

# Justification, Choice, and Promise: Three Devices of the Consent Tradition in a Diverse Society

*Gerald Gaus*

## 1. CONSENT IN THE SOCIAL CONTRACT TRADITION

The twin ideas at the heart of the social contract tradition are that persons are naturally free and equal, and that genuine political obligations must in some way be based on the consent of those obligated. In the *Second Treatise* Locke rejects the claim of Filmer — and the entire conception of the Great Chain of Being on which the medieval order was founded — that some people have natural authority over others, and thus those over whom authority is exercised have a natural duty to obey. “Men being...by Nature, all free and equal and independent, no one can be put out of this estate and subjected to the Political Power of another, without his own *Consent*” (Locke, 1960: §95. Emphasis in original). Almost right from the beginning the ideas that persons are free and equal, and that political authority must be based on the consent of the governed, went hand-in-hand.

Over the last thirty years A. John Simmons has brought home a simple but fundamental truth: while consent can indeed reconcile our status as free and equal with submission to political authority, the undeniable fact is that few actually consent to political authority. If we understand the “legitimacy” of the state as “its exclusive right to impose new duties on subjects by initiating legally binding directives, to have those directives obeyed,” and thus generating corresponding obligations on its citizens, and if such legitimacy requires that people have actually consented to such a state, it turns out that there are no states that possess general legitimacy *vis a vis* their citizens. (Simmons, 2001: 137, 156). This gives rise to a certain sort of “philosophical anarchist” position according to which, while it is conceptually possible that there be such a thing as a genuinely authoritative state that imposes obligations on its citizens, as a matter of fact no present state generally meets the necessary condition — general consent has not been obtained — and so in that respect all states are illegitimate (Simmons, 2001: chap. 6). Such philosophical anarchists maintain that citizens should “act morally” towards each other but not accept moral obligations to obey the state (Simmons, 2001: 153).

Simmons (2001: 140) acknowledges that Kantian-inspired views do not lead to this result. The Kantian stresses that that since each person has good reason to endorse the existence of states, because they are “necessary for the realization of freedom and rights and justice,” each

has an obligation “enter civil society and accept the duties society imposes.” Thus, for a Kantian such as Thomas Nagel (1991: 33), “[t]he task of discovering the conditions of legitimacy is traditionally conceived as that of finding a way to justify a political system to everyone required to live under it.” There is a type of rational consent here, since the justification is owed to each person, appealing to her ends and values. Thus, on a typical contemporary Kantian view, justification requires the hypothetical consent of all idealized persons, who endorse principles and institutions that conform to their own reasons. Simmons, in reply, insists that Kantians conflate the justification of the state (are states necessary for justice?) with their legitimacy (are we obligated to obey the laws of our own state?). Even if each could be shown that there are good reasons to endorse the existence of states, and even the existence of their own state, it would not follow that they are bound to acknowledge that the state has a moral power over them to create new obligations. Simmons (1991: 147) concludes that “[a]ppeals to hypothetical choice, acceptability, or reasonable nonrejectability have a very different moral basis and force than do appeals to actual choice....Even appeals to what ought to be chosen in the light of the individual’s own interests and values are quite different from appeals to that individual’s actual choices.” Justification via the consent of idealized agents who respond only to their reasons to endorse or reject is very different from legitimacy based on real people’s actual choices.

In this essay I seek to bring the Kantian and Lockean contract traditions together. Kantian rational justification and actual choice are complementary devices through which our freedom and equality can be reconciled with moral and political authority. We should not think that there is simply one way by which relations of moral and political authority can be reconciled with our status as free and equal. I defend three distinct devices through which freedom and authority may be reconciled: justification to others, social choice, and promise. All three are aspects of the “consent tradition” broadly construed. As all-too-often occurs, philosophers have succumbed to endless disputes about *which one* must ground authority and its power to obligate when, we shall see, all three are important ways in which a free and equal person can live under the authority of others.

## 2. MORAL AUTHORITY AMONG FREE AND EQUAL PERSONS: THE ROLE OF JUSTIFICATION

### 2.1 *The necessity of authority: the instrumental and constitutive interpretations*

As Simmons notes, according to Kant the source of political obligation is our duty to enter civil society.

Although experience teaches us that men live in violence and are prone to fight one another before the advent of external compulsive legislation, it is not experience that makes public lawful coercion necessary. The necessity of public lawful coercion does not rest on a fact, but on an a priori Idea of reason, for, even if men to be ever so good natured and righteous before a public lawful state of society is established, individual men, nations and states can never be certain they are secure against violence from one another because each will have the right to do what *seems just and good to him*, entirely independently of the opinion of others (Kant, 1999: 116, emphasis added).

Kant goes on to insist that justice is absent in the state of nature because each relies on his own judgment, and thus “when there is a controversy concerning rights (*jus controversum*), no competent judge can be found to render a decision having the force of law” (1999: 116). Thus Kant argues that individuals are obligated to abandon the “state of nature” and enter into a “juridical state of affairs” (1999: 114-5).

