Oxford Studies in Political Philosophy

Volume 7

Edited by
DAVID SOBEL, PETER VALLENTYNE,
AND STEVEN WALL

OXFORD
UNIVERSITY PRESS

8

Why Strict Compliance?

Simon Căbulea May

Introduction

John Rawls (1971/1999) defends the principles of justice as fairness as the appropriate terms of association for a well-ordered society. He presents this defense as an exercise in ideal theory (1971/1999, 8). Ideal theory has two key features in his account. First, it rests on a strict compliance assumption: the institutions of a well-ordered society are perfectly just and every person acts as justice requires in upholding these institutions. Second, ideal theory takes priority over non-ideal theory insofar as it underpins the moral norms that govern responses to injustice in societies that are not well ordered.[1] The combination of these two features creates a puzzle: why should moral norms for responding to the widespread injustice of the actual world depend on an account of a hypothetical world in which every agent complies with the requirements of justice? Rawls's approach may seem excessively perfectionist. Solutions to the urgent problems of everyday life do not generally rely on appeals to any standard of perfection. Moral arguments for particular tax codes, high-way speed limits, and workplace safety regulations, for instance, should not presume that every relevant agent will comply with the rules once they are implemented. So why should claims about how to build a more

---

[1] Rawls also defines ideal theory in terms of an assumption of favorable circumstances (1971/1999, 216). Although this is an important element of Rawlsian ideal theory, my present argument concerns only the strict compliance assumption. I do not discuss alternative concepts of ideal theory that do not employ a strict compliance assumption. See Valentini (2012) for discussion of various accounts of ideal theory. Nor do I discuss the full array of issues associated with Rawlsian ideal theory. For a comprehensive discussion of these issues, see Stemplowska and Swift (2014).

just world ultimately depend on any assumption of strict compliance in a perfectly just society?

In this chapter, I present a distinctly Rawlsian interpretation of ideal theory, one that provides a response to the strict compliance puzzle.[2] This *cooperative interpretation* is grounded in the idea of society as a fair scheme of cooperation, which Rawls (2001, 5) describes as the most fundamental idea of justice as fairness. A key element of the Rawlsian idea of cooperation, I claim, is that the individual participants of a genuinely cooperative scheme—whatever its scale—are morally account-able to each other for complying with the scheme's rules. This means that each participant has the moral standing to demand of the others that they comply with the rules. I argue that the logic of these moral demands requires that the scheme's rules be worked out on the basis of a strict compliance assumption. The ideal theory of cooperation takes priority over non-ideal theory because responses to unfairness presuppose some conception of fair cooperation between agents who comply with the rules of the scheme. In justice as fairness, society as a whole is a grand scheme of cooperation. The principles of justice constitute the moral terms of association for this cooperative scheme, and hence define the moral demands that citizens, as such, may make of one another. Thus, these principles of justice must likewise be worked out on the basis of a strict compliance assumption. Once the principles of justice have been estab-lished, moral norms for responding to injustice can be developed.

The cooperative interpretation is *deontic* insofar as it is grounded in the way that the participants of a genuinely cooperative scheme are obligated to follow its rules.[3] The most obvious interpretation suggested by Rawls's own statement of justice as fairness, in contrast, is *telic*. On this standard interpretation, the function of ideal theory is to specify a political telos—a realistic utopia—that should orient political decision

---

[2] The interpretation is distinctly Rawlsian in the sense that it depends on certain basic ideas that are either an explicit part of or closely related to Rawls's broader statement of justice as fairness. I do not claim that the interpretation is in fact Rawls's own view of ideal theory.

[3] The cooperative interpretation is only one possible deontic interpretation of ideal theory. Blain Neufeld (2017) argues that ideal theory is required by a principle of equal civic respect, a principle that underpins the idea of public reason. Despite some important similarities with Neufeld's view, the cooperative interpretation does not in itself support the argument for public reason. See also Laden (2013).

making in the actual world. Ideal theory takes priority over non-ideal theory because it identifies the ultimate end to be realized, and this end must be identified before the appropriate means to it can be determined. Ingrid Robeyns aptly captures the spirit of this telic interpretation:

> Ideal theory functions as a mythical *Paradise Island*. We have heard wonderful stories about Paradise Island, but no one has ever visited it, and some doubt it truly exists. We have a few maps that tell us, roughly, where it should be situated, but since it is in the middle of the ocean, far away from all known societies, no one knows *precisely* where it is situated. Yet we dream of going there, and ask ourselves how we could get there, and in which direction we should be moving in order to eventually reach Paradise Island. . . . We don't know whether it can be reached and no one has ever set foot on the Island. Yet since it is our dream to go there, reaching Paradise Island is our ultimate goal.
>
> (2008, 344–45)

My aim in this chapter is not to criticize the telic interpretation, but rather to establish the cooperative approach as a viable alternative response to the strict compliance puzzle. On the cooperative interpretation, the idea of a well-ordered society in justice as fairness may play a telic role, but only because it is already a quite specific substantive ideal of social cooperation. The strict compliance assumption is ultimately warranted, not because the Rawlsian well-ordered society serves as the telos for political decision making, but because this society is a grand scheme of fair cooperation between citizens who have standing to press demands of justice against one another.

The chapter contains six sections. The first four sections present the cooperative interpretation of ideal theory. Section I outlines Rawls's cooperative conception of society. Section II presents an account of the nature of moral demands that participants in a scheme of cooperation may make of one another. Section III explains why a moral theory of cooperation, so conceived, must incorporate a strict compliance assumption. Section IV extends the argument to the political domain of justice and the idea of a well-ordered society. The last two sections contrast the cooperative interpretation with the standard telic approach. Section V

discusses an anti-utopian objection to the telic interpretation of ideal theory and explains why this objection does not threaten the cooperative alternative. Section VI explains how the cooperative interpretation allows for the possibility of a non-utopian ideal theory. A moral theory of cooperation need not assume that the cooperative scheme in question is anything more than a means to some other end. Paradise Island does not have to exist for the strict compliance assumption to be warranted. And even if the ideal theory of justice as fairness is ultimately utopian, it is not true that it is ideal because it is utopian.

I. The Cooperative Conception of Society

Rawls presents justice as fairness as an alternative to utilitarianism and draws several contrasts between the two doctrines. Some of these contrasts concern the different institutional implications the doctrines have, such as for the protection of individual liberties and the structure of a just economic system. These differences are important in determining just public policy, but the deepest philosophical contrast between the two doctrines is the difference in their underlying conceptions of society. Rawls claims that a conception of justice is part of an overall social ideal:

> A social ideal in turn is connected with a conception of society, a vision of the way in which the aims and purposes of social cooperation are to be understood. The various conceptions of justice are the outgrowth of different notions of society against the backdrop of opposing views of the natural necessities and opportunities of human life. Fully to understand a conception of justice we must make explicit the conception of social cooperation from which it derives.   (1971/1999, 11)

Justice as fairness and utilitarianism derive from quite different normative conceptions of society (1971/1999, 29–30). Justice as fairness conceives of a well-ordered society as a scheme of cooperation between free and equal persons for their reciprocal advantage. Utilitarianism, in contrast, conceives of a well-ordered society as one in which social interactions and resources are efficiently administered to maximize the

satisfaction of desires. Rawls's normative conception of society provides the first premise for the cooperative interpretation of ideal theory:

(1) A just political society constitutes a grand scheme of social cooperation between free and equal persons for reciprocal advantage.

As it stands, the cooperative conception of society is a somewhat nebulous idea. Two steps help specify its meaning more determinately: a fuller account of the contrast with the utilitarian administrative conception, and an explanation of the idea of genuine social cooperation it incorporates.

1. Essential Plurality and Relational Constraints

The contrast between the cooperative and administrative conceptions of society finds expression in at least two ways. First, justice as fairness is essentially pluralistic in the sense that it requires that the content of the principles of justice depend on their role in setting the terms of association between a plurality of individuals with competing interests. Utilitarianism can allow that the concept of justice is defined by the role of the principles in determining the appropriate distribution of benefits and burdens across a group of people. So, for both doctrines, the problem of justice is essentially pluralistic. But the utilitarian solution to this problem is not. Rawls claims utilitarianism "does not take seriously the distinction between persons" because it extends the principle of rationality for a single person to society as a whole (1971/1999, 24). The utilitarian legislator regards individuals in much the same way as an entrepreneur regards her various investments: what matters is how the allocation of rights and duties to these individuals affects the bottom line; that is, maximal aggregate utility. The principle of utilitarian legislation is the same whether or not it is the interests of a single individual at stake or those of many millions of people.

Rawls's "separateness of persons" objection characterizes utilitarianism as, in one sense, insufficiently individualistic (1971/1999, 26). But Rawls also claims that the doctrine's use of the principle of rationality for

a single person demonstrates that it is, in another sense, overly individualistic:

> [The] assimilation of justice to a higher order executive decision ... is central to classical utilitarianism; and it also brings out its profound individualism in one sense of this ambiguous word. It regards people as so many *separate* directions in which benefits and burdens may be assigned; and the value of the satisfaction or dissatisfaction of desire is not thought to depend in any way on the moral relations in which individuals stand, or on the kinds of claims which they are willing, in the pursuit of their interests, to press on each other.
>
> (Rawls 1958, 187; 1999a, 66)

The second contrast between the cooperative and administrative conceptions of society, then, is that the principles of justice as fairness depend on the nature of the proper moral relations between individual citizens. The idea here is not simply that principles of justice should have a certain substantive content that defines what these relations should be. On that reading, citizens' relations would be purely derivative of the principles, so there would be no reason in utilitarianism why they should constrain the value of desire satisfaction. Rather, Rawls means some account of the moral relations between citizens serves as a constraint on the principles of justice. Refer to this feature of Rawls's view as a *relational constraint*.

A relational constraint entails that the moral principles governing some sphere of social life depend on an account of the appropriate moral relations between the people in that sphere. Consider, for example, the problem of determining which moral principles govern the domestic life of a family. These principles plausibly depend on assumptions about the appropriate relations between members of the same family: parents and children have associative obligations of care toward each other that they do not have toward outsiders. Similarly, principles of family life should accommodate and legitimate the love and affection family members are normally expected to feel for each other: upsetting one's parents matters in a way that upsetting the other members of a student dormitory does not, precisely because they are one's parents and not merely

fellow residents of a shared accommodation. But justice as fairness cannot incorporate a relational constraint in the exact same way as the ethics of family life. This is because the Rawlsian well-ordered society is not a community dedicated to the pursuit of some overarching conception of the good, much less a family united by the bonds of love. There is no place in justice as fairness for the idea that relations between citizens should mirror the thick ethical relations within families or other social associations. Justice as fairness may invoke only a thin account of citizens' moral relations. The resources for this thin theory must be provided by the idea of social cooperation between free and equal persons in itself.

2. Genuine Social Cooperation

Rawls distinguishes genuine social cooperation from socially coordinated activity. Both involve schemes of social interaction, but they differ in two key respects, each of which underpins a relational constraint. The first respect is an explicit feature of justice as fairness. Rawls claims social cooperation embodies certain values, whereas mere coordinated activity need not: "The idea of cooperation includes the idea of fair terms of cooperation [and these] specify an idea of reciprocity, or mutuality: all who do their part as the recognized rules require are to benefit as specified by a public and agreed-upon standard" (2001, 6). Genuine social cooperation is fair and reciprocal, whereas efficiently coordinated activity need not be. The first relational constraint, then, is that the principles of justice are justified only if they constitute fair and reciprocal relations between the citizens of a well-ordered society. Thus, Rawls defends his difference principle in part on the grounds that it better expresses the value of reciprocity than a utilitarian alternative (1971/ 1999, 64, 76).[4]

---

[4] See also Rawls (2001, 96): "Between [the social contract and utilitarian] traditions there is a basic contrast: the idea of society as a fair scheme of cooperation is quite naturally specified so as to include the ideas of equality . . . and reciprocity (of which the difference principle is an example). By contrast, the idea of society organized to produce the most good expresses a maximizing and aggregative principle of political justice. In utilitarianism, the ideas of equality

The second respect in which genuine social cooperation differs from socially coordinated activity is less explicitly developed in Rawls's state-ment of justice as fairness, but it is pivotal to the cooperative interpret-ation of ideal theory. The participants of a cooperative scheme, I claim, are mutually accountable to each other in the sense that they have standing, as participants, to make demands of one another in accordance with the rules of the scheme. The rules define not only the appropriate content of the demands but also the nature of the standing the partici-pants have to press them. Socially coordinated activity, in contrast, does not require any such relations of mutual accountability. A central authority could impose various rules on the participants and take responsibility for enforcing those rules. The rules are sent out by the authority to each participant and it is to this authority alone that the members are accountable. In essence, its relations of accountability are fundamentally vertical, rather than horizontal.[5]

To illustrate the idea of relations of mutual accountability, consider the difference between two normative conceptions of the firm. In the first conception, the firm's workers determine the rules governing their joint labor and, crucially, are ultimately responsible to each other for complying with these rules. This basic relation of mutual accountability exists even if the workers should choose to employ various hierarchical structures for the sake of efficiency. In the second conception, the firm's management determines the rules governing the workers' labor and it is to management alone that the workers are accountable. No basic relation of mutual accountability between the workers exists, even as the management might implement various team-building exercises and peer evaluation mechanisms. The managerial firm fails to be a genuinely cooperative scheme because its members lack standing, just in virtue of being part of

and reciprocity are accounted for only indirectly, as what is thought to be normally necessary to maximize the sum of social welfare."

[5] The idea of a central coordinating authority here resembles Joel Feinberg's (1970) idea of a sovereign rights monopoly, but the two ideas are importantly distinct. Feinberg's concern is with individuals' standing to demand the performance of duties *owed to them*, since this standing is part of what explains the value of rights. But not every duty a person has standing to demand is a duty that is owed to or directed toward that person. Both standing and direction are important components of claim-rights, but they are distinct. I discuss the concept of directed duties in May (2015) and provide an account of the difference between standing and direction in a separate paper project.

the firm, to demand that the other members comply with its rules. It is instead a more or less complicated form of socially coordinated activity.

The second relational constraint, then, is that the principles of justice are justified only if they establish a fundamental horizontal relation of moral accountability between citizens. Utilitarianism does not satisfy this constraint since it does not explain why it is justifiable for citizens to have the moral standing to demand sacrifices of each other in pursuit of the goal of maximal desire satisfaction. The doctrine only claims society should be efficiently arranged so that these sacrifices are in fact made. The form relations of accountability take in a utilitarian society is, accordingly, entirely derivative of the principle of utility.

In contrast to utilitarianism, justice as fairness incorporates mutual accountability as a fundamental constraint on principles of justice. As an example, consider the criterion of reciprocity Rawls invokes in his defense of political liberalism. According to this criterion, proposed terms of association are fair only if their advocates may reasonably expect other citizens to reasonably accept them.[6] On one articulation of the criterion, it is enough that the advocates *sincerely believe* other citizens can accept the terms "as free and equal, and not as dominated or manipulated, or under the pressure of an inferior political or social position" (2005, xlii). But on this articulation, the criterion does little to constrain the terms of association, since optimistic citizens could sincerely believe (even if, perhaps, incorrectly) that others might one day be persuaded to accept terms of association grounded in a comprehensive conception of the good. A sharper articulation of the criterion is that proposed terms are fair only if it is reasonable for their advocates to expect of other citizens that they accept the terms. Here, the reasonable expectation is normative rather than descriptive—it is not merely a sincere belief about what others might endorse, but a *justifiable demand* that they accept the terms in question. On this articulation, the criterion has much greater potential to constrain fair terms of association. Sincere

---

[6] Rawls's liberal principle of legitimacy applies the criterion of reciprocity to the political structure: "the exercise of political power is fully proper only when it is exercised in accordance with a constitution the essentials of which all citizens as free and equal may reasonably be expected to endorse" (2005, 137). I criticize the liberal principle as a conception of political legitimacy in May (2009).

beliefs do not themselves require moral justification—they are either true or false, supported by the evidence or not, and so on—but relations of mutual accountability do. Comprehensive conceptions of justice plausibly violate the criterion, given the burdens of judgment (Rawls 2005, 54–58), because it is not morally justifiable to demand of other citizens, on the political liberal view, that they endorse any particular comprehensive moral doctrine. Since citizens are not accountable to each other, as citizens, for endorsing a comprehensive doctrine, they cannot be accountable to each other for accepting any conception of justice that depends on such a doctrine.

II. Moral Demands in Cooperative Schemes

I have claimed that the cooperative conception of society that underpins justice as fairness includes the idea that participants in a fair scheme of cooperation are bound together by horizontal relations of mutual accountability: they all have standing to demand of each other that they comply with the rules governing the scheme. This provides the second premise for the cooperative interpretation of ideal theory:

(2)   A scheme of genuine social cooperation exists only when individual participants are morally accountable to each other in that they each have standing to demand that the other participants of the scheme comply with its rules.

What does a horizontal relation of mutual accountability add to the basic idea that the scheme ought to be effectively governed by the rules in question? In this section, I advance an account of the nature of moral demands in cooperative schemes in response to this question.

To fix ideas, consider an example where one person plausibly lacks moral standing to demand that another perform her moral duty:

Strangers on a Train
Anne and Bruno are two strangers sitting next to each other on a train. Anne is talking on her phone, describing in great detail her plans to ruin her

friend Guy's romantic relationship by spreading deceitful rumors. Bruno follows her conversation for some time and forms the correct belief that Anne has a moral duty to not carry through with her malicious meddling.

I assume here that Bruno has no moral standing to demand that Anne abandon her plans, despite the truth of his belief. It would be out of place, in some sense, for him to turn to her and insist, "Anne, you must not ruin Guy's romantic relationship!" Her immoral meddling is, intuitively, none of his moral business.[7] The reason for this lack of standing is not epistemic. Bruno might at first lack sufficient information about Anne's friendship with Guy to know whether any moral considerations could justify her action. But as the conversation progresses, it can become apparent the deceit is entirely unjustified. Nor is the problem that Bruno's assertion would be a separate action from whatever Anne chooses to do, and hence subject to moral evaluation itself. Bruno does not lack standing because his assertion would be a rude intrusion into her private conversation. Anne might invite him to offer his honest opinion, in which case his comment would not be rudely intrusive. Yet Anne would still not be morally accountable to him for her actions. Three proposition types can accordingly be distinguished:

