral norms must change the behavior though a shift in the utility function. ... The analysis here does not require that persons seek to change *their own* preferences. While it may be empirically descriptive to say, with Frank Knight, that persons really "want better wants," we do not need to take this additional step here. We need no such bootstrap ethics. My model suggests, much more restrictively, that persons rationally will "want others to have better wants," or, specifically, that others will behave more cooperatively toward them in social discourse (2001, 17: 223).

4.2 More of a Convention than a Norm

Note that on this analysis for Alf to internalize the “ γ norm” does not imply that he accepts that he ought to γ even though he would benefit by defecting, or that he is committed to following the γ norm despite his self-interested x -inclining preferences. “Individuals behave in accordance with norms of cooperation in social interaction because of the dictates of their preference orderings. They ‘do not want to steal,’ even when opportunities exist” (2001, 17: 230). Such a view of internalization reflects the thought expressed by Lincoln: “When I do good, I feel good. When I do bad, I feel bad” (Bowles and Gintis, 2011: 169). Yet, though this certainly picks up in a feature of moralization, if that is the entire story we have something closer to a convention than a moral norm.¹⁰ Given the preference change in Game 2, coordination on γ is secured through entirely “self-serving” action; each has no other concern than securing his or her best outcome, given the choices of the other. Once empirical expectations about the γ rule are secured, the underlying incentive to comply is the same as a convention: the “ γ rule” would be entirely self-enforcing (without threats of punishment), since neither Alf nor Betty would have any incentive to defect — they are each doing what they most want to do.

What is missing from Buchanan’s account of moral norms is the sense of “internal ought” that we find in moral philosophers such as Hobbes, Hume, and Hutcheson (Darwall, 1995). A capacity for normative guidance is not simply a disposition to form a pro-social cooperative attitude such as “I don’t like to harm people,” but a rational preference to follow rules even when one sees that such following does not yield what, on the basis of one’s core interests and goals, one wants (§2.1). Interestingly, Kliemt (1990) — in a way similar to social theorists such as Sen (1982) and Baier (1995: chap 5) — extends Buchanan’s framework so that moralized individuals possess both self-interested and “moral” sub-orderings, such that the power of an individual’s normative guidance can be modeled in terms of the meta-ordering (or trade-off rate) between the two sub-orderings. The moral individual will defer to the moralized preference ordering over the self-interested one. This approach provides a more promising way of modeling guilt: when an individual gives into temptation and acts on her self-interested, rather

than moralized, preferences, she experiences guilt. On the basic model introduced by Buchanan, there is no reason why Alf should feel guilty if he fails to play γ in Game 2 — he simply failed to maximize his own preference satisfaction.

4.3 Preaching and Attitude Change

A deeper worry about Buchanan's model of the origin of moral norms stems from the proposed mechanism of moralization: preaching. It seems a dubious origin for norms. Studies indicate that "preaching" or "exhorting" others to, say, donate to charity, has relatively little effect on behavior. On the other hand, when others observe people *being* charitable, they are much more likely to engage in charitable acts (Henrich and Henrich, 2006: 29-30). Subjects who experience inconsistency between what a person preaches (what she states are her normative expectations), and what she actually does, tend to have a harder time recalling the content of the exhortation; those who witness conformity of preaching and action better recall the content of the preaching message. This suggests that statements of normative expectations that do not mesh with observed behavior are washed out, while the congruence of normative expectations with action increases the subjects' awareness of the normative expectations. So, as in our example, it is critical that Betty acts on the γ norm before she preaches to Alf.¹¹ Normative behavior must precede effective preaching. In an important experiment, Bicchieri and Erte Xiao (2009) studied the role of empirical expectations — what people *do*, rather than what they say *should* be done. They found that, indeed, expectations about what others will do are crucial for norm following. In dictator games, normative ideas about what others think *ought to be done* do not have a significant impact when they run counter to empirical expectations about *the expected behavior of others*. In their experimental work on public goods games among two small-scale societies (the Machiguenga and the Mapuche), Joseph Henrich and Natalie Smith (2004: 13) also found that "the primary indicator of what a subject will do is what the subject thinks the rest of the group will do." There thus seems good evidence to support the claim that people often act on social and moral norms when they expect that a sufficiently large subset of others will. Unless people form the empirical expectation that cooperation is the "done thing" in a group,

individuals do not have sufficient reason to comply with the normative exhortations of their fellows. Consequently, simply being preached to that I should adopt a rule, and even the true reports of others that they endorse it, does not tell me what I really need to know: do people actually follow it?

Attitude change is, moreover, often costly — imbuing people with new attitudes typically requires sustained persuasion, and to be effective we often need to know something about the target's preexisting beliefs and attitudes. Not only is this often costly but, as we saw in section 2.1, even if attitudes are changed, this is an uncertain indicator as to whether behavior will be. Given that attitude change tends to be a rather costly activity that is of dubious reliability, investing in it is generally rational only if one will have sustained relations with the target, and inducing his cooperative actions will have fairly significant payoffs to the preacher. In small group settings where the interests of the participants are tightly bound, the investment may be rational. Knowledge of the other's preexisting attitudes may be readily available, and sustained interaction is apt to provide many opportunities for preaching. In this case some of the conditions for direct reciprocity and iterated games can employ preaching; when others fail to act on the cooperative norm, we respond by preaching internalization (Kliemt, 1990). But as the size of the cooperative network expands to include many strangers, this dynamic becomes impossible to sustain (see §5). The number of interactions increases rapidly, and opportunities for effective preaching to any one individual are drastically reduced. Moreover, as networks become heterogeneous, it becomes more difficult to know for any given target what an effective sermon would be. In these cases preaching takes on the features of a public good: let others devote themselves inculcating pro-social attitudes, while I devote myself to maximizing my gains within the existing attitude parameters. If, as most have argued in the social-moral norm literature, the evolution of social norms is a key factor in scaling up human cooperation beyond the small band (Gaus, 2015), then the attitude change model seems unable to do precisely what is required of a norm-focused account. The rational strategy in large groups is to refrain from investing in norm change.¹²

It is important that in the Hayekian “Great Society” (which Buchanan takes as equivalent to his “moral order”) individuals interact with strangers, with whom they often do not expect to meet again. Thus the rule network of the moral order is not simply an outcome of the interaction of many small-scale norm networks in which preaching may be efficient. Indeed, a recurring finding is that small-scale norm networks (“moral communities”) are based on a strong distinction between “Us” and “Them.” It seems that the evolution of humans as norm-followers may be at the root of this distinction: in learning to follow norms it was important to know who was one of “Us,” so that one knew what norms and behaviors to comply with — ours, not the outsiders (Henrich, 2016: 204; Boyd and Richerson, 2005: 113-17). But this very sensitivity to “Us-ness” is one of the most serious challenges to truly mass moral orders — it all-too-easily produces the sort of “tribalism” against which Hayek fought (Greene, 2015). The breakout from small-scale to large-scale norm networks is a fundamental moment in the evolution of the free society, and was not accomplished simply through the interaction of small-scale networks.

4.4 Preaching and norm stabilization

While of course they no doubt have some role, preaching and attitude change are, I think, inadequate grounds for an account of the origin of norms in large-scale settings, especially without a centralized “norm change agent.” Until people expect others to follow a norm N , most preaching will have little effect. However, because preaching is more effective when behavior and sermon are consistent, preaching can have a significant role in stabilizing existing norms. This hypothesis is implied in Bicchieri’s most recent statement of her theory of social norms. On her account, individuals are tempted to defect from norms to secure opportunistic self-interested gains. For any norm N and any given individual, she will have a certain “sensitivity,” k_N , to the norm, which determines how much weight she gives it in conflicts with opportunistic gains.

Sensitivity in this case refers to how much a person adheres to what the norm stands for.^[13] Norm sensitivity embodies one’s personal reasons for adhering to the norm. A highly sensitive individual could list several good, important reasons for why a particular norm should be enforced, whereas a low sensitivity individual, who doesn’t care much about what the norm stands for, may only list the fact that, since the norm is widespread, it makes sense

for her to obey it (to avoid the sanctions that transgressions incur). Let's call a person's sensitivity to a particular norm, N , k_N . For example, a person who is not very convinced of the advisability of child marriage will have very low sensitivity to that norm (in other words, a very low k_N), whereas a person who is convinced that child marriage is the best way to protect a child's honor will be highly sensitive to such a norm (Bicchieri, 2016: 165).

While preaching is unlikely to be sufficient to get the target to adopt a norm, it does increase people's sensitivity to current norms.¹⁴ In contrast to norm adoption, which has significant threshold properties (one must be "convinced enough" to adopt a norm), sensitivity is a continuous variable, so even if preaching is not terribly powerful, we can expect marginal increases in preaching to have some marginal effects in increasing sensitivity to a norm. And when a person's sensitivity to a norm is increased, she is less tempted by opportunistic cheating.¹⁵

4.5 Normative Expectations (and Expressing Them)

Buchanan's model tends to collapse two functions of preaching: (i) changing the attitudes of others, which we have thus far been examining, and (ii) alerting others to our normative expectations of them. When Betty preaches to Alf about why he should comply with the norm and not set back her interests, she is expressing her normative expectations to him. As Brennan and Lomasky (1993) have famously argued, expressive action is very low cost (indeed, we usually positively enjoy it), and is apt to occur in many-person interactions when the agent is very unlikely to be significant in securing a consequential result. When Betty observes Alf inflicting "moral externalities" on her — in the sense that she observes him flouting a current norm in a way that sets back her interests — she certainly is apt to convey to him that he has disappointed her normative expectations, and this can matter. People's norm-based behavior is driven by both their empirical expectations and their beliefs about what others normatively expect of them.

One of the most striking "social experiments" based on this insight about the importance of expressing normative expectations was that of Antanas Mockus, mayor of Bogotá in the late 1990s and early 2000s (Mockus, 2012; Tognato, 2017). Bogotá was characterized by a very high rate of traffic fatalities in the mid-1990s, with widespread disregard for traffic regulations. Mockus distributed 350,000 "Thumbs Up/Thumbs Down" cards that drivers could display in response to driving that violated the formal

rule, to drive home the message that such behavior violated the normative expectations of other drivers. Along with related programs, Bogotá witnessed a 63% decrease in traffic fatalities between 1995-2003. The normative expectations of others do matter to us; signaling that we have them can stabilize compliance and reduce opportunistic cheating.

5 THE IMPLAUSIBILITY OF A COMPREHENSIVE HOBBSIAN ACCOUNT OF THE EMERGENCE OF THE MORAL ORDER

That Buchanan's preaching model of the origin of moral norms fails does not, of course, show that the Comprehensive Hobbesian project fails. Perhaps other essentially Hobbesian accounts, such as those of Binmore (2005) or Moehler (2014, 2016) succeed in showing the origins of informal norms lie simply in self-interest. Obviously, we cannot go into the details of these interesting accounts here. Yet we have good reason to suspect that any account of the origin of norms that presupposes only Hobbesian maximizers will fail to result in the emergence of large-scale moral norms. A generalized argument for the implausibility of any Hobbesian account of the origins of large-scale moral norms is suggested by an early essay of Buchanan's (1965), in which he explores the special difficulties of rule following in large populations. I alter, and extend, his presentation.

We now suppose that Alf has a choice between two rules of behavior; a rule "commits"¹⁶ to certain actions (Buchanan, 1965: 2). Suppose Alf has a choice between two rules: a moral norm N that secures the benefits of cooperation and social order (for example, by securing Paretian gains in mixed motive games) and R , which is Alf's personal maxim of behavior that allows him to secure what he sees as the best results in an unconstrained way (Buchanan, 1965: 2).¹⁷ Now true Kantian-inclined agents are willing to follow N even in the face of large-scale defections; at a limit, the true Kantian would commit to N regardless of the behavior of others.¹⁸ However, for Hobbesian agents the decision to follow a moral norm involves securing for themselves the benefits of social cooperation, and these benefits can generally only be secured if enough others are acting on it. Let $\mu_a(N_x)$ be the utility Alf receives if x number of others act on the cooperative norm N , and $\mu_a(R_x)$ be the utility Alf would receive if, instead, he acted on his own personal rule (where x remains the number of others acting on N). Now what the

Hobbesian would like say is that Alf should act on N if $\mu_a(N_x) > \mu_a(R_x)$. Obviously, the Hobbesian would say, Alf should not act on N if most others do not (x is low), because only a few acting on N will at best secure meager cooperative benefits, so it would be better for Alf to act on his own personal rule. Alf and a few others cannot bring about social cooperation. But if the number of cooperators begin to rise, Alf might find that the benefits of acting on N exceed that of acting on his personal rule, R , which allows unconstrained behavior.

The attraction of this set-up is that it opens a way for the Hobbesian to explain the emergence of norms though appeal to cascades, based on sensitivity to empirical expectations of the behavior of others. Suppose that we have a population of diverse Hobbesian agents, who place a wide range of value on the benefits of social cooperation as well as the value of acting on personal unconstrained rules, and who vary in their estimates of how many people need to act on the norm to secure significant cooperative benefits.¹⁹ Consider group α , composed of people who deem the benefits of cooperation great, the number needed to produce it modest, and the costs in terms of forgone opportunities to act on personal rules low. Thus for even low values of x , they will tend to hold that $\mu_\alpha(N_x) > \mu_\alpha(R_x)$, and so act on N . Suppose that a “moral entrepreneur” (Buchanan, 2001, 17: 442) or a “trendsetter” (Bicchieri, 2016: chap. 5) in the α group decides that there are enough α s to get the norm going, and succeeds in gaining allegiance to N in this pro-norm group. This will change the calculations of group β , who, say, were slightly more skeptical that sufficient others were willing to join in, or who believed a somewhat greater number is needed to secure significant cooperative benefits; but now that α has adopted N , group β might revise their judgments, and conclude that for them, given this new level of participation y , $\mu_\beta(N_y) > \mu_\beta(R_y)$. We can see how a cascade might get underway, as each slightly more less- N inclined group comes to the conclusions that its “threshold” for participation has been reached (Gaus, forthcoming).

The unstated assumption of the above cascade model — which is truly a model of norm emergence — is that Alf can only secure the cooperative benefits of the norm if he acts on N himself. So Alf keeps on looking at the cooperative gains of acting on N given the number of others doing so, and switches at some point to secure these gains when

they have surpassed his threshold value (i.e., they are greater than $\mu_a(R_x)$). However, none of this applies if Alf receives $\mu_a(N_x)$ whether or not he acts on N himself. In that case his choice is between acting on N and receiving $\mu_a(N_x) - \mu_a(R_x)$, or acting on R and receiving $\mu_a(R_x) + \mu_a(N_x)$ — hardly a difficult choice. As Buchanan (1975) argues, in this case for any level of participation x , the latter strictly dominates the former. Recall further (§4.1) that the gains to Alf of x others acting on N comes from eliminating the ethical externalities they impose *on him*. His own acting on N provides him with no direct benefits whatsoever, and thus does not in any way increase $\mu_a(N_x)$.

Unless, that, is, others make their acting on N toward Alf conditional upon his acting on it. As Buchanan (1975) stresses, if others can withhold from Alf the benefits of their norm following (say, by continuing to impose externalities on him if he fails to act on N), then this conditionality imposes on Alf a necessary cost for receiving $\mu_a(N_x)$; he must forgo $\mu_a(R_x)$. This is by now the familiar lesson of the Folk Theorem in repeated games (Binmore, 2005: 79-82). If one can deny the benefits of cooperation to defectors, then a stable cooperative equilibrium can be achieved. The Folk Theorem formalizes this idea of “direct reciprocity.” Suppose Alf and Betty are playing a series of games where cooperation yields mutual benefits, but unilateral defection is the best payoff in any single play (e.g., a series of Prisoner’s Dilemmas as in Figure 1). Alf may be tempted to defect in game 1 to gain the unilateral defection payoff, $\{D_a, C_b\}$; but if (given certain constraints about payoffs, expectations for future games and discounting the future) Betty threatens to respond with her own defection in game(s) 2 (or onwards), pushing Alf’s payoff to $\{D_a, D_b\}$ (i.e., his “maximin payoff”) in those subsequent games, it will be rational for Alf to cooperate rather than defect in game 1.²⁰ In small-scale settings (dyads, triads, etc.) with long histories of play, direct reciprocity can sustain cooperation, but it has severe problems scaling up to even medium-sized groups (Henrich and Henrich 2007: 48ff; Bowles and Gintis, 2011: 63-8). The “bookkeeping” costs become excessive, as each must keep track of the history of all her dyadic interactions, and the costs and benefits of each. These demands are so excessive that there is increased skepticism about whether direct reciprocity is more than a minor basis for human cooperation (Boehm, 2012: 60ff).

Much more plausible is “indirect reciprocity,” or generalized reputation. Indirect reciprocity does not require detailed information on the history of Alf’s play with Betty, but more general information, which can be shared among norm compliers, as to whether Alf is in general a norm complier or a non-complier. This information is much easier to secure and disseminate. The problems, as I have pointed out elsewhere (Gaus, 2011: 90-6), is that the information is rather blunt, subject to strategic manipulation, and sensitive to being degraded by false gossip. It is blunt because, given the number of norms that a person may or may not be following at any given time, reputational information tends to sum up whether the person “is a good group member” or a “reliable person.” But this can be distorting. As Henrich and Henrich discovered in their ethnographic studies, a person can secure a reputation as “unreliable” if he has flouted an important group norm that is not really indicative about his cooperativeness — such as marrying into a different ethnic group (Henrich and Henrich, 2007: chap. 4). When general reputations are important, people come to strategize over gaining reputations, where this may not cohere with genuine conditional cooperation. Again, as Henrich and Henrich discovered, when general reputations are critically important, people are very reluctant to withdraw cooperation from anyone, since this might be falsely interpreted by third parties as a defection rather than a reply to another’s defection. The costs of being mistakenly branded a defector are high, thus one tends to cooperate with anyone who is not already widely known as a defector. This, then undermines the effectiveness of reputations, as many opportunistic defectors can, after all, gain the benefits of cooperation. Moreover, if the aim is to secure reputations, people invest significant resources in broadcasting — and overemphasizing or simply lying about — their cooperative accomplishments. Lastly, malicious gossip can quickly undermine the reputations of cooperators (Vanderschraaf, 2007). Again, all these problems become debilitating when the moral order is one among strangers for whom any reputational information is, at best scant.

In some models, withholding the benefits of cooperation is understood as a sort of punishment, but in the present model we should recall that if Alf is not complying with N , he is imposing ethical externalities on others; for them to withdraw is to allow

themselves to impose externalities on him. This need not be costly to them given his current non-compliance. Whether or not we deem this to be “punishment,” we should distinguish it from acts where Betty, at a positive additional cost to herself, imposes sanctions on Alf with the aim of reducing or eliminating his gains from noncompliance. Although, as I pointed out (§2.1), punishment is necessary to stabilize norms, it is very difficult to plausibly model how it can arise in a large populations of Hobbesian agents. When x , the number complying with the norm, is large, the benefits to any one individual non-complier from defection are apt to be high. Alf receives $\mu_a(N_x)$, the benefits of x individuals not imposing ethical externalities on him as he goes about his business. For Betty to alone effectively punish him she would essentially have to impose negative (punishment) utility that would drive his utility below what he would have receive if he complied with the norm, so erasing all his gains from defection. That is apt to be very costly to her, and it would seem irrational for her to do so. More plausible is to model punishers as forming a quorum — a number that is sufficient to punish without high costs to any member (Bowles and Gintis, 2011: 150-63). But in extensive societies the punishing group also would have to be extensive to minimize punishing costs on any individual. Among Hobbesian agents this would require group norms about punishing to ensure that each punisher does her part. This, though, would be to suppose the presence of a norm to explain their emergence. Generally, it is extraordinarily difficult to model Hobbesian agents acting as punishers in large groups.²¹

After reviewing this general model, Buchanan concludes that there is “elemental truth” in “the old adage, ‘Never trust a stranger’” (1975: 8). If, though, the aim was to provide a Comprehensive Hobbesian account of the origins of moral order, which was to undergird the constitutional political economy project, this admission is unsettling. Hobbesian models probably can yield the emergence of norms in small groups, but Buchanan’s account of the origins of the moral order was meant to push beyond the limits of “moral community” (and, we might say, tribalism) to account for a large-scale moral order among strangers (§3). It seems very doubtful that he succeeded.

6 RETHINKING THE CONSTITUTIONAL POLITICAL ECONOMY PROJECT

6.1 *Rule-Following v. Best Response*

Whatever their differences, the disciplines of economics and philosophy embrace *Homo Rationalis* as their basic model. *Homo Economicus* is one articulation, depicting the rational individual as a maximizer, making her best response to the maximizing choices of others. Hayek refused to endorse *Homo Economicus* and best response as the mainstay of his analysis. Explicitly drawing on cultural anthropology, he (1973: 74-81) argued that our reasoning abilities were an adaptive, evolutionary, response to the challenges of human group life, and that norm-following and purpose seeking are equally basic to rationality. Hayek repeatedly stressed that humans are learners — we learn to follow rules, the functions of which we do not fully understand. This feature of our rationality is fundamental to Vernon Smith's (2008: 13-42) inquiry into ecological rationality and Douglas North's (2016) learning-based institutionalism. The last twenty years has witnessed an increasing recognition of the importance to economic analysis of humans as cultural beings (Storr, 2013). Indeed, there has been an explosion of work in cultural anthropology and cultural evolution supporting the Hayekian insight (Boyd and Richerson, 1985, 2005; Richerson and Boyd, 2005; Henrich and Henrich, 2007; Boehm, 2012). Henrich (2016) summarizes a large body of these findings, stressing that in a wide variety of contexts those who rely on cultural norms radically outperform those who seek to individually judge for their best option for themselves.

Once we view humans as fundamentally cultural beings, individual maximizing rationality no longer uniquely defines us. Indeed, there is reason to think that our true excellence may be in copying and imitating. In a fascinating series of experiments, Call, Carpenter and Tomasello (2005), compared learning in chimps (median age 11 and a half years) and two-year old human children. Chimps were found to be better at emulating a successful process; observing what worked, they reproduced it, observing what did not, they ignored it. Humans, in contrast, simply copied the actions of the "teacher." Horner and Whiten (2005) arrived at similar conclusions, providing evidence that Chimps rely more on causal reasoning (the glory of *Homo Economicus*) while children basically copied whatever they observed, causally efficacious or not. A child, some have said, is essentially a "cultural sponge" (Mesoudi, 2011: 15). Other experiments indicate that

adult humans excel at copying (Henrich, 2016: 112). We tend to copy those who strike us especially successful or who have high prestige: seeing the way they do things, most of us easily imitate their actions — without much causal understanding of what is going on. Henrich (2016: 165) echoes Hayek: “in many cases people don’t understand how or why their norms work or that their norms are even ‘doing’ anything.”

Being rule-guided follows from being cultural copiers. When a child sees a bit of behavior by a person judged to be competent, she must form a rule: in situations like *C*, do such-and-such. As Hayek (1952), recognized, categorization of behavior implies rules. To copy we must categorize, and so form behavioral rules. And children automatically convert these categorizations into norms: seeing that doing such-and-such in circumstances is associated with good results, children convert this into a rule “Do such-and such in these circumstances!” and “It is wrong not to” (Schmidt and Tomasello, 2012). Whereas imitation is inherently classification and rule-governed, best response maximizing is inherently act-focused: what ought I to do right now that gives me the best results, given what others are doing? The attractions of best-response reasoning are so powerful that they have led generations of philosophers and economists down the garden path of seeking to reduce rule-governed behavior to enlightened best response. The results have been impressive displays of creativity and ingenuity but, I would venture, there is no adequate solution to the puzzle how maximizers can use maximizing reasoning to get themselves not to maximize (Gaus, 2011: chap. 2).

Economists and philosophers typically resist (to understate matters) this line of thinking: it severely limits the scope of “rational choice reductionism” (Buchanan, 2001, 17: 439), as well as the aspiration for thoroughgoing rational justification of our moral and social practices (Gaus, 2017). As Hayek so often stressed, rational constructivist social contracts are replaced by evolutionary mechanisms and self-organizing systems. Political philosophy no longer can give comprehensive blueprints for social and political orders. However, what we have seen here is that acknowledging these limits allows us to ground politico-legal orders on a powerful account of the underlying moral order. And as our insight into the moral order deepens, best-response reasoning will have only a subsidiary role in its explanation.

6.2 Reimagining Constitutional Politics Economy

Constitutional political economy thus confronts a fundamental dilemma. To stay true to Comprehensive Hobbesianism's commitment to rational choice all the way down, it would have to embrace the fatal lacuna (§§1, 2.3), seeing government and law arising out of a thoroughly disordered state of nature populated by egoists, who manage to see that their true best response to the strategies of others is to agree to the rule of laws and their enforcement. It is not perhaps surprising that so many see Buchanan's constitutional political economy project in this light — it is the natural way to understand Comprehensive Hobbesianism. Yet, I have argued, it is an unattractive option that runs counter to our increasing knowledge of the normative basis of order. And despite the popular impression, Buchanan did not take this route, but consistently explored ways to ground the state on a moral order which was itself grounded on Comprehensive Hobbesianism. This more sophisticated Hobbesian project, I have argued, also fails. In the end Buchanan's constitutional political economy project is torn between its commitment to Comprehensive Hobbesianism and its acceptance of a large-scale moral order. It cannot have both.

In reimagining constitutional political economy without Comprehensive Hobbesianism we might identify two visions. A modest revision would more fully recognize that the appropriate baseline for political negotiation is often the moral order, not simply a clash of individual interests. On this reimagining the distinction between the protective and productive state (Buchanan, 1975: 95-98) becomes even sharper. As the Protective State focuses on basic individual rights, its baseline will be the norms in the moral order about, say, harm to others and various jurisdictional rights, including some contours of property rights (Gaus, 2011: chap. 6). These define the accepted moral baseline, and the formal state apparatus would seek Pareto-improvements on it. Individual rights would not be solely — perhaps not even primarily — the outcome of individual interest-based bargaining, but would be critically about interpretations of the demands of the moral order (though, as Hobbes (1651: 180) recognized, these disputes are always affected by interests). Much of this, I think, is nascent in Buchanan's work (e.g., 1975: chap. 1), but it has often not been well-integrated into the constitutional

political economy project, which so often supposes an amoral baseline. In contrast, because the Productive State's concern is supplying public goods that are in the interests of individuals to supply, Buchanan's (2016) fundamental insights about a Wicksellian-inspired politics as exchange endures as the core of this part of constitutional political economy.

However, when we reflect not simply on the fact that a preexisting moral order obtains, but on the theory of rationality which accounts for it (§5), a more radical reimagining seems called for. *Homo Economicus* and its assumption of self-interest are no longer general accounts of rational action (Buchanan, 2016), but become domain-specific. The rational basis of individual action is complex, and only sometimes can it be best understood as maximizing self-interest. That, however, calls into question the import of Buchanan's key "argument from symmetry," which insists that "it is illegitimate to restrict *Homo Economicus* to the domain of market behavior while employing widely different models of behavior in nonmarket settings, without any coherent explanation of how such a behavioral shift comes about" (Brennan and Buchanan, 1986: 50).²² If *Homo Economicus* is inherently domain-specific, we ought not to begin with the supposition that it is a general model and we need an explanation why *not* to apply it; rather, we should endeavor to better understand its legitimate domain. When, say, do people maximize by playing subgame perfect equilibria and when do they trust others even when doing so fails to make the best (i.e., equilibrium) response (Smith, 2016)? When do they bargain about interests and when will they sacrifice interests for ideologies? And when will they act selfishly, and when will they act fairly and contribute to public goods (Henrich and Smith, 2004; Smith, 2008: chaps. 10-11)? I certainly do not believe that in the face of these types of questions the constitutional political economy project withers, but it becomes much more complex. It demands a more nuanced model of individual action, individual learning, and their relation to the overall decision-making context or, we might say, its ecology (Smith, 2008). The upshot is perhaps closer to Elinor and Vincent Ostroms' (2004) IAD model rather than the more elegant version we have learned from Buchanan.

Constitutional political economy is not a completed artifice to be admired and defended, but an ongoing project, constantly refining its assumptions and analysis. Such is the nature of science. Discovering tensions moves us forward, pointing us toward the next step — and a better account of social cooperation and public choice. James Buchanan got us on our way. It is up to us to move further.

Notes

¹ “Commonly understood,” as it is debatable whether this is Hobbes’s considered view (Gaus, 2013). In this essay I assume this common view of the Hobbesian project.

² These games can be combined. See Vanderschraaf (2006); Chung (2015).

³ See Kliemt (2009: 46ff); for a criticism, see Gaus (2008: 19ff).

⁴ I do not distinguish here social and moral norms, though distinctions certainly can be drawn (Brennan et al., 2013: 57ff.; Platteau, 2000: 291ff.) As Bicchieri observes (2006: 21), her notion of social norms is essentially what Hume means by “justice.” I have employed the term “social morality” to refer to rule-based moral claims, and to distinguish such claims for personal moral claims that do not suppose social uptake (Gaus, 2011: 2ff., 163ff.)

⁵ In ultimatum games one person has the role of proposer and another responder. In a common version proposer is given an amount of money; he can propose any division he wants. Responder then can either accept or reject. If responder accepts, both parties get what the proposer offered; if responder rejects, no one gets anything. Offers in which the responder gets less than 20% are routinely rejected. The typical offer gives the responder between 50% and 40%. In dictator games, the proposer simply divides the money in any way she wants, and the game’s over — not really much of a game.

⁶ See Gaus (2011: chap. 4). For cognitive modelling of norm internalization processes, see Andrighetto, Villatoro and Conte (2010). See further section 4 below.

⁷ Though he allows that there may be more than one type of explanation of the basis of the moral order (2001, 17: 231).

⁸ I am simplifying Buchanan’s presentation. In his paper on “Moral Externalities” (which Buchanan cites at 2001, 17: 215n as providing a more detailed analysis), Hartmut Kliemt develops an account of moral externalities in which the “moral externality” arises out of an agent *seeking to impose a moral rule* on another (1990: 37). In the paper being discussed here (“The Economic Origins of Ethical Constraints”), however, Buchanan sees moral rules as attempts to limit damages to one caused by “nonconstrained behavior” of others (2001, 17: 216). So on Kleimt’s interpretation the moralist imposes an externality, while in this essay by Buchanan, ethical rules are responses to externalities engendered by unconstrained pursuit of self-interest. I am analyzing the latter idea here.

⁹ I assume that there is no inducement Betty would be willing to offer that would get Alf to contract to play γ_a . Note that Betty also imposes an externality on Alf by playing γ_b rather than x_b . I shall simply focus on Alf’s externality here. In Buchanan’s more complex model, which involves an assumption of symmetry, both Alf and Betty must change their move to secure the “ethical equilibrium.”

¹⁰ Although it does not pick out one equilibrium in a multiple equilibria game, as do standard conventions. But the compliance dynamic is the same.

¹¹ This was not the case in Buchanan's (2001, 17: 219-24) example, where both switch play, from the $\{Y_a, Y_b\}$ equilibrium to a new $\{X_a, X_b\}$ equilibrium. The simpler case considered here is actually a more plausible account of norm emergence, since here Betty is already acting on the Y rule, and so her preaching would have more effect. But that asymmetry is only for the sake of simplicity of presentation.

¹² This is not to deny that some forms of preaching are effective once a central agency takes over the task of supplying the "public good." Norm-change technologies in large-scale societies such as media campaigns and "edutainment" of well-designed soap operas may well transform attitudes. For example, students of large-scale norm change have recently focused on so-called "edutainment" such as soap operas produced in Latin America, seeking to change gender and workplace norms. See Bicchieri, 2016: chap. 5.

¹³ See here Bicchieri, 2006: 52.

¹⁴ "Persistence can thus be explained without explaining emergence" (Brennan et al., 2013: 150).

¹⁵ For similar reasons preaching is no doubt important in the process of norm abandonment by reducing sensitivity. But without a change in the empirical expectations — which preaching alone cannot accomplish — norm change will not occur. See Bicchieri, 2016: 119-27.

¹⁶ In this model, then, Buchanan is supposing that ethical rule following involves a weak type of commitment strategy; but we shall also see he allows constant reevaluation of the choice of which commitment strategy to adopt.

¹⁷ Buchanan insists that the choice should be between two rules, not rule v. simply unconstrained behavior. In the present context nothing rides on the distinction.

¹⁸ I have modeled populations composed of such quasi-Kantian agents and agents who have stronger conditional preferences for norm following. Under a surprising range of parameters, even the quasi-Kantians opt for norms that are followed by all (Gaus, forthcoming).

¹⁹ As Hobbes (1651: 23) points out, individuals not only have divergent interests, but divergent judgments.

²⁰ Buchanan (2000) draws on iterated game results to argue that Hayekian cultural evolution is not necessary to explain rule-following behavior.

²¹ Not to say that many do not try. See Hampton, 1986, 176ff.

²² See further Gaus, 2010. This essay has, alas, often miffed followers of Buchanan. I hope the present effort will prove to be less irritating.

REFERENCES

- Andrighetto, Giulia, Daniel Villatoro and Rosaria Conte. (2010). "Norm Internalization in Artificial Societies." *AI Communications*, vol. 23: 325–339.
- Baier, Kurt. (1995). *The Rational and the Moral Order: The Social Roots of Reason and Morality*. La Salle, IL: Open Court, 1995.
- Bicchieri, Cristina. (2006). *The Grammar of Society: The Nature and Dynamics of Norms*. Cambridge: Cambridge University Press.
- . (2016). *Norms in the Wild: How to Diagnose, Measure, and Change Social Norms*. New York: Oxford University Press.
- and Alex Chavez. (2009). "Behaving as Expected: Public Information and Fairness Norms." *Journal of Behavioral Decision Making*, vol. 22: 191–208.
- and Alex K. Chavez (2013). "Norm Manipulation, Norm Evasion: Experimental Evidence." *Economics and Philosophy*, volume 29 (Special Issue 2, July): 175-98.
- and Erte Xiao. (2009). "Do the Right Thing: But Only if Others Do So." *Journal of Behavioral Decision Making*, vol. 22: 191-208.
- Binmore, Ken. (2005). *Natural Justice*. Oxford: Oxford University Press.
- Boehm, Christopher. (2012). *Moral Origins: The Evolution of Virtue, Altruism and Shame*. New York: Basic Books.
- Bowles, Samuel and Herbert Gintis. (2011). *A Cooperative Species: Human Reciprocity and its Evolution*. Princeton: Princeton University Press.
- Boyd, Robert and Peter J. Richerson. (1985). *Culture and the Evolutionary Process*. Chicago: University of Chicago Press.
- . (2005). *The Origin and Evolution of Cultures*. New York: Oxford University Press.
- Brennan, Geoffrey, Lina Eriksson, Robert E. Goodin and Nicholas Southwood. (2013). *Explaining Norms*. Oxford: Oxford University Press.
- Brennan, Geoffrey and James Buchanan (1985). *The Reason of Rules: Constitutional Political Economy*. Cambridge: Cambridge University Press.
- Brennan, Geoffrey and Loren Lomasky. (1993). *Democracy and Decision*. Cambridge: Cambridge University Press.
- Buchanan, James M. (1965). "Ethical Rules, Expected Values and Large Numbers." *Ethics*, vol. 76 (October): 1-13.
- . (1975). *The Limits of Liberty: Between Anarchy and Leviathan*. Chicago: University of Chicago Press.
- . (2000). "The Legacy of Friedrich von Hayek: vol. 7 Morality and Community in the Extended Market Order." (Audio). Liberty Fund: Indianapolis.

- <<http://oll.libertyfund.org/titles/2153>>
- . (2001). *The Collected Works of James M. Buchanan*. Indianapolis: Liberty Fund (cited by volume).
- . (2016 [1986]) “The Constitution of Economic Policy.” In *Mainline Economics: Six Nobel Lectures in the Tradition of Adam Smith*, edited by Peter J. Boettke, Stefanie Haeffele-Balch and Virgil Henry Storr. Arlington, VA: Mercatus Center: 43-59.
- and Gordon Tullock. (1965). *The Calculus of Consent*. Ann Arbor: Ann Arbor Paperbacks.
- Call, Josep, Malinda Carpenter and Michael Tomasello. (2005). “Copying Results and Copying Actions in the Process of Social Learning: Chimpanzees (*Pan Troglodytes*) and Human Children (*Homo Sapiens*).” *Animal Cognition*, vol. 8: 151–163.
- Chung, Hun. (2015) “Hobbes’ State of Nature: A Modern Bayesian Game-Theoretic Analysis.” *Journal of the American Philosophical Association*, vol. 1 (October): 485-508.
- Darwall, Stephen. (1995). *The British Moralists and the Internal “Ought.”* Cambridge: Cambridge University Press.
- Friedman, Daniel (2008). *Morals and Markets: An Evolutionary Account of the Modern World*. New York: Routledge.
- Gaus, Gerald (2008). *On Philosophy, Politics, and Economics*. Belmont, CA: Thomson-Wadsworth.
- . (2010). “The Limits of *Homo Economicus*.” In *Essays on Philosophy, Politics & Economics*, edited by Christi Favor, Julian Lamont and Gerald Gaus. Stanford: Stanford University Press: 14-37.
- . (2011). *The Order of Public Reason: A Theory of Freedom and Morality in a Diverse and Bounded World*. Cambridge: Cambridge University Press.
- . (2013) “Hobbesian Contractarianism, Orthodox and Revisionist.” In *The Bloomsbury Companion to Hobbes*, edited by S.A. Lloyd. New York: Bloomsbury: 263-278.
- . (2015). “The Egalitarian Species.” *Social Philosophy and Policy*, vol. 31: 1-27.
- . (2017). “Hayekian ‘Classical’ Liberalism.” In *The Routledge Handbook of Libertarianism*, edited by Jason Brennan, David Schmidtz and Bas van der Vossen. New York: Routledge.
- . (forthcoming). “Self-Organizing Moral Systems: Beyond the Social Contract.” *Politics, Philosophy and Economics*.
- Green, T. H. (1889 [1986]). *Lectures on the Principles of Political Obligation*. In *Lectures on the Principles of Political Obligation and Other Writings*, edited by Paul Harris and John

- Morrow. Cambridge: Cambridge University Press, 1986: 13–193.
- Greene, Joshua (2013). *Moral Tribes: Emotion, Reason, and the Gap Between Them and Us*. New York: Penguin.
- Hampton, Jean. (1986). *Hobbes and the Social Contract Tradition*. Cambridge: Cambridge University Press.
- Hayek, F. A. (1952 [1976]). *The Sensory Order: An Inquiry into the Foundations of Theoretical Psychology*. Chicago: University of Chicago Press.
- . (1973). *Rules and Order*. Chicago: University of Chicago Press.
- . (1988). *The Fatal Conceit: The Errors of Socialism*, edited by W. W. Bartley. Chicago, IL: University of Chicago Press.
- Henrich, Joseph. (2016). *The Secret of Our Success*. Princeton: Princeton University Press.
- and Natalie Smith. (2004). “Comparative Evidence from Machiguenga, Mapuche, and American Populations.” In *Foundations of Human Sociality: Economic Experiments and Ethnographic Evidence from Fifteen Small-Scale Societies*, edited by J. Henrich et al. Oxford: Oxford University Press: 125–67.
- Henrich, Natalie and Joseph Henrich. (2007). *Why Humans Cooperate: A Cultural and Evolutionary Explanation*. Oxford: Oxford University Press.
- Hobbes, Thomas. (1651 [1994]). *Leviathan*, edited by Edwin Curley. Indianapolis: Hackett.
- Hopfensitz, Astrid, and Ernesto Reuben. (2009). “The Importance of Emotions for the Effectiveness of Social Punishment.” *The Economic Journal*, vol. 119 (540): 1534–1559.
- Horner, Victoria and Andrew Whiten. (2005), “Causal Knowledge and Imitation /Emulation Switching in Chimpanzees (*Pan Troglodytes*) and Children (*Homo Sapiens*).” *Animal Cognition*, vol. 8: 164–181.
- Kitcher, Philip. (2011). *The Ethical Project*. Cambridge, MA: Harvard University Press.
- Kliemt, Hartmut. (1990). “Moral Externalities.” In his *Papers on Buchanan and Related Subjects*. Munich: Accedo Verlagsgesellschaft: 37-60.
- . (2009). *Philosophy and Economics I: Methods and Models*. Munich: Oldenburg.
- Mackie, Gerry (2017). “Effective Rule of Law Requires Construction of a Social Norm of Legal Obedience.” In Carlo Tognato, *Cultural Agents Reloaded: The Legacy of Antanas Mockus*. Cambridge, MA: The Cultural Agents Initiative at Harvard University.
- Mesoudi, Alex. (2011). *Cultural Evolution*. Chicago: University of Chicago Press.
- Mockus, Antanas. (2012). “Building ‘Citizenship Culture’ in Bogotá.” *Journal of International Affairs*, vol. 65 (Spring / Summer 2012): 143-6.

- Moehler, Michael. (2014). "The Scope of Instrumental Morality." *Philosophical Studies*, vol. 167): 431-451.
- . (2016). "Orthodox Rational Choice Contractarianism Before and After Gauthier." *Politics, Philosophy and Economics*, vol. 15 (May): 113-31.
- Montesquieu, Baron De. (1748 [1949]). *The Spirit of the Laws*, translated by Franz Neumann. New York: Hafner Press.
- North, Douglass C. (2016 [1993]). "Economic Performance Through Time." In *Mainline Economics: Six Nobel Lectures in the Tradition of Adam Smith*. edited by Peter J. Boettke, Stefanie Haeffele-Balch and Virgil Henry Storr. Arlington, VA: Mercatus Center: 81-99.
- Ostrom, Elinor and Vincent Ostrom. (2004). "The Quest for Meaning in Public Choice." *American Journal of Economics and Sociology*, vol. 63 (January): 105-47. Special issue on the "Production and Diffusion of Public Choice Political Economy: Reflections in the VPI Center."
- Platteau, Jean-Philippe. (2000). *Institutions, Social Norms and Economic Development*. Amsterdam: Harwood Academic Publishers.
- Richerson, Peter J and Robert Boyd. (2005). *Not by Genes Alone: How Culture Transformed Human Evolution*. Chicago: University of Chicago Press, 2005.
- . (2008). "The Evolution of Free Enterprise Values." In *Moral Markets: The Critical Role of Values in the Economy*, edited by Paul Zak. Princeton: Princeton University Press: 107-41.
- Robinson, Paul. (2000). "Why Does the Criminal Law Care What the Layperson Thinks Is Just? Coercive versus Normative Crime Control." *Virginia Law Review*, vol. 86 (Symposium: The Legal Construction of Norms): 1839-1869.
- Rose, David C. (2011). *The Moral Foundations of Economic Behavior*. (New York: Oxford University Press.
- Schmidt, Marco F. H. and Michael Tomasello, (2012). "Young Children Enforce Social Norms." *Current Directions in Psychological Science*, vol. 21(4) 232–36.
- Schwab, David and Elinor Ostrom. (2008). "The Vital Role of Norms and Rules in Maintaining Open Public and Private Economies." In *Moral Markets: The Critical Role of Values in the Economy*, edited by Paul Zak. Princeton: Princeton University Press: 204-27.
- Sen, Amartya. (1982). "Rational Fools." In his *Choice, Welfare and Measurement*. Cambridge, MA: Harvard University Press: 84-106.
- Smith, Vernon (2008). *Rationality in Economics: Constructivist and Ecological Forms*. Cambridge: Cambridge University Press.

- Smith, Vernon (2016). "Constructivist and Ecological Rationality in Economics." In *Mainline Economics: Six Nobel Lectures in the Tradition of Adam Smith*, edited by Peter J. Boettke, Stefanie Haeffele-Balch and Virgil Henry Storr. Arlington, VA: Mercatus Center: 103-90.
- Storr, Virgil Henry. (2013). *Understanding the Culture of Markets*. New York: Routledge.
- Stuntz, William J. (2000). "Self-Defeating Crimes." *Virginia Law Review*, vol. 86 (Symposium: The Legal Construction of Norms): 1871-1899.
- Thrasher, John and Gerald Gaus. (forthcoming). "Calculus of Consent." In the *Oxford Handbook of Classics in Contemporary Political Theory*, ed. Jacob Levy. Oxford: Oxford University Press.
- Tognato, Carlo, ed. (2017). *Cultural Agents Reloaded: The Legacy of Antanas Mockus*. Cambridge, MA: The Cultural Agents Initiative at Harvard University.
- Tyler, Tom R. (1990). *Why People Obey the Law*. New Haven: Yale University Press.
- Vanderschraaf, Peter. (2006). "War or Peace? A Dynamical Analysis of Anarchy," *Economics and Philosophy*, vol. 22: 243–79.
- . (2007). "Covenants and Reputations." *Synthese*, vol. 157: 167–95.
- Young, Iris Marion. (2011). *Responsibility for Justice*. Oxford: Oxford University Press.