The claim that there cannot be justice among free and equal persons without a state is open to two interpretations: the instrumental and constitutive. Locke clearly presents an *instrumentalist* view of the injustice that arises in the state of nature when each acts on his own private judgment about justice. According to Locke (1960: §87), in the state of nature, self-interest, passion, and desire for revenge lead people to misapply the law of nature, and over-punish perceived violations against them (1960: §§124-26). Consequently, to secure a condition in which justice obtains, “all private judgment of every particular Member” must be “excluded, [and] the Community” must live under an “Umpire, by settled standing Rules, indifferent, and the same to all Parties....” A system of political authority, we might say, provides an impartial system of justice, and so overall a more just order. If this is how the authority of the state provides for the justice lacking in the state of nature, Simmons’ sharp distinction between justifying the state, and submitting to the authority of the state, makes perfect sense. To justify the state is to show that it performs the task of providing an impartial umpire, and so an effective system of justice. Realizing that, one can think the state that does its job is a good thing and be glad when we have such a state. It would not follow, however, that one should submit to its authority in the sense of acknowledging an obligation to do as it instructs (just because it instructs it).

Kant points to a much deeper sense in which justice is absent from the state of nature (an “a priori idea”), one which leads us back to the problem of authority among free and equal moral persons. Stephen Darwall (2006: 4) has recently stressed the way in which interpersonal moral

claims involve “authority relations that an addresser takes to hold between him and his addressee.” When I make a moral claim on you not to  $\phi$ , I am not making a request that you refrain from  $\phi$ , or calling attention to my opinion that  $\phi$  is immoral: I am issuing an imperative that you must not  $\phi$ . If I decide that your action falls under the purview of social morality, I see it as my business that you refrain from  $\phi$ -ing (Baier, 1965: xviii-xix). I claim authority over you in the sense of claiming a *standing* to direct your action. And I suppose that you are under obligation to comply (Gilbert, 2006: 103ff, 147ff, 245ff).

And here we confront the core problem: how can one have the standing to command the performances of others while treating them as free and equal moral persons? You are claiming that they are obligated to *obey you*. Recall that for Locke, fundamental to our natural freedom and equality is that “the *Natural Liberty* of Man is to be free from any Superior Power on Earth, and not to be under the Will or Legislative Authority of Man, but to have only the Law of Nature for his Rule” (1960: §22). To be morally free is to have *only* the law of nature as one’s rule. This implies, though, that one is not also ruled by the private judgment of others as to the demands of the law of nature. As a morally free person one employs one’s reason to understand the requirements of the law of nature, and one submits to rule by that law, which is the rule of reason. Moral freedom is not freedom from morality, but freedom to directly employ your (private) reason to determine what morality requires. Such freedom appears straightforwardly at odds with the claims by others to authority over you, which is fundamental to the practice of interpersonal moral authority.

Unless there is an authority endorsed by public reason — the reason of all — a moral claim is simply the attempt of one person’s private judgment to rule over others. For the constitutive view, unless there is an authority that expresses a public reason — a judgment that you and I both share — the practice of interpersonal morality is inconsistent with our fundamental status as free and equal. Notice that on the constitutive view, philosophical anarchism leads to either moral nihilism or moral authoritarianism. If you say that public authority is well and fine, but you see no reason why one should submit your judgment to public reason, you are no longer able to proceed with moral demands that treat others equally. You seem to have two options. First, you might forgo moral claims entirely. So far as interpersonal morality is concerned, this leads to a sort of nihilism: you neither make nor recognize interpersonal moral claims. To be sure, one can still have personal moral ideals, such as “I will respect others” but such purely first-personal ideals do not constitute interpersonal moral claims — claims that others must do as you demand. The second option is to insist that these purely first-personal ideals based on one’s private reason do indeed have moral authority over one’s fellows, thus leading to

authoritarianism. Because the philosophical anarchist does not appreciate that all moral claims are exercises of authority over others, his effort to reject submission to authority while keeping morality is ultimately incoherent.

## 2.2 *The public authority of justified moral norms*

The social contract tradition makes a very quick move from the conflict of private judgment about morality to the need for an umpire with political authority to judge all disputes about right according to an authorized public reason, to which all submit (Gaus, 2011c). However, as T.H. Green (1986: §113) pointed out in his review of the social contract tradition, Hobbes, Locke and Rousseau overlooked the fact that between individual judgments about individual rights, and the determinations of public political authority, lies the social authority of practices, norms, and conventions. We might think of this as taking seriously Kant's idea that the solution to the conflict of private judgment is to enter into *civil society* — while we need to avoid a condition in which each does what “seems just and good to him, entirely independently of the opinion of others,” this need not immediately drive us into *political society*. When a moral rule  $R$  is a social norm it is commonly recognized as a rule to be followed within the group and in fact there is general conformity to it (Gaus, 2011a: §10). It is common knowledge that it is the rule of our group: each knows that it is our rule, and knows that others know it to be.

Now when a moral rule that is accepted as a social norm is also endorsed by the private reason of each free and equal moral person — each moral person's deliberative rationality — moral authority is reconciled with our freedom and equality. Suppose rule  $R$  is endorsed by the members of a social group  $G$  in the sense that  $G^*$  — somewhat idealized members of  $G$  who deliberate solely on their relevant values, ends, and personal moral intuitions — all have sufficient reason to endorse  $R$  over all the alternatives. It is true that what is endorsed by  $G^*$  will not be the same as what is actually endorsed by, or would if asked be actually endorsed by,  $G$ . Different theories of public reason relate  $G$  and  $G^*$  different ways, but in all plausible versions a member of  $G^*$  deliberates on the reasons to endorse a moral norm that her counterpart in  $G$  has, but to which her actual counterpart does not always pay attention. This is, I think, a far less controversial idea than Lockean such as Simmons would have us believe. We are a complex combination of selfish and moral creatures: the moral system, we might say, has developed on top of an earlier selfish set of motivations (Richerson and Boyd, 2008: 114; Freidman, 2008: chap. 1). Morality is required just because our will does not always align with our reasons to restrain our will and abide by social rules. As a rational moral agent, to know whether a rule is one to which my private reason accords moral authority, a test which consults my actual present will

("Do you now choose the rule?") is inherently flawed, for it supposes that our moral commitments and reasons will be dominant in our thinking (rather than, say, narrowly self-interested reasons that induce us to bargain and bluff).

If  $R$  is a social rule that is endorsed by the relevant reasons of all, when I demand that you conform to  $R$  I am not merely employing my private reason to instruct what you must do: I am appealing to a commonly recognized rule that our reason endorses as a rule to live by. In familiar Kantian terms, the rule expresses legislation in the commonwealth of ends, in which the law comes from all and applies to all. In addition, when  $R$  is a social rule, it is common knowledge among us that  $R$  is the rule to which all submit. Because the authority of the rule is self-imposed, the claim to authority over others is reconciled with the status of all as free and equal.

### 3. MORAL AUTHORITY AMONG FREE AND EQUAL PERSONS: THE ROLE OF CHOICE

#### *3.1 The challenge of diversity for reconciliation through justification*

Justification as a device of reconciling moral authority with recognition of others as free and equal falls short of a fully adequate reconciliation. In a diverse social world, in which individuals have a plethora of values, ends, and moral ideals, it is most unlikely that any candidate social rule will be endorsed by all members of  $G^*$  as the uniquely best. On any plausible account of the deliberations of members of  $G^*$ , they will endorse very different rankings of candidate moral rules. The only way to induce unanimity is to greatly, indeed radically, idealize the members of  $G^*$  so that they essentially evaluate candidate moral rules in the same way, as does Rawls' argument from the original position. The veil of ignorance excludes "knowledge of those contingencies which set men apart..." (1999: 17). Rawls then attributes to the parties a concern with primary goods that provides a basis for their common deliberation. Insofar as we consider ourselves as agents devoted to some ends, they are what we need. When abstracted to the common status of agents devoted to their own (unknown) evaluative standards (values, comprehensive conceptions of the good and so on), because "everyone is equally rational and similarly situated, each is convinced by the same arguments" (1999: 120). So although the original position begins by posing a problem of collective choice, the problem is reduced to a choice by one person, and so there is no disagreement (1999: 120-1). Although a degree of idealization that allows us to model a person's deliberation about her relevant values, ends, and moral ideals is appropriate, an extreme idealization that reduces the

deliberations of members of a diverse society to that of a representative, single, abstract, person ignores our real problem: the justification of authority under conditions of wide disagreement.<sup>1</sup>

### 3.2 *The eligible set*

In a diverse society, then, self-legislation inevitably leads to disagreement. If we take this problem seriously, we must revise the typical Kantian-inspired original justification situation. Instead of constructing a deliberative justificatory model in which all members of  $G^*$  concur on the same rule, we should suppose that the outcome of any plausible deliberative model among members of  $G^*$  will yield a set of proposed moral rules.

Let us assume that a member of  $G^*$ ,  $G_i^*$ , proposes the rule  $R_i$ , regulating some area of social life that, given her values, ends, and moral ideals, is optimal. If she were the moral dictator — if social morality conformed perfectly to her private reason (as many moral philosophers seem to suppose it should) — this would be the rule that all would endorse as authoritative. The problem is that under any remotely realistic level of idealization, other members of  $G^*$  are almost certain to disagree that this is the optimal rule given *their* values, ends, and moral ideals. Other members of  $G^*$  must consider how well  $R_i$  stacks up against other proposals. To simplify, let us suppose that each member of  $G^*$  is able to order all proposals based on her relevant values, ends, and moral ideals. Having constructed such an ordering she has evaluated the proposals from best to worse in relation to how well they advance or express her values, ends, and moral ideals.

Not all the proposals, however, will qualify as eligible moral rules. *First*, following Kurt Baier, R.M. Hare, Rawls and many others, we must suppose that there are some formal constraints on what can constitute a *bona fide* moral rule. An eligible rule must have a certain generality, the rule must be public at least in the sense that its content can be taught to new members of the group, it must be determinate enough to resolve conflicts within the group about what is the correct action, it must generally be accepted that its requirements override personal aims and inclinations, at a minimum the rule must not be viewed as hostile to the basic good or well-being of any member of the group, and a person must be able to endorse the rule whatever role such occupies under it (Gaus, 2011a: §15.2). *Secondly*, it is important to recall that to accept a rule as authoritative entails that others can *demand* that you comply regardless of your personal aims and inclinations, appropriately rebuke and punish you for not complying, and that you appropriately feel guilty for failing to comply. These are real costs of moralization: one allows that one's actions become the business of the public, and that you are to conform your actions to public norms regardless of your own proximate aims, and are answerable to

others for failing to do so. Now a member of  $G^*$  will not only deem ineligible any rule that fails to conform to the formal constraints on moral rules, but as she reviews the proposed rules in terms of her reasons to endorse them, she must take into account the costs of moralizing an activity. At some point she will almost certainly decide that, given the extent to which some  $R_i$  advances her overall values, ends and moral ideals, these benefits are exceeded by the costs of moralization. She will thus deem any such rule ineligible as a genuinely morally authoritative rule that she can endorse.

Each member of  $G^*$  will, then, divide her ranking of proposals into an eligible and an ineligible set. Let us call  $S$  the socially eligible set — the set of all proposed rules that are not in the ineligible set of any member of  $G^*$ . If  $S$  is null there is no eligible rule on this matter: that would be simply to say that no rule  $R_i$  conforms to the formal principles of right and for each member of  $G^*$  the reasons to endorse are outweighed by the costs of moralization. Of course any Kantian theory must worry that this will generally be the result of moral justification — that would entail the dismal result that there is no such thing as a publicly justifiable moral rule. Kantian theory has, in my view plausibly, argued against this dismal result (where all we are left with is the choice between moral authoritarianism and nihilism), but then has pushed on to a much more controversial claim:  $S$  is a singleton. Given evaluative diversity this is most unlikely. A plausible Kantian view under conditions of wide-ranging diversity will be left with a socially eligible set  $S$  that is neither null nor a singleton. No social order has been able to exist without moral rules covering, for example, basic rights of the person, harm to others, and property rights. Effective moral rules on these matters are a great good, and this would be recognized by an even moderately idealized  $G^*$ . But there is no reason to think that there is only one acceptable moral rule to regulate each of these matters.

### *3.3 Moral equilibrium*

Can a group of free and equal people come to accept a common rule, even though plausible justificatory models are indeterminate? Suppose that members of  $G^*$  take the results of the justificatory argument, and provisionally assume that it is the limit of the morality dictated by impartial reason. Morality, we might say, is indeterminate, specifying only a set  $S$ . Insofar as justification has had an outcome, it is  $S$ : what is common to  $S$  defines the set of actions that are certainly morally prohibited, required, or permitted. Beyond that, each is free to act on the moral rule within  $S$  she chooses. Now each person has two concerns: (i) to induce the group to come to adopt his preferred rule and (ii) to come to act on a common moral rule with others. The first concern is obvious; a moral agent has good reason to induce others to adopt the rule

that, given his private reason, he thinks best. In addition, though, moral persons are concerned with converging with others on a common moral rule. Until there is a shared common rule, known by all to be the rule of the group, there cannot a real moral life that respects all as free and equal. Should all come to accept the same eligible rule, then, finally, moral authority is reconciled with freedom and equality.

Two members of  $G^*$  can be modeled as playing an impure coordination game. Display 1 gives the game in its simplest form, a simple 2x2 game (numbers indicate ordinal utility, highest being most preferred).<sup>2</sup>

		<b>Betty</b>	
		$R_A$ in $S$	$R_B$ in $S$
<b>Alf</b>	$R_A$ in $S$	1 2	0 0
	$R_B$ in $S$	0 0	2 1

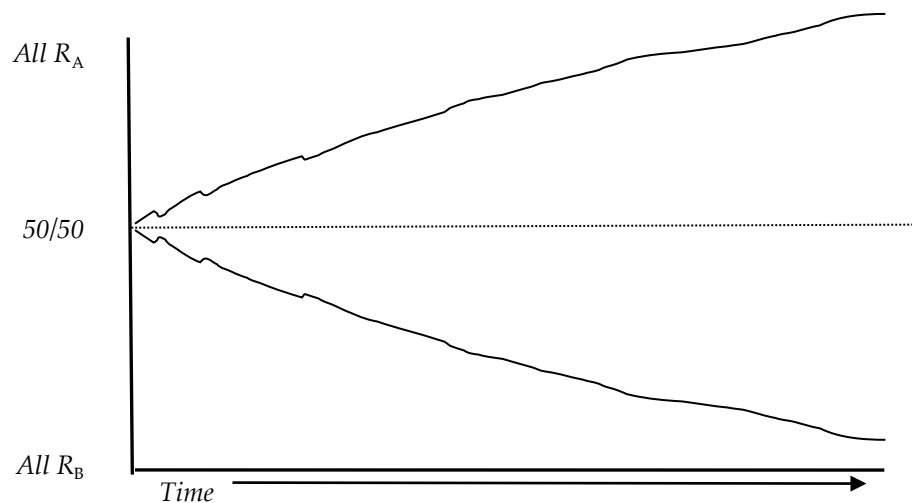
Display 1

Should Alf and Betty find themselves coordinating on  $R_A$  neither would have reason to change his or her action. For each, given his or her overall evaluative standards, each has the most reason to act on  $R_A$ . Should they instead find themselves at  $R_B$ , each will then have most reason (given his or her evaluative standards) to act on  $R_B$ . Note that in neither case is either induced by some external consideration to conform to a rule that is not, from his or her perspective, optimal: *consulting simply his or her own evaluative standards, each has decisive reason to freely endorse whichever eligible moral rule on which they have coordinated.* When coordinating on  $R_A$  Alf can demand that Betty conform and, consulting only her own values, ends, and moral ideals, she will have decisive reason to conform; and if they have coordinated on  $R_B$  Betty can demand that Alf conform, and he will have decisive reason to act on it, considering only what he thinks his important — his own “utility function.” And this even though, from the initial deliberative perspective, neither had reason to act on the other’s preferred moral rule.<sup>3</sup>

A one-shot two-person game can give us some insight, but it is clearly an inadequate way to model the selection of a moral rule from the eligible set. The relevant coordination problem is not a single-play game, but an iterated game. We have a number of encounters with others, and each can be understood as a play in a series of impure coordination games over many options. Now in an iterated game a person’s utility (this is defined solely in terms of her relevant values,

ends, and moral ideals) is a combination of her utility in this play, plus her expectations for utility in future games. Thus a person might sacrifice utility in one play to induce play in future moves that will yield her a more favored result. Now in large iterated games a bandwagon effect manifests itself. As I have argued elsewhere, such large-person iterated coordination games exhibit a strong increasing returns effect: the more people come to embrace a particular rule, the more reason others have to also embrace it (Gaus, 2011b).<sup>4</sup> Even if one ranks a rule near the bottom of  $S$ , under a wide range of conditions one will have strong reason to embrace it if most others are acting on it.

This dynamic is illustrated in Display 2. Starting out with a population evenly split between advocates of  $R_A$  and of  $R_B$ , individual choices and random events can lead the population to all  $R_A$  or all  $R_B$  equilibria. Which equilibrium emerges will be path-dependent: at time zero there is no reason why one or the other should emerge as the *unanimously-chosen rule*. But once they have arrived at such a convergence, each member of  $G^*$ , consulting only her own values, ends, and moral ideals will freely act on the rule (which will be a Nash equilibrium). For our purposes what is crucial is that the contingent way in which large groups can come to coordinate on a common practice is no bar to there being a determinate morality that all can endorse given their evaluative criteria *once it has been arrived at*.

Display 2<sup>5</sup>

### 3.4 From justification to choice (and back again)

This very abstract model points to an important insight: a specific authoritative social morality that treats all as free and equal is to a large extent the result of the choices of its members. It is

not *merely* chosen: not just any rule that a group might converge on is a genuinely moral rule. Bad norms, backed by punishment, can be in equilibrium: each may have sufficient reasons to conform simply in order to avoid punishment (Bicchieri, 2006: 42). Punishment can stabilize just about any norm, good or bad (Boyd and Richerson, 2005: chap. 9). Social equilibrium does not imply moral equilibrium, and the fact that a bad norm is in social equilibrium can make it very difficult to dislodge. However, when a rule in equilibrium is within the socially eligible set, it is both chosen by the members of a society and is acceptable from what Baier (1965) called “the moral point of view.” The rule passes the tests of impartiality, generality, teachability, and so on, and is endorsed by the values, ends, and moral ideals of all. Each, consulting only his own values, ends, and moral ideals has reason to act on it.

A moral rule in equilibrium is not a collective “we” choice; the group, *as a group*, does not choose its moral rules (Cf. Gilbert, 2006: Part II). Nevertheless, a moral rule in equilibrium is a social fact that arises from the interdependent choices of its members. As interdependent individuals, we choose our moral rules — at least, if we are lucky enough to actually possess genuine moral rules (§4). This is *a real choice of actual people*; not a hypothetical choice, or one dictated by impartial reason. To be sure, it is the choice of social creatures, who must take into account what others are doing, and cannot dictate to the group their preferred rule. Importantly, although the rule is chosen by members of society, taken individually in their interdependent interactions, and unless some rule was so chosen by the society there would be no effective morality, the authority of the morality does not depend on each and every person choosing the rule. Once society has chosen a rule, if the rule in equilibrium is also a member of the socially eligible set, *we have created through our actual interdependent choices what impartial reason could not deliver: a uniquely justified rule*. For once the moral rule is in social equilibrium (and is a recognized social norm), then all have conclusive moral reason to act on this rule rather than any other in *S*. Thus, having created a justified rule through our interdependent choices, we can then insist that all conform to this rule, for all free and equal persons now have conclusive reason to conform to this rule, rather than any other. That genuine moral authority is created by our choices does not entail that one who holds out, refusing to choose our rule, escapes the moral authority of our justified social-moral rule. In this fundamental sense, society consents to the moral rule to which we are all bound.<sup>6</sup>

Filmer long ago pointed out that, the “plausible and dangerous” notion that all are free and equal appears to require that each and every person consents to authority, but it is impossible to imagine that all do so. And, Filmer insisted, it is not enough to say that the majority, even the overwhelming majority, have consented: “unless it can be proved by some law of nature that

the major, or some other part, have power to overrule the rest of the multitude, it must follow that the acts of multitudes not entire are not binding to all but only such as consent unto them” (1991: 21). This has always been the Achilles’ Heel of consent theory (see Gilbert, 2005: chap. 5). On the account advanced here, the consent of the majority — their actual choices in selecting a rule from  $S$  — has the power to bind the rest. Once society has consented to this rule, it becomes the sole rule that is uniquely rationally justified: it now is the one rule that all have reason to endorse. It is the only rule that all can endorse and which is able to provide the basis of shared moral life in which all are treated as free and equal. Each, consulting only her set of relevant values, ends, and moral ideals, has reason to act on that rule over any alternative. What was *ex ante* indeterminate, is *ex post* uniquely justified. In game theoretic terms, in the *ex ante* situation there are many possible Nash equilibria (in fact, every member of  $S$ ); once a rule has been selected by society, it is in Nash equilibrium, and so all have reason to act on it.

#### 4. POLITICAL AUTHORITY, JUSTIFICATION, AND CHOICE

##### *4.1 The state as the protector, sustainer, and developer of social morality*

T. H. Green (1986: §134) insisted that “a state presupposes other forms of community, with the rights that arise out of them, and only exists as sustaining, securing, and completing them.” For Green, the great error of social contract theory was that it supposes that individuals somehow confront each other simply with their individual rights before the advent of the state, where in fact all states arise out of a pre-existing system of social morality and authority, and at least in the first instance their task is to protect, sustain, and develop this system of morality. Indeed, unless there is a definite pre-existing social morality, there is no reason to think that government can solve our problems. Until we can be confident that those with political authority will themselves abide by the norms of social morality, instituting political authority may simply leave us open to new forms of exploitation — those with authority may simply use the institutions to promote their narrow self-interest (Schwab and Ostrom, 2008) or impose their personal moral ideals. Once we understand that interpersonal morality already supposes a system of social authority, the question is not whether free and equal individuals choose to be subjected to the authority of others, but whether, while endorsing social moral authority, they pull back and refuse to endorse political authority.

Although on many matters a society of free and equal individuals can coordinate on a common interpretation of a moral rule (and so, a rule specifying their rights), social

coordination fails to fully secure a common authoritative social morality among free and equal persons for a number of reasons (Martin, 1993: 166).

(i) Given social and technological changes, new cases may arise for which the social rule has no clear answer. Although we may expect that over a long enough time a new social equilibrium on this matter may arise, this may entail a long period of disagreement and moral disorder.

(ii) The rules of social morality can conflict. Again, although some rules of priority are no doubt part of social morality (we usually know what to do when a promise conflicts with saving a life), many of these conflicts may be unresolved by social morality, and so a more formal system of adjudication may be required.

(iii) The informal punishments of social morality may be insufficient to enforce moral rules when the temptations to cheat are high.

(iv) Our social order may be stuck at a social equilibrium outside the socially eligible set — that is, we may have an oppressive social morality that some free and equal moral persons do not have reason to endorse and abide by.

An agency that protects, sustains, and develops our social morality is justified by our very devotion to social morality conjoined with the knowledge that problems (i)-(iv) are endemic features of life — at least in large-scale modern societies.

Recall that Simmons is ready to accept the Kantian claim that states are justified: what he challenges is that this type of justification implies a political obligation to obey (§1). Note, though, that when states are justified as harmonizers and developers of social morality, this justification implies that the judgments of the state articulate social morality, which is authoritative and obligatory. Insofar as the task of the state is to, as it were, continue on with the development of social morality by political means, when it fulfills these tasks it imposes obligations via the authority of social morality. The social contract tradition was wrong in holding that a political umpire or adjudicator is always a necessary, constitutive, element of justice among individuals (§2.1), but it was correct that in many instances it is necessary. And this necessity is not simply an instrumental necessity: the state's actions as developer of social morality is constitutive of some of our moral relations with others (§2.1). Social morality both constrains the state (where we have a social equilibrium on a member of the eligible set, the demands of morality are clear) and yet it is also completed by the state, when it provides the means for adopting a member of the eligible set.

#### 4.2 *The role of social consent*

The state is morally justified if (a) it is required to protect, sustain, and develop social morality. We also should add that for any particular state to be justified it must (b) be within the socially eligible set of all institutional structures that fulfill the role specified in (a). For a state *T* to have moral authority qua sustainer and developer of social morality, both conditions must be met. Again, we have arrived at an eligible set that is not a singleton. Members of  $G^*$  would almost certainly be unable to settle on a uniquely best regime. Although some form of democracy, understood as a representative system with rights of participation and political pluralism, is surely abstractly justified, members of  $G^*$  will not concur on which system is best. Democracies differ in fundamental ways: whether they are majoritarian or consensual, unicameral or bicameral, whether they are proportional representation, are parliamentary or presidential, have a unitary or federal structure, possess a written or an unwritten constitution, have a strong or weak system of judicial review — just to name some of the more obvious factors. Members of  $G^*$  will have reasonable differences about how to rank different democratic regimes. Whether we take these issues as a bundle (members of  $G^*$  choose regimes types), or divide them up into a series of choices, members of  $G^*$  will order the options differently.

As with social morality itself, different societies arrive at different equilibria on these matters, largely on the basis of a path-dependent political history. Just as the justification of the rules of social morality from the impartial deliberative perspective of the commonwealth of ends ( $G^*$ ) is insufficient to yield a determinate publicly justified social morality, so too is the abstract justification of democracy and its attendant political rights insufficient to yield the justification of a system of governance. Public reason does not mandate a specific democratic regime (how could it?). To be sure, those committed to a democracy that secures publicly justified outcomes may well advance proposals for institutional design that, in their view, do a better job than other arrangements, but rational and good-willed individuals will disagree about the merits of these proposals. Political authority too relies on informal social authority — an evolution of a political culture leading to the selection of one of a wide range of acceptable political systems.

Understood as an institutional system that is required to protect, sustain, and develop social morality (and its rights), the authority of an existing, justified state, and the corresponding obligation of its citizens to obey, is indeed based on the consent of the governed. For the state thus understood is itself a development of social morality, and so like all social morality it is the result of numerous individual choices over a long process that leads to a the specific social

equilibrium. Again, this is not a collective “we choice” that we make together, or a one-time social contract or constitutional convention. It is an ongoing social choice, arrived at by a path-dependent history, and continually reaffirmed by the choices of its members. Filmer (1991: 142) was correct: “Mankind is like a sea, ever ebbing or flowing, every minute one is born and another dies. Those that are the people this minute, are not the people the next minute. In every instant and point of time there is variation.” But this does not preclude the important idea that political authority rests on the ongoing consent of the governed.

##### 5. THE ARBITRARY WILL AND THE POWER OF PROMISING

The social contract’s reliance on consent seems to invoke consent qua promising: one is bound to obey the terms of the contract (to obey the law) because one has promised to do so. I have been at pains to show that consent need not be in the form of a promise. The members of the socially eligible set are *justified* authoritative rules or institutions; a society exercises real *choice* in deciding on which of these they shall equilibrate. No promise, explicit or tacit, is involved. But can promises truly bind, and could they form the basis of political obligation?

Some doubt that actual consent, even if it could be obtained, would yield true political obligations (Gilbert, 2006: 75-83; Estlund, 2008: 49-50). The actual consent view seems to allow that one can be bound just because no one has yet thought of the decisive problem with an institution; once it occurs to someone and she withdraws her consent, then the institution immediately is de-legitimized. It would seem that actual consent cannot provide the basis for assured continuing obligations — there is, we might say, justificatory instability. Now relying on consent qua promising avoids this problem: to promise is not just to agree for now, but to bind oneself. Even if it later occurs to you after the promise that perhaps it was not such a great idea to promise, you are, generally, nevertheless bound. On Hobbes’ (1994: 85) view of promising and contract, unless a crucial new piece of information becomes available between the time of promising and the time of performance, one is bound to perform, regardless of second thoughts.

This response to the problem of justificatory instability seems simply to lead to a deeper problem, stressed by Hegel (1952: 157) who charged that Rousseau’s general will, being based on actual and not rational wills, therefore founds the state on citizens’ “arbitrary wills,” and “and their capriciously given consent.” Hegel is entirely right that a fundamental aspect of promissory obligations is our moral power to bind our future action to our present “arbitrary wills.” Many find this power mysterious, and seek to base promising on more fundamental, and

non-arbitrary, moral principles (e.g., Scanlon, 1998: chap. 7). I believe this is a mistake. One of the most basic features of moral agency, especially in a diverse society in which people fundamentally disagree about values, ends, and moral ideals, is the ability of individuals to, as it were, extend the bounds of morality through their individual choices and agreements. A framework of common moral rules is necessary for ordered moral relations, but insufficient. We not only need to invoke the common authority of social morality in our relations with strangers, but we need specialized moral claims with respect to those with whom we share ends and ideals. One way — though not the only way — to extend moral obligations is through our distinctive moral power as agents to create moral obligations through promising. Just as a distinctive legal power of a legislature is to create obligations, a distinctive self-legislative power of a free and equal moral person is to create obligations on herself. This power follows from our very status of self-legislative agents. In terms of the categories I have employed, it is not our legislative power qua members of  $G^*$  (in Hegel's terms, our rational wills), but our legislative power simply as members of  $G$  (our actual, arbitrary, wills) that legislate promise-based morality.

An effective social morality needs to respect the self-legislation of arbitrary, actual, wills. We can say some things about what members of  $G^*$  would endorse, but this is, I have argued, surprising modest. Often it is simply unclear what a person has reasons to endorse — what his  $G^*$  counterpart would agree to. A person's reasons are complex, and his deliberations are often inconclusive. As an actual agent in the world, he must often act on his own, often incomplete and imperfect, deliberations about what he has reason to do. Fundamental to being an actual agent is to have the authority to decide in these sorts of cases what one has reason to do. If I had deliberated on your beliefs and values, I would have decided that  $\phi$  is the thing to do, but you deliberate and decide that  $\gamma$  is. This is not simply a disagreement, in which we both have competing claims. That you have decided  $\gamma$  rather than  $\phi$  is normally decisive: to be an agent is to be guided by your deliberations about what you should do, even if from an impartial epistemic point of view, my decision is superior. In Hegelian terms, actual agents are typically guided by their actual, even if arbitrary, wills. Because a purely rational morality is impossible (or, at least, we have no idea what it would be), real moralities must respect our arbitrary wills and, indeed, must allow that our moral self-legislation is often based on them. Our power to bind by promising is perhaps the most important way this is accomplished in our social morality.

The problem, then, is not that arbitrary wills cannot morally bind themselves. The real difficulties for the promissory view of political obligations are two. The first is recognized by all:

it is simply implausible to think that many have bound themselves in this way to any state, so obligation through promise is never a plausible ground for a case that there is an obligation of almost all citizens to obey the law, just because it is the law. However, even if that obvious — and as far as I can see insurmountable — difficulty could be overcome, an account of political obligation based solely on promissory obligations would remain objectionable. Hegel was basically correct: to base obligations *simply* on the arbitrary wills of a group of people is insufficient to assure us that our basic framework is not unjust and irrational. If all happen agree we may leave well enough alone, but the worry lingers: is the entire framework based on errors, superstitions, and the effects of social indoctrination? Unless we are convinced that the foundations of the framework are set in good reasons that all can endorse, the worry gnaws that our moral and political lives are nothing but an arbitrary, perhaps oppressive, social equilibrium.

#### 6. CONCLUSION: CONSENT THEORY AS DEVICES OF RECONCILIATION

I have stressed in this essay how the authority of social morality and political obligation are the result of a complex interplay of our rational and actual wills, or what we might think of as hypothetical and actual consent. Arguments from the hypothetical consent focus on an idealized group  $G^*$ , seeking to showing what authoritative rules and institutions all members of  $G$  have reasons to endorse. As Kant stressed, until we have common rules that reflect public reason there simply is no public justice: we have only different people employing their private reason, each asserting what he thinks justice is. But, as I have stressed, to respect others as free and equal is inconsistent with claiming such authority of one's private reason over them. Only rules endorsed by public reason can reconcile authority with our status as free and equal. However the reasoning from this perspective (that of  $G^*$ ) is indeterminate: in a diverse society no single moral rule over any area of social life is endorsed by public reason. Here, I have argued social morality employs a second device of reconciliation: a social choice of one member of the eligible set on which we coordinate. This is a real choice. In Hegelian terms, we might say that it is based on partly on our rational, and partly on our joint arbitrary wills. Insofar as we are rational, we restrict our choice to the socially eligible set; but within that set a society chooses via the interaction of its members' arbitrary wills, producing a social equilibrium. Lastly, I have argued, the device of promising grants each free and equal person the moral power to employ her arbitrary will to grant others moral authority over her. It allows self-

legislation by actual free and equal moral persons, based on their full set of ideals, values, and ends.

The consent tradition and liberalism developed together. Liberalism's devotion to the moral freedom and equality of all can only be reconciled with moral and political authority through forms of consent — all of which are exercises of self-legislation, some rational, some quite arbitrary. The problem has been the insistence that one form of consent is the unique device of reconciliation. Different aspects of the consent tradition perform some tasks of reconciliation well, but all are implausible when pressed as the sole method of reconciliation. Once we appreciate that "consent theory" is not a foundational commitment, but an array of devices for reconciling our status as free and equal with relations of authority, we can stop battling about which is correct, and investigate what tasks each can perform.

*Philosophy*  
*University of Arizona*

## NOTES

- <sup>1</sup> We should not think that even Rawls holds that the justification of principles of justice can simply ignore the issues about which we disagree. Although the free-standing argument from the original position is based simply on shared models of the person, Rawls (1996: 386) insists that this is a "pro tanto" or "as far as it goes" justification. What he calls "full justification" requires that the principles of justice be evaluated in light of one's entire conception of value.
- <sup>2</sup> It is important to stress that "utility" is a mathematical representation of how an agent's judgment as to how well a rule satisfies her values, aims, and moral ideals. It is not either itself a unique value (such as welfare), nor it is a goal to be pursued.
- <sup>3</sup> Again, we should not be misled by the language of "preference." To prefer  $R_A$  to  $R_B$  is simply to rank  $R_A$  over  $R_B$  for purposes of choice; in our terms one's evaluative standards indicate reason to rank  $R_A$  over  $R_B$  — this is all that is implied by saying one has a preference for  $R_A$  over  $R_B$ .

- <sup>4</sup> The path-breaking work on increasing returns was done by W. Brian Arthur (1994). For the technically-minded, this convergence can be analyzed in terms of positive network externalities.
- <sup>5</sup> Based on Arthur: 1994: 3.
- <sup>6</sup> It is important to distinguish this notion of consent from promissory obligations, which I consider in section 5. Compare Gilbert (1986: 55): “Actual consent theory invokes an agreement as opposed to a contract in law.”

### References

- Arthur, W. Brian. (1994) *Increasing Returns and Path Dependency in the Economy*. Ann Arbor: University of Michigan Press.
- Baier, Kurt. (1965) *The Moral Point of View*, abridged edn. New York: Random House.
- Bicchieri, Cristina. (2006) *The Grammar of Society: The Nature and Dynamics of Norms*. Cambridge: Cambridge University Press, 2006.
- Boyd, Robert and Peter Richerson. (2005) *The Origin and Evolution of Culture*. Oxford: Oxford University Press.
- Darwall, Stephen. (2006) *The Second-person Standpoint: Morality, Respect and Accountability*. Cambridge, MA: Harvard University Press.
- Estlund, David. (2008) *Democratic Authority*. Princeton: Princeton University Press.
- Filmer, Robert. (1991) *Patriarcha* in *Patriarcha and Other Writings*, edited by Johann P. Sommerville. Cambridge: Cambridge University Press.
- Freidman, Daniel. (2008) *Morals and Markets: An Evolutionary Account of the Modern World*. New York: Palgrave Macmillan.
- Gaus, Gerald. (2011a) *The Order of Public Reason: A Theory of Freedom and Morality in a Diverse and Bounded World*. Cambridge: Cambridge University Press.
- (2011b) “The Demands of Impartiality and the Evolution of Morality” in *Partiality and Impartiality*, edited by Brian Feltham and John Cottingham, Oxford: Oxford University Press.

- (2001c) "The Property Equilibrium in Our Liberal Social Order." *Social Philosophy & Policy*, vol. 28 (Summer, no. 2).
- Gilbert, Margaret (2006) *A Theory of Political Obligation*. Oxford: Clarendon Press.
- Green, T.H. (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.
- Hegel, G.W.F. (1952) *The Philosophy of Right*, translated by T.M. Knox. Oxford: Oxford University Press, 1952.
- Hobbes, Thomas. (1994) *Leviathan*, edited by Edwin Curley. Indianapolis, IN: Hackett.
- Kant, Immanuel. (1999) *The Metaphysical Elements of Justice*, 2nd edition, edited and translated by John Ladd. Indianapolis: Hackett, 1999.
- Locke, John. (1960) *The Second Treatise of Government* in *Two Treatises of Government*, edited by Peter Laslett. Cambridge: Cambridge University Press, 1960.
- Martin, Rex. (1993) *A System of Rights*. Oxford: Clarendon Press.
- Nagel, Thomas. (1991) *Equality and Partiality*. New York: Oxford University Press. Oxford University Press.
- Rawls, John. (1996) *Political Liberalism*. New York: Columbia University Press.
- Rawls, John. (1999) *A Theory of Justice* Cambridge, MA: Harvard University Press.
- Richerson, Peter J. and Robert Boyd. (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.
- Scanlon, T.M. (1998) *What We Owe Each Other*. Cambridge, MA: Belknap Press of Harvard University Press.
- Schwab, David and Elinor Ostrom. (2008) "The Vital Role of Norms and Rules in Maintaining Open Public and Private Economies" in *Moral Markets*, edited by Paul Zak. Princeton: Princeton University Press.
- Simmons, A. John. (2001). *Justification and Legitimacy: Essays on Rights and Obligations*. Cambridge: Cambridge University Press.