The fool and the franchiser: formal justice in the political theories of Hobbes and Rawls

Jan Niklas Rolf*
Faculty of Society and Economics, Rhine-Waal University of Applied Sciences, Germany

Abstract
Thomas Hobbes and John Rawls are usually portrayed in antagonistic terms. While Hobbes, one of the first scholars to translate Thucydides, is often held to be an archetypal realist, Rawls, a self-proclaimed follower of Kant, is frequently said to argue from an explicit normative position. In this paper, I try to demonstrate that the two philosophers have more in common than is generally thought. Drawing on Hobbes’s answer to the fool and Rawls’s analogy of the franchiser, I suggest that there is a powerful link between the two philosophers that can tell us something valuable about their theories of formal justice. Against Brian Barry’s characterization of Hobbes as an advocate of justice as mutual advantage and Rawls as a proponent of both justice as mutual advantage and justice as impartiality, I argue that the two philosophers adhere to one and the same tradition of justice, justice as reciprocity, which bases obligations of reciprocity not only on explicit express, but also on tacit acceptance of benefits.

Keywords: John Rawls; Thomas Hobbes; formal justice; justice as mutual advantage; justice as impartiality; justice as reciprocity

Since its publication in 1651, there has been extensive writing about Thomas Hobbes’s Leviathan. John Rawls’s oeuvre, on the other hand, ‘has generated more secondary literature than perhaps any other work of twentieth century political philosophy’. While Hobbes and Rawls may have attracted sufficient interest in their own right, it is the claim of this paper that the relationship between the two has not
been adequately thought through. In her article ‘Reading Rawls and Hearing Hobbes’, Rosamond Rhodes draws attention to a *New York Times* book review of Rawls’s *Political Liberalism* in which the author, John Gray, tries to establish a link between Hobbes and Rawls, but to her knowledge nothing more has been written on the subject.\(^2\)

While Joseph Grcic’s article ‘Hobbes and Rawls on Political Power’ should not go unmentioned,\(^3\) my own review of the literature has shown that, apart from these three papers, no article, and certainly no monograph, has been published that puts Hobbes and Rawls into comparative perspective.\(^4\) As Grcic notes, the differences between their philosophies seem to be too obvious to justify a comparison.\(^5\)

One fundamental difference concerns their contrasting ideas of sovereignty. Not least because of the different times in which they were writing, Hobbes advocates an unlimited sovereign, whereas Rawls is committed to a liberal democratic system where power is effectively balanced. In fact, Hobbes’s view of government as having absolute power over its subjects could not be more at odds with Rawls’s vision of government as the body that maximizes individual liberties. A further difference pertains to their conflicting understandings of justice. This difference, as I will argue below, has less to do with what justice requires than with the conditions under which it matters. While for Hobbes, justice comes into play only after the institution of the social contract, for Rawls, questions of justice arise whenever the circumstances of justice obtain. A third area of disagreement has to do with the nature of social cooperation. Hobbes holds a negative conception of cooperation in that each person gains from the reciprocal renunciation of his or her right of nature, whereas Rawls embraces a positive conception of cooperation in that the gains stem from an economic surplus acquired by collaboration. Here again the difference is in detail rather than in substance, which is that social cooperation leads to mutually beneficial outcomes, be it ‘in a directly Hobbesian or anonymously Rawlsian form’.\(^6\)

Although Hobbes and Rawls differ in central aspects of their philosophies, it is possible to make out a number of striking similarities between them. Unfortunately, most of the resemblances identified in the literature are either very general (and thus liable to being simplistic) or extremely specific (and thus liable to being pointless).\(^7\) One of the more interesting claims made by Rhodes is that although ‘Rawls is noted for his focus on justice and Hobbes is noted for his attention to order, in fact, they each have a dual agenda and recognize that one end cannot be achieved without the other’.\(^8\) To make good this claim, Rhodes cites a passage from *Leviathan*, where Hobbes suggests that justice is one end for which the state is instituted, and refers to *Political Liberalism*, where Rawls comes to appreciate the importance of stability for a just society. It remains true, though, that Hobbes is primarily concerned with order, while Rawls’s interest lies with justice. When Hobbes speaks of justice, he tends to do so in the context of securing order, and when Rawls speaks of a well-ordered society, he means a society that is regulated by a public conception of justice.

The argument that I am most interested in is that ‘both Hobbes and Rawls take their project to be the creation of a just and stable society in the face of general egoism’.\(^9\) As Rhodes explains, ‘neither Hobbes nor Rawls believes that reasonable people are moved by a conception of the general good, but so that each may benefit’.\(^10\) Building
The fool and the franchiser

on this, the present paper demonstrates that Hobbes and Rawls do in fact reject the strong moral demands of justice as impartiality. But it goes further to show that the two philosophers also reject the short-sighted egoistic approach that is justice as mutual advantage. Instead, I argue that Hobbes and Rawls commonly adhere to the tradition of justice as reciprocity, which, for both philosophers, entails not only an obligation to reciprocate where one has made an agreement, but also where one accepts the benefits of others’ actions.

In section one, I will devise a table based on the specific game structure of the three traditions of formal justice. A discussion of Hobbes and Rawls in section two will be instrumental toward determining their place in this table in section three.

CONCEPTIONS OF JUSTICE

Justice has been given both a purposive and a formal meaning. In its purposive version, a just act is one that serves good ends. Purposive justice roughly corresponds to what Hobbes calls distributive justice and to what Rawls calls justice for institutions. In its formal version, a just act is one that conforms to certain rules. Formal justice roughly corresponds to what Hobbes calls commutative justice and to what Rawls calls justice for individuals. According to Terry Nardin, a just act must ‘be understood as having both a purposive and a formal aspect—that is, as being at the same time action directed toward some substantive good and action relative to certain rules’. In other words, purposive justice prescribes ends, whereas formal justice governs the pursuit of these ends. The focus of this paper will be on formal justice, as Hobbes and Rawls made multiple references to it.

In the first volume of his treatise on justice, *Theories of Justice*, Brian Barry identified two traditions of justice: justice as mutual advantage and justice as impartiality. One of Barry’s central claims was that Rawls’s theory of justice contains elements of both traditions. This has been questioned by Allan Gibbard, who argued that neither justice as mutual advantage nor justice as impartiality could be associated with Rawls. Instead, Gibbard believes Rawls to stand in the tradition of justice as reciprocity. In the second volume of his unfinished treatise, *Justice as Impartiality*, Barry came to recognize justice as reciprocity as a distinctive tradition, but insisted on his earlier characterization of Rawls as a proponent of justice as mutual advantage and justice as impartiality. In order to make our own judgment, I will examine the three traditions with a view to their specific game structure.

According to Barry, justice as mutual advantage consists ‘in the constraints on themselves that rational self-interested people would agree to as the minimum price that has to be paid in order to obtain the cooperation of others’. This brings out the idea that people expect to benefit, not from their own act of cooperation, but from that of their fellows. But this carries the risk of people no longer cooperating as soon as they obtain the cooperation of others. As Barry acknowledges in the second volume of his treatise, justice as mutual advantage is a very unstable conception with the structure of a prisoner’s dilemma: player 1 will not only not cooperate when he or she believes player 2 not to cooperate; he or she will also not cooperate when he or
she believes player 2 to cooperate. Consequently, justice as mutual advantage is located in the lower two rectangles in Table 1.

The basic idea of justice as impartiality can be expressed in different ways, whereby ‘the details of these differences can produce profound effects on the outcomes that the theory generates’. While Barry cannot give a general account of what justice as impartiality comprises, he provides us with the motive for behaving in conformity with it, which ‘is the desire to act in ways that can be defended to oneself and others without appealing to personal advantage’. Such a ‘desire to act justly’, it appears, is independent of other people’s desires. Cooperation, then, is a dominant strategy: player 1 will cooperate irrespective of whether he or she believes player 2 to cooperate or not. This is why justice as impartiality appears in the upper two rectangles in Table 1.

Finally, justice as reciprocity contains—as its name suggests—an element of reciprocity. As such, it has essentially the structure of an assurance game: player 1 will cooperate when he or she believes player 2 to cooperate, and he or she will not cooperate when he or she believes player 2 not to cooperate. Accordingly, justice as reciprocity is included in the upper left and lower right rectangles in Table 1.

**THEORIES OF JUSTICE**

The task that remains is to situate Hobbes and Rawls in the Table 1 above. I begin with a discussion of Hobbes’s theory of justice, before I turn to Rawls’s theory of justice.

Barry takes Hobbes to be the ‘greatest expositor’ of the tradition of justice as mutual advantage. Hobbes’s idea that people, in laying down their right of nature, escape the dreadful state of nature does in fact correspond to the tradition’s presentation of self-restraint as a rational sacrifice of part of our good for the sake of a greater good. Recall, however, that justice as mutual advantage implies that whenever people can reap the greater good without having to make a sacrifice, or at least no immediate one, they will not make that sacrifice. Accordingly, a person moved by justice as mutual advantage will not make the sacrifice of laying down his or her right of nature where the other side has made that sacrifice already. It is at this point that Hobbes departs from justice of mutual advantage.

In order to see why this is the case, consider Hobbes’s distinction between a contract and a covenant. A contract, Hobbes defines as ‘the mutual transferring of right’. However, when one of the parties delivers ‘the thing contracted for on his part, and leave[s] the other to perform his part at some determinate time after, and
in the meantime be trusted; . . . then the contract on his part is called . . . covenant’. 28
Now, if it was possible to make an all-embracing contract—an agreement to deliver the thing contracted for at the same time—problems of compliance could hardly arise. For Hobbes, though, agreements in the state of nature do not only assume the form of a contract but also that of a covenant—an agreement to perform successively. But then the laws of nature may oblige only in foro interno, that is, we wish to obey them, but not in foro externo, that is, we do not need to obey them in practice. 29 ‘For he that performeth first’, Hobbes explains, ‘has no assurance the other will perform after’. 30 Rather, he makes ‘himself a prey to others, and procedure[s] his own certain ruin, contrary to the ground of all laws of nature, which tend to nature’s preservation’. 31

From this, several scholars have concluded that Hobbes’s contract theory possesses the structure of a prisoner’s dilemma. 32 According to these scholars, people in the state of nature not only have a good self-interested reason not to observe the laws of nature when everyone else does not (yet) observe them; they also have a good self-interested reason not to observe them when everyone else does observe them. Why, these scholars ask, should a person renounce his or her right of nature, that is, his or her liberty to do whatever he or she wants at any given moment, when everyone else’s renunciation makes him or her secure?

Yet at no point in Leviathan does Hobbes indicate that his contract theory has the structure of a prisoner’s dilemma. In the paragraph subtitled ‘Justice Not Contrary to Reason’, Hobbes takes issue with the fool who claims that ‘not [to] keep covenants, was not against reason, when it conduced to one’s benefit’. 33 What the fool is essentially questioning is Hobbes’s third law of nature, specifying that people perform their covenants made. Since a law of nature is defined as a precept found out by reason, Hobbes needs to reject the claim that keeping of covenant can be against reason. 34 ‘[W]here one of the parties has performed already’, Hobbes says in all clarity, ‘there is the question whether it be against reason, that is, against the benefit of the other to perform, or not. And I say it is not against reason’. 35

On what grounds does Hobbes make this claim? Answering the fool, Hobbes points out that the person

that breaketh his covenant, and consequently declareth that he thinks he may with reason do so, cannot be received into any society, that unite themselves for peace and defence, but by the error of them that receive him; nor when he is received, be retained in it, without seeing the danger of their error; which errors a man cannot reasonably reckon upon as the means of his security: and therefore if he be left, or cast out of society, he perisheth; and if he live in society, it is by the errors of other men, which he could not foresee, nor reckon upon; and consequently against the reason of his preservation; and so, as all men that contribute not to his destruction, forbear him only out of ignorance of what is good for themselves. 36

Provided that people commit no error, Hobbes believes that a person violating his or her covenant would be denied membership to society. But social exclusion can by no means be in a person’s interest: if a man is ‘cast out of society’, Hobbes says, ‘he perisheth’, because he ‘can in reason expect no other means of safety, than what
can be had from his own single power'. From this, Hobbes concludes that keeping of covenant is a rule of reason. Consequently, his theory does not qualify as an example of justice as mutual advantage. As Edwin Curley has pointed out correctly, the situation Hobbes contemplates in reply to the fool is not a prisoner’s dilemma but an assurance game, where one does whatever the other player does. 

This is also evident from Hobbes’s laws of nature, and the second law in particular, which holds

that a man be willing, when others are so too, … to lay down this right to all things [the right of nature]; and be contented with so much liberty against other men, as he would allow other men against himself …. But if other men will not lay down their right, as well as he; then there is no reason for any one, to divest himself of his.

The first part of the law dictates that a person lays down his or her right of nature, provided that other people do so as well. If other people do not lay down their right of nature, the second part of the law applies, which postulates that a person does not have to lay down his or her right of nature. Note that to comply when others comply, and not to comply when others do not comply, is nothing else than to reciprocate. At the bottom of Hobbes’s second law of nature then lies a pattern of reciprocity. This pattern is also characteristic of Hobbes’s third law of nature, in which ‘consisteth the fountain and original of justice’. It postulates ‘that men perform their covenants made’, which, after all, is to reciprocate. Accordingly, when Hobbes speaks of justice he means justice as reciprocity.

But the obligation to reciprocate not only applies to situations in which one has made a covenant, that is, an explicit agreement to reciprocate, but also to situations in which one benefits from the actions of others without having made such an agreement. This is apparent from Hobbes’s fourth law of nature, the law of gratitude, which states that ‘a man which receiveth benefit from another of mere grace, endeavour that he which giveth it, have no reasonable cause to repent him of his good will’. What Hobbes is essentially saying here is that whenever we enjoy the benefit of other people’s actions, we have an obligation to reciprocate, that is, to act in such a way that the person who benefits us could not wish that he or she had never benefited us. This idea is stated even more rigorously in On the Citizen, where Hobbes argues that in order for people to be obliged they need to have given their ‘personal consent’ whereby this consent, ‘if it is not explicit, it must at least be implied, as when they accept the benefit of a person’s power and laws for protection and preservation of themselves against others’.

What this shows is that Hobbes bases obligations of reciprocity not only on explicit express, but also on tacit acceptance of benefits. In his fourth law of nature, Howard Warrender has rightly pointed out, ‘Hobbes is moving from the principle that promises oblige towards the principle that benefits oblige’. Accordingly, our obligation to reciprocate extends from situations in which we have made a promise—or a covenant, for ‘a promise is equivalent to a covenant’—to situations in which we accept the benefit of other’s actions. But then obligations of reciprocity must arise
in a multitude of circumstances in human life. For Hobbes, justice as reciprocity is all but a bounded theory.

Having shown that justice as reciprocity lies at the heart of Hobbes’s moral philosophy, what are we to make of Barry’s characterization of Rawls as a proponent of both justice as mutual advantage and justice as impartiality? Let us consider the two traditions in turn.

Rawls, recognizing that ‘while we normally think of moral requirements as bonds laid upon us, they are sometimes deliberately self-imposed for our advantage’, certainly draws the connection between self-restraint and advantage that is also central to justice as mutual advantage. The controversy is one that arises once the bonds are laid upon us. As a utility maximizing theory, justice as mutual advantage implies that it is in my interest ‘that everybody else adheres to rules that are mutually advantageous if generally adhered to and I break them whenever it is to my advantage to do so’. This leads to the question of whether justice as mutual advantage is a theory of justice at all. This, however, is a question that does not have to concern Rawls, because he holds that once a cooperative agreement has been set up it is irrational to break the rules of the agreement. In a footnote that has received little to no attention, Rawls draws an interesting analogy between a private cooperative venture and the cooperative venture of society that is worth citing in full length.

Consider the following analogy, which I owe to Peter Murrell. A franchiser (say Dunkin’ Donuts) is deciding what kind of terms to put in its contract with its many franchisees. Suppose there are two strategies it may follow. The first is to try to make a separate contract with each franchisee, hoping to take a higher percentage of the return in better-situated franchises, as well as increasing the percentage when particular franchises become more profitable. The second strategy is to set once and for all a fixed percentage that seems fair throughout the franchise and to require of franchisees only certain minimum standards of quality and service so as to preserve the franchiser’s reputation and the good will of the public, on which its profit depend. Here I assume that the minimum standards of quality and service are quite clear and can be enforced without seeming arbitrariness. Note that the second strategy of setting a fixed percentage for all franchises and enforcing minimum standards has the advantage that it fixes once and for all the terms of agreement between the franchiser and the franchisees. The franchiser’s interest in its reputation is secured, while at the same time franchisees have an incentive to meet the franchiser’s minimum standards and to increase their own return, thus strengthening the franchise as a whole. They know the franchiser will not try to increase its return should they become more prosperous. Thus, given the very great initial uncertainty the franchiser faces, the great uncertainty in cooperative relations between franchiser and franchisee that the first strategy would perpetuate, and the continuing suspicion and distrust which that uncertainty would cause, the second strategy is superior. From the point of view of the franchiser’s own interests it is more rational to try to create a climate of fair cooperation based on clear and fixed terms that strike all parties as reasonable than to try for adjustable fine-tuned contracts that might enable the franchiser to increase profits as particular opportunities arise. There is some evidence that in fact successful franchisers follow the second strategy.
What is it that makes the franchiser prefer acting fairly and not increasing profits to acting unfairly and increasing profits? Rawls says that following the second strategy has ‘the advantage that it fixes once and for all the terms of agreement between the franchiser and the franchisees’. By means of this, ‘the franchiser’s interest in its reputation is secured while at the same time franchisees have an incentive to meet the franchiser’s minimum standards and to increase their own return, thus strengthening the franchise as a whole’. What appear to be two advantages for the franchiser—good reputation and realized minimum standards—is in effect just one, because three sentences earlier Rawls defines realized minimum standards as an explanatory variable of good reputation. What makes the second strategy superior to the first one, then, is that franchisees have an incentive to meet the franchiser’s minimum standards (thereby securing the reputation of the franchise, which, in turn, yields higher returns for the franchiser) and to increase their own returns (thereby further increasing the returns of the franchiser).

Rawls follows that in the cooperative venture of a franchise it is more rational to comply with the terms of the agreement one has made than to push for better terms should the opportunity arise, and, by analogy, that in the cooperative venture of society it is more rational to comply with the terms of the agreement one has made than to push for better terms should the opportunity arise. But insofar as Rawls’s theory is not one of justice as mutual advantage, is it one of justice as impartiality?

The answer to this, I suggest, lies in the term ‘the reasonable’, a ‘difficult’ term that ‘easily becomes vague and obscure’ but that ‘will not be explicitly defined’. This term appears for the first time in Rawls’s 1958 essay ‘Justice as Fairness’, where he comes up with the idea that self-interested people decide over the principles of justice that should govern the basic institutions of their society. To guarantee this choosing situation to be fair, Rawls describes the people as being subject to certain constraints. Later, in A Theory of Justice, Rawls presents these constraints, inter alia, in the form of the veil of ignorance. However, at the time he wrote ‘Justice as Fairness’, he had not developed this device yet. Instead, he wants us to imagine the constraints ‘analogous to those of having a morality’, where ‘having a morality must at least imply the acknowledgement of principles as impartially applying to one’s own conduct as well as to another’s’. Being constrained in this way, ‘rational and mutually self-interested persons are brought to act reasonably’. From this, we might conclude with Catherine Audard that the reasonable embraces an idea of impartiality.

Rawls’s primary concern in Political Liberalism, a collection of lectures he gave in the 1980s, is to demonstrate that the principles of justice are recognized from within the perspective of citizens’ comprehensive moral doctrines. But in order to show that the reasonable can connect with the motivations of citizens, Rawls has to free it from some moral content. ‘For the purposes of a political conception of justice’, Rawls writes in Political Liberalism, ‘I give the reasonable a more restricted sense [than it is given in Kant].’ He associates it with a readiness ‘to propose principles and standards as fair terms of cooperation, and to abide by them willingly,
given the assurance that others will likewise do so'.\footnote{Note that compliance with these principles is not unconditional, but conditional on the compliance of others. Reciprocity, not the wider notion of impartiality, connects with the idea of the reasonable. Rawls could not be much clearer about this: ‘[T]he idea of impartiality’, he says, ‘is altruistic (as moved by the general good)’.\footnote{However, reasonable agents ‘are not moved by the general good as such but … insist that reciprocity should hold … so that each benefits along with others’}.\footnote{A little later, Rawls puts the point even more forcefully: ‘[T]he reasonable (with its idea of reciprocity) is not the altruistic (the impartial acting solely for the interests of others) nor is it the concern for self (and moved by its ends and affections alone)’.\footnote{But when Rawls’s theory is neither one of justice as impartiality nor one of justice as mutual advantage, what is it then?}} As should be clear by now, Rawls advances a theory of justice as reciprocity, with reciprocity being ‘a moral idea situated between impartiality … on the one side and mutual advantage on the other’.\footnote{We must conclude that Gibbard had every reason to question Barry’s association of justice as impartiality and justice as mutual advantage with Rawls. As Rawls notes himself: ‘Barry thinks justice as fairness hovers uneasily between impartiality and mutual advantage, where Gibbard thinks it perches between on reciprocity. I think Gibbard is right about this’.\footnote{This is also evident from Rawls’s principles of justice that apply to individuals (as opposed to the principles of justice that apply to institutions, which are the two principles of justice). One of these principles is the principle of fairness. It holds that one is required to do one’s part as defined by the rules of an institution when, firstly, this institution is just, that is, it satisfies the two principles of justice, and secondly, one has taken advantage of the opportunities it offers to further one’s interests.\footnote{What emerges from this principle is that we acquire obligations by our voluntary acts, whereby these acts ‘may be the giving of express or tacit undertakings, such as promises and agreements, but they need not to be, as in the case of accepting benefits’.\footnote{Already in ‘Justice as Fairness’ we read that it is an ‘unfortunate mistake’ of proponents of the idea of a social contract to suppose that political obligation requires ‘a deliberate performative act in the sense of a promise, or contract’.\footnote{Rather, ‘it is sufficient that one has knowingly participated in and accepted the benefits of a practice acknowledged to be fair’.\footnote{This shows that Rawls has essentially the same account of obligation as Hobbes. The idea that promises oblige, falling under Hobbes’s third law of nature, and the idea that benefits oblige, falling under Hobbes’s fourth law of nature, are brought under a single principle, the principle of fairness.\footnote{This principle ultimately amounts to an idea of reciprocity: ‘We are not to gain from the cooperative labors of others without doing our fair share’,\footnote{as this would result in the weakening of the cooperative venture (or, in the case of Hobbes, in the exclusion from present and future cooperative ventures).\footnote{With Rawls’s definition of society as a ‘cooperative venture for mutual advantage’, we must expect to be bound by obligations of reciprocity in our daily lives.\footnote{For Rawls, as for Hobbes, justice as reciprocity is all but a bounded theory.}}}}}}}}}}}}
Having considered advocates of all three traditions of formal justice, we can now fill in the rectangles in Table 2. Let us first determine the place of Hobbes and Rawls in this table. As mentioned above, Hobbes’s second law of nature requires that people be willing to lay down their right of nature when others are so too. Similarly, for Rawls, the reasonable demands that we act on certain principles of justice, provided others can be relied on to do the same. ‘If we cannot rely on them’, Rawls tells us, ‘then it may be irrational or self-sacrificial to act from those principles’. Likewise, ‘if other men will not lay down their right’, Hobbes says, ‘then there is no reason for any one, to divest himself of his: for that were to expose himself to prey’. In that case, the laws of nature oblige only in foro interno. Equally, for Rawls, when others cannot be relied on to reciprocate, ‘the reasonable may be suspended’, although it ‘always binds in foro interno, to use Hobbes’s phrase’. What this shows is that for Hobbes and Rawls there will be no situation in which it is rational for only one player to cooperate. Either both players cooperate, or neither one does. Accordingly, the names of Rawls and Hobbes can be included in the upper left and lower right rectangles.

Barry joins Rawls and Hobbes in the upper left rectangle by virtue of his claim that the motive for just behavior is a ‘desire to act justly’. But this claim also puts him in the upper right rectangle, for Barry provides no reason, at least not in his earlier work, why people should make their desire dependent on the desire of others. Finally, the fool appears in the lower left rectangle by virtue of his claim that it may not stand against reason to free-ride on the compliance of others. Of course, the fool would never permit others to free-ride on his own compliance, which puts him also in the lower right rectangle.

What do these findings hold for scholarship on Hobbes and Rawls? For one thing, they cast doubt on the general perception of Hobbes and Rawls as spiritual opponents. Jodi S. Kraus, for example, places Hobbes and Rawls on opposite ends of a continuum ranging from theories that are minimally idealized to theories that are highly idealized. By moving the two philosophers to the center, this paper helps to make their work more relevant to a world that, arguably, is neither the self-centered egoistic world of the realists nor the self-sacrificing altruistic world of some idealists. Indeed, with their notion of reciprocity, Hobbes and Rawls are part of the rich philosophical tradition stretching from ancient times (Plato and Aristotle), to the enlightenment (Joseph Butler and Francis Hutcheson), to more recent times (Thomas Hill Green and Max Weber) that has grounded morality in rationality.
ACKNOWLEDGEMENTS

I would like to thank Nathan Widder, Royal Holloway University of London, for his help and advice.

NOTES

1. Chris Brown, “The Construction of a “Realistic Utopia”: John Rawls and International Political Theory’, Review of International Studies 28, no. 1 (2002): 5–21, 9.
2. Rosamond Rhodes, ‘Reading Rawls and Hearing Hobbes’, The Philosophical Forum 33, no. 4 (2002): 393–412.
3. Joseph Grcic, ‘Hobbes and Rawls on Political Power’, Ethics & Politics 9, no. 2 (2007): 371–92.
4. There are, needless to say, several volumes on political theory, and social contract theory in particular, that feature essays on Hobbes and Rawls, among many other thinkers. Most of these volumes have a brief introductory chapter that tries to put these thinkers into conversation. See, for example, David Boucher and Paul Kelly, The Social Contract from Hobbes to Rawls (London: Routledge, 1994); Andrew Levine, Engaging Political Philosophy from Hobbes to Rawls (Oxford: Blackwell, 2002); and Mark E. Button, Contract, Culture, and Citizenship: Transformative Liberalism from Hobbes to Rawls (University Park: Pennsylvania State University Press, 2008).
5. Grcic, ‘Hobbes and Rawls on Political Power’, 379.
6. Amartya Sen, The Idea of Justice (London: Penguin Books, 2009), 206.
7. Regarding the former, consider Grcic’s finding that both Hobbes and Rawls make use of the concept of the state of nature (Grcic, ‘Hobbes and Rawls on Political Power’, 379), or Rhodes’s discovery that in both cases the state of nature is a hypothetical construct (Rhodes, ‘Reading Rawls and Hearing Hobbes’, 397–402). With regard to the latter, consider Rhodes’s argument that Rawls’s ‘spirit of compromise’ and ‘readiness to meet others halfway’ resemble Hobbes’s sixth law of nature, stating ‘that upon caution of the Future time, a man ought to pardon the offenses past of them that repenting, desire it’, and the spirit of his seventh law, prescribing ‘that in Revenges, Men look not at the greatness of the evill past, but the greatnesse of the good to follow’ (ibid., 403).
8. Ibid., 395.
9. Ibid., 394, my emphasis.
10. Ibid., 395–6.
11. Terry Nardin, Law, Morality and the Relations of States (Princeton, NJ: Princeton University Press, 1983), 259.
12. Thomas Hobbes, Leviathan (Oxford: Oxford University Press, 1998), 99–100; and John Rawls, A Theory of Justice (Oxford: Oxford University Press, 1973), 108–9.
13. Nardin, Law, Morality and the Relations of States, 259.
14. Hobbes, Leviathan, 99–100; and Rawls, A Theory of Justice, 108–9.
15. Nardin, Law, Morality and the Relations of States, 266–9.
16. Ibid., 259.
17. Brian Barry, Theories of Justice (London: Harvester Wheatsheaf, 1989).
18. Allan Gibbard, ‘Constructing Justice’, Philosophy and Public Affairs 20, no. 3 (1991): 264–97.
19. Brian Barry, Justice as Impartiality (Oxford: Clarendon Press, 2002).
20. The three traditions roughly parallel David Gauthier’s account of the prudent, the trustworthy, and the prudent but trustworthy man (David Gauthier, ‘Morality and Advantage’. The Philosophical Review 76, no. 4 (1967): 460–75).
21. Barry, Theories of Justice, 7.
22. Ibid., 51.
23. Ibid., 363.
24. Ibid., 361.
25. Ibid., 363.
26. Barry, _Justice as Impartiality_, 31.
27. Hobbes, _Leviathan_, 89.
28. Ibid., 89.
29. Ibid., 105, italics in original.
30. Ibid., 91.
31. Ibid., 105.
32. Rawls, _A Theory of Justice_, 269, figuring most prominently among those, holds Hobbes's state of nature to be the 'classical example' of the prisoner's dilemma. Others arguing in this vein include Brian Barry, _Political Argument_ (London: Routledge, 1965); Michael Taylor, _Anarchy and Cooperation_ (London: Wiley, 1976); and Levine, _Engaging Political Philosophy from Hobbes to Rawls_.
33. Hobbes, _Leviathan_, 96.
34. Aloysius P. Martinich, _Hobbes_ (London: Routledge, 2005), 103.
35. Hobbes, _Leviathan_, 97.
36. Ibid.
37. Ibid.
38. See Gregory Kavka, _Hobbesian Moral and Political Theory_ (Princeton, NJ: Princeton University Press, 1986), 137–56, 387, 405; Edwin Curley, 'Introduction to Hobbes' _Leviathan', in Thomas Hobbes: _Leviathan_, ed. Edwin Curley (Cambridge: Hackett, 1994), xxvi–xxvii; Jean Hampton, _Hobbes and the Social Contract Tradition_ (Cambridge: Cambridge University Press, 1995), 64–5; and Noel Malcolm, _Aspects of Hobbes_ (Oxford: Clarendon Press, 2002), 438. For Hobbes, it is a matter of accident that a person violating his or her covenant is not discovered. He consequently cannot deny that there are situations in which a person can achieve great gain by violating his or her covenant. What he can deny, though, is that a person can with right reason expect that violation of covenant will conduce to his or her gain (David Gauthier, _The Logic of Leviathan_ (Oxford: Clarendon Press, 1969), 84–5). As Hobbes, _Leviathan_, 97, puts it: '[W]hen a man doth a thing, which notwithstanding anything can be foreseen and reckoned on tendeth to his own destruction, howsoever some accident, which he could not expect, arriving may turn it to his benefit; yet such events do not make it reasonably or wisely done'.
39. Curley, 'Introduction', xxiv–xxviii.
40. Hobbes, _Leviathan_, 87.
41. Ibid., 95. It is also implicit in the first law—the fundamental law of nature from which the second law is derived—holding 'that every man, ought to endeavour peace, as far as he has hope of obtaining it; and when he cannot obtain it, that he may seek, and use, all helps, and advantages of war' (Hobbes, _Leviathan_, 95), and in the negative form of the golden rule, which summarizes all nineteen laws of nature: 'Do not that to another, which thou wouldest not have done to thyself' (Ibid.).
42. Ibid., 95.
43. Ibid., 100.
44. Thomas Hobbes, _On the Citizen_ (Cambridge: Cambridge University Press, 2003), 159.
45. Howard Warrender, _The Political Philosophy of Hobbes_ (Oxford: Clarendon Press, 1957), 235. Or, in the words of Barry, _Theories of Justice_, 465, Hobbes is moving from justice as fidelity, that is, carrying out one's side of a bargain voluntarily entered into, to justice as requital, that is, making a fair return for benefits received.
46. Hobbes, _Leviathan_, 90.
47. Rawls, _A Theory of Justice_, 347.
Barry, *Justice as Impartiality*, 51.

Barry (ibid., 45–6) answers in the affirmative but shows that justice as mutual advantage suffers from several drawbacks.

John Rawls, *Justice as Fairness: A Restatement* (London: Harvard University Press, 2001), 118–19.

John Rawls, *Political Liberalism* (New York: British Columbia Press, 1993), 48; and Rawls, *Justice as Fairness*, 82.

John Rawls, ‘Justice as Fairness’, in *John Rawls: Collected Papers*, ed. Samuel Freeman (London: Harvard University Press, 1999), 47–72.

Ibid., 54, my emphasis.

Ibid., 54, my emphasis.

Catherine Audard, *John Rawls* (Stocksfield: Acumen, 2007), 77.

Rawls, *Political Liberalism*, 49.

Ibid., 49, my emphasis.

Ibid., 50.

Ibid.

Ibid., 54.

Ibid., 77.

Ibid., 17. After being criticized by Gibbard and Rawls for not having identified justice as reciprocity as a distinctive alternative, Barry, in *Justice as Impartiality*, tries to show that what he meant by justice as mutual advantage and justice as impartiality in *Theories of Justice* is in line with justice as reciprocity: ‘[M]y account in *Theories of Justice* . . . of the aspect of Rawls’s theory that I claimed to fall under the theory of justice as mutual advantage was intended to fit in with the idea that the motive for compliance would be a sense of justice as ‘fair play,’ i.e. a duty to do one’s part to sustain a mutually advantageous institution. I regret not making this more explicit’ (Barry, *Justice as Impartiality*, 48). This claim is quite astonishing, as it is only in the conclusion of *Theories of Justice* that Barry sketches a ‘sophisticated version’ of justice as mutual advantage under which justice ‘consists in playing one’s part in mutually advantageous cooperative arrangements’ (Barry, *Theories of Justice*, 361). On the other hand, justice as impartiality, Barry now argues, ‘has the structure of an assurance game. If I am motivated by a desire to behave fairly, I will want to do what the rules mandated by justice as impartiality require so long as enough other people are doing the same’ (Barry, *Justice as Impartiality*, 51). Note the two fundamental changes Barry makes to his prior statement of justice as impartiality: firstly, he no longer speaks of a desire to act justly, as in *Theories of Justice*, but of a desire to behave fairly—a term that Rawls uses to describe his theory of justice as reciprocity—and secondly, this desire now exists only as long as enough other people have the same desire—a condition not to be found in *Theories of Justice*. But when both justice as mutual advantage and justice as impartiality articulate an idea of fair play and reciprocity, one is left to wonder why Barry, in *Theories of Justice*, defines justice as impartiality ‘more or less as the obverse’ of justice as mutual advantage (Barry, *Theories of Justice*, 361).

Rawls, *A Theory of Justice*, 111–12.

Ibid., 113.

Ibid., 60.

Ibid.

This, in fact, is a point that Rawls, ‘Justice as Fairness’, 61, criticizes: ‘Hobbes, . . . when invoking the notion of a ‘tacit covenant,’ appeals not to the natural law that promises should be kept [the third law of nature] but to his fourth law of nature, that of gratitude. . . . While it is not a serious criticism of Hobbes, it would have improved his argument had he appealed to the duty of fair play [the third law of nature]. On his premises he is perfectly entitled to do so’.
Rawls’s argument against defection can be illustrated by David Hume’s famous example of two people rowing a boat that neither can row alone: if only one person fails to do his or her part, the cooperative venture collapses with both persons worse off. As a result, each person has an incentive to row when the other person rows (David Hume, *A Treatise of Human Nature* (Oxford: Clarendon Press, 1978), 490). Hobbes’s argument against defection, on the other hand, can be exemplified by slightly modifying Hume’s example: in this case it is possible to row the boat with one manpower alone. When there are two people in the boat, both rowing, the boat goes faster than if there was only one person in the boat. Meanwhile, when there are two people in the boat, one rowing and the other not rowing, the boat, because of the additional burden, goes slower than if there was only one person in the boat. The person rowing then has an incentive to throw the person not rowing overboard.

Hobbes, *Leviathan*, 87, my emphasis.

Rawls, *A Theory of Justice*, 53–54, my emphasis.

Rawls, *A Theory of Justice*, 54, my emphasis.

Hobbes, *Leviathan*, 87, my emphasis.

Hobbes, *Leviathan*, 105, italics in original.

Rawls, *A Theory of Justice*, 54, italics in original.

Barry, *Theories of Justice*, 363.

Jodi S. Kraus, *The Limits of Hobbesian Contractarianism* (Cambridge: Cambridge University Press, 2002), 30–1.

David Gauthier, *Morality and Rational Self-Interest* (Englewood Cliffs: Prentice-Hall, 1970), 111.