

REPRESENTATION AND OBLIGATION IN RAWLS'S SOCIAL CONTRACT THEORY

Simon Cushing
University of Southern California

There are two main roles for an account of justification in political philosophy. The first is that of political obligation, of settling the question of which society I am uniquely obligated to.¹ The second role for an account of justification is that of establishing whether or not a social system is a *just*, or more vaguely, a *legitimate* one. That the two justificatory roles are distinct can be seen by noting that even in a world of many perfectly just societies with perfectly legitimate laws, one would still presumably only be a citizen of one of those countries. These roles are not clearly distinguished in the classic social contract theories of Locke and Rousseau in particular, but there the contract fills both roles, as I will show. Rawls's hypothetical social contract seems designed only to address the second role, but Rawls does in fact explicitly tackle the issue of political obligation in *A Theory of Justice*,² where he tries to remedy the flaws of the classic social contract *obligation* account with a *duty-based* account. In this paper I shall assess the criticisms of his account offered by A. John Simmons in his book *Moral Principles and Political Obligations*³ and a possible response to the criticisms that Samuel Freeman has implicitly offered in his work on Rawlsian social contracts. I conclude, however, that this response rests on an equivocation in *Theory* and thus fails.

1. The Founding Contract

Consider the following model, intended as a very rough sketch of a literal reading of Locke's social compact: Art, Billie, Charlie and Dinah decide to form a club. They openly discuss rules of behaviour for members of the club, the scope of the rules, positions of authority in the club, rules for election into those positions, term limits, etc. Having come to a consensus on all points, the four of them jointly agree to found the club, which currently comprises the four of them. Call the agreement they reach the Founding Contract, and the sketch in general the Club Model.⁴

The Founding Contract provides an *obligation* account of the club members' moral requirement to follow the rules: Art et. al. are all bound to follow the rules, *because* they agreed to them. Consider Ella, for example, who sat in on the discussion of rules, contributed many valuable suggestions which the others all agreed made the rules better, but declines to join in the

actual act of agreement. Nobody would say, I think, that Ella is a member of the club or bound to follow its rules. The actual act of agreeing to be a member (whether it take the form of swearing an oath, shaking hands or whatever agreed form of acknowledgment) is required to found obligation.

What about legitimacy? The source of legitimacy of the rules *for those members* is the fact that they openly discussed them and came to an agreement on them (you might say this was the bargaining stage of the Founding Contract, without which it is not a contract but simply an agreement).

Thus if the social contract approaches our rough sketch, then different aspects of the same contract serve both to legitimize the social ordering and bind members to it. Granted, the Club Model is a crude sketch, but it does serve to highlight the strengths and weaknesses of the idea of the founding contract as found in Locke, for example. As an account of political obligation it has these strengths. An actual agreement by an actual citizen to abide by the laws of her society, provided it takes place in the right, non-coerced circumstances (admittedly a big proviso) is intuitively a plausible sufficient condition for a moral obligation. Furthermore, the fact that it is a social contract means that the justificatory strategy respects the *autonomy* and *individuality* of the members of the club, which is what particularly appeals about the use of a social contract in political philosophy, in contrast, in particular, to various versions of utilitarianism.

However, the major flaws are of course the following: first, it is by no means clear that founders of actual countries agreed in this way. Second, even if they did, their consent does not bind us. And third, although we could be bound, like Ella, by an act of *joining* consent, in real instances such a clear, intentional sign of agreement, in conditions of full knowledge of alternatives, is almost never given. And the flaws in the alternative notion of *tacit* consent (say, on the grounds that we live or own property within the borders of a country) are well-documented.⁵ Thus, the Founding Contract fails to be *general* enough to account for the obligation of at least the large majority of citizens, who live peacefully within a society and indeed if asked would regard themselves as citizens. It is this lack of generality that prompts Rawls's alternative, duty-based account, to which we now turn.

2. Rawls's Duty-Based Account

In *A Theory of Justice*, Rawls states:

From the standpoint of justice as fairness, a fundamental natural duty is the duty of justice. This duty requires us to support and to comply with just institutions that exist and *apply to us*... Thus if the basic structure of society is just, or as just as it is reasonable to expect in the circumstances, everyone

has a natural duty to do his part in the existing scheme. Each is bound to these institutions independent of his voluntary acts, performative or otherwise. [p. 115, my emphasis]

Rawls's account of political obligation⁶ is in terms of a natural duty that we have to comply with just institutions. Thus it is only obliquely contractarian, in that the contract is a way to test whether or not an institution is just, its justice being a necessary condition for the application of the duty to comply. Let me sketch a second model to illustrate roughly how Rawls's social contract contrasts with the Club Model.

In this case, Art, Billie, Charlie and Dinah are particularly concerned that the rules of their club (or, to be more precise, the general principles that will frame the rules) be *fair*, but are worried that if they bargain over the rules themselves, their personalities and relationships will distort the principles in favour of the stubbornest or most forceful personality among them.⁷ So, instead of working out the rules of their club themselves, they each hire an agent, and get the agents together jointly to choose from a set of conceptions that our prospective club members themselves provide as exhausting the most reasonable alternatives to produce principles to order their rules. The agents are not informed about any particular features of their clients, but each consultant is told to represent *only* her client's interests. Assuming that Rawls is correct, and that the perfect rationality of the agents (the parties of his original position) combined with the veil of ignorance that divides the agents from their clients (which models perfect equality) ensures that the political conception the agents choose is fair (and by extension, *just*), it seems that on Rawls's picture the four *clients* are obligated to follow the principles chosen by their *agents* because of each client's natural duty of justice. Call this model the Consultant Model. I will postpone further comment until I have laid out Simmons's criticisms.

Note that in his description of the natural duty of justice Rawls includes the stipulation that the just institutions *apply to us*. This seems a reasonable enough stipulation, for it preserves the distinction between political obligation and an account of justice, without which we could not explain why we are citizens of one just state over another.⁸ Simmons seizes on just this clause, however, as the basis of his criticism that Rawls's duty account is inadequate to ground a bond of citizenship. He analyses the possible senses in which an institution could be said to "apply to" an individual, and suggests that they fall into three rough categories: the "weak" sense, the "territorial" sense, and the "strong" sense.

Simmons argues, and I think he is right, that in *Theory* Rawls assumes the "territorial" sense of application but that this does not intuitively

distinguish a just institution to which one is not bound from a just institution to which one is:

[D]oes it follow from this institutions 'applying to me' in this sense that I am *morally required* to follow its rules if the institution is just?... I think not, for the thing which *makes* the institution apply to me here is the simple fact of my birth and growth in a territory within which the institution's rules are enforced: but my birth is not an act I perform, or something for which I am responsible. [pp. 149-50]

There is a sense in which I have a moral duty to obey *any* rules that are just, but this applies equally to all the just rules of any country in the world. The only kind of application that will provide the moral bond necessary for political obligation to *my particular* government, is strong application, which requires in addition to territorial application that

I am an active participant in the activities of the institution, and am a member in the full sense of the word. I have given my express consent to be governed by its rules, or perhaps I have held office in the Institute, or *accepted*... substantial benefits from the institution's workings. [p. 150]

However, the ground of my moral bond is now the fact that I performed some *deliberate consensual act*, and thus, contends Simmons, the duty account collapses into an obligation account once more. To illustrate, let us return to my Consultant Model. My intuition is that even after the agents have made their choice of political conception (from which are derived the principles to order the rules of the club), the conception does not *apply* to the clients in a strong enough sense that they must abide by it. That would only happen if Art and co. agreed to *accept* the conception chosen. A Rawlsian could perhaps respond that the agreement had already been made in the *hiring* of the agents. I will say more on this below, but suffice to say that Ella is not bound by the choice of the agents even if she shares her four friends' concern for fairness and was considering joining the club, even despite the fact that the veil of ignorance means that any of the four agents could easily be taken to represent her interests as well as they do the interests of the clients who hired them.

Given Rawls's own doubts about the adequacy of an obligation-based account, and in particular his dissatisfaction with its *lack of generality*, Simmons's conclusion that this is what his duty-based account amounts to is an unfortunate one. What response can Rawls, or a Rawlsian, make?

3. A Rawlsian Response

All accounts of justification work from certain fundamental ideas. In classical social contract theories, these crucially include the conception of individuals as *free and equal*.⁹ For Rawls, the conception of the person follows from an even more basic idea:

the fundamental organizing idea of justice as fairness, within which the other basic ideas are systematically connected, is that of *society as a fair system of cooperation* [PL p. 15, my emphasis]

Given this conception of society, the question arises: how are the fair *terms* of cooperation to be established? Rawls's answer to this question embodies his contractarianism.

Are they, for example, laid down by God's law?... Or are these terms established by an undertaking among those persons themselves in the light of what they regard as their reciprocal advantage?....

Justice as fairness recasts the doctrine of the social contract and adopts a form of the last answer: the fair terms of social cooperation are conceived as agreed to by those engaged in it, that is, by free and equal citizens who are born into the society in which they lead their lives. [PL, pp. 22-3]

Here is an apparent response to Simmons's challenge. If *the citizens themselves* were "agreeing" on the fair terms of cooperation, then this would appear to meet Simmons's requirements for "strong application" of the fair terms of cooperation (or at least the principles that govern them), and Rawls's natural duty would succeed in providing a general account of obligation for all citizens. However, more needs to be said to support this bold claim that the citizens have "agreed" to the terms. To see how this might be justified, we can draw on the article "Reason and Agreement in Social Contract Views"¹⁰ by Samuel Freeman, who is sympathetic to what he calls the *right-based* social contract approach, of which Rawls's theory is the most sophisticated example.

Freeman's article covers much ground, but what interests us here is his contention that Rawls's contract is a "general precommitment" given by the citizens of a society whose basic structure is ordered by the principles chosen by the parties in the original position. To begin with, Freeman asserts that all free and equal moral persons have, in their sense of justice¹¹ a shared social interest in cooperating with one another on publicly justifiable terms that express their conception of themselves as free and equal. He goes on:

To realize this shared social interest, free and equal citizens mutually precommit themselves (*through their representatives in the Original Position*) to principles, appropriate to their self-conception, for the design of institutions and the regulation of their individual pursuits....

Members of a democratic society make this general precommitment, not the parties in the Original Position. The parties choose from interested motivations and have no basic concern for justice and equality. Here Rawls's argument from the "strains of commitment" plays an essential role. Because the parties conceive of themselves as free and able to control their future, each sees himself as responsible to the future claims of the self. Each knows that he will be held to his decision by the others, so each aims

to commit himself once and for all to principles that he will not reproach himself for choosing even should the worst transpire. Commitment by the parties to principles that secure equal citizenship and enable each freely to pursue his own good corresponds to the precommitment of free and equal citizens in society. [pp. 143-145, my emphasis]

I take Freeman's account here to be a fleshing out of Rawls's view that citizens "agree" to the fair terms of social cooperation, and can thus be said to be "active participants" (in Simmons's words) in the institutions of their society, and thus the institutions apply to them in the strong sense, and their duty of justice binds the citizens to those institutions. However, Freeman's sketch fails to meet the criteria of strong enough participation because it relies on arguments resting on an ambiguity in *Theory* concerning the relationship between the parties and the citizens of the society ordered by the principles the parties choose. In *Theory* Rawls alternates between describing the parties as *being* the citizens of the well ordered society (identification) and describing them as acting *for* the citizens (representation). In PL however, in response perhaps to criticisms that the veil of ignorance strips the parties of all features that could be used to identify them with any particular citizen, the parties are always described as representatives of the citizens.¹²

The "strains of commitment" to which Freeman refers are the strains that a citizen would feel if there were a clash between her legitimate interests and the demands of the principles of justice ordering her society. (These strains would be intolerable, say, if the principles were utilitarian and demanded she sacrifice herself for the greater good.) Rawls claims that because of the common knowledge of a shared sense of justice, which means that citizens will hold each other to the terms of the initial agreement, the parties cannot agree to something that "they" (implying that the citizens and the parties share identities) would not accept later, and thus such strains of commitment must be avoided.

Two points can be made about this argument. First is that, with the understanding that the parties are *representatives* of the citizens, the commitment of the parties is more like the commitment of the agents in my Consultant Model than the commitment of the club members themselves in my Club Model. It is not like the precommitment that Ulysses makes, in having himself tied to the mast, to constrict his future desires to be with the sirens (to which Freeman compares it) because that is the commitment of an earlier phase of the same self, while the parties are distinct (artificial) individuals. Furthermore, there is not even the bond between the parties and the citizens that there is in my consultant model, because there the parties deliberately employed the agents.

The second is that there is no need for the *parties* to share a sense of justice, as Rawls suggests in *Theory* and as Freeman assumes, only for them to know that the citizens of the well-ordered society will have one and will all hold each other to fair terms. Once this admission is made, I think it becomes clear that the sense of justice of the citizens is doing the work of obligation. Even if we grant Rawls that all citizens should be understood to have this capacity to be reasonable, the problem remains of deciding *which* fair terms "apply" to each particular citizen. Rawls claims in PL that an overlapping consensus of reasonable comprehensive doctrines will emerge over his principles of justice, but even if this were so, without a deliberate consensual act of acceptance, we would still be in the situation of Ella in my Consultant example. There is still the crucial gap between *fair* terms and *my* fair terms, which Rawls's duty-based account has failed to bridge.

It remains for the Rawlsian to throw up his or her hands and proclaim that someone who recognizes that the original position produces fair principles and possesses a sense of justice such that they are committed to accepting fair terms, but still refuses to accept a moral bond between themselves and the principles of justice is simply failing to recognise what is implicit in the political conception they already accept. But if our intuitions are drawn from our "shared public political culture" as are the conceptions of society and citizen that Rawls deems basic to justice as fairness, then, if my Consultant Model is right, we would not be being true to those shared ideals if we conceded that we are bound without the kind of clear consent an obligation account requires.

Notes

¹ Strictly speaking, political obligation is obligation to a *government* rather than a society, so I have conflated it here with the question of *citizenship*, which society is *my* society. However, for the purposes of this paper, I think the distinction is inessential, and I shall continue to treat political obligation as described.

² Cambridge, Mass.: Harvard University Press, 1971, henceforth *Theory*.

³ Princeton: Princeton University Press, 1979.

⁴ In this paper I ignore the debate, covered at length in the *Political Liberalism* (New York: Columbia University Press, 1993, henceforth PL) lecture "The Basic Structure as Subject," over whether or not associations are appropriate models for societies. I believe Rawls's points are correct, and address them in another paper, but feel they can be put to one side for the purposes of this discussion.

⁵ See, for example, Chapter IV in *Moral Principles and Political Obligations*.

⁶ Rawls distinguishes obligations from duties (obligations are normally owed by and to definite individuals and come about as a result of voluntary acts, while duties are more general) so strictly speaking it is not a political *obligation*, but I am

using the common phrase and mean it generally as a normative bond between citizen and political structure.

⁷ My example models Rawls's (and Rousseau's) concern that a compact like Locke's would favour those in a superior bargaining position. See Rousseau's contention in his *Discourse on the Origin of Inequality* that the rich would thereby influence laws in their favour, and the poor would be tricked into complying with a system that favours the propertied.

⁸ Rawls could perhaps respond that he is not trying to provide an account of *citizenship* here, but simply to explain why we should obey *any* institution, but the fact that he does insist on the "application" clause in the first part of the duty of justice, but does not in the second - "to further just arrangements not yet established" - suggests that citizenship is what he had in mind.

⁹ Thus the classic contractarians' opposition to such theories as the Divine Right of Kings, Sir Robert Filmer's version of which Locke criticized in both of his *Treatises on Government*.

¹⁰ In *Philosophy and Public Affairs 19* (Spring 1990): 122-57. As Rawls himself cites Freeman's article approvingly in PL (53*n*), and is thanked by Freeman for his help in its writing, I will assume that Freeman faithfully represents Rawls's position.

¹¹ Recall that Rawls claims that the two moral powers of citizens comprise their capacity for a sense of justice and their capacity for a conception of the good. For the purposes of this paper, I will assume that it is legitimate for Freeman to assume an *actual* sense of justice, rather than a mere capacity, although I cannot say if this is true to Rawls's intent.

¹² For example, in *Theory* Rawls writes of the parties: "no one knows *his* place in society" (p.12), while in PL he writes: "the parties are not allowed to know the social position of *those they represent*." (P.24). Furthermore, it is clear in PL that when Rawls says the parties "represent" the citizens, he means represent in the way the agents in my Consultant Model represent their clients, and not in the sense that the veil of ignorance "represents" strict equality (or that *c* can represent the speed of light in $e=mc^2$). See for example PL p.106, where Rawls writes of the parties: "They are doing what trustees are expected to do for the persons they represent." Mylan Engel alerted me to the possibility of the alternative reading of "represent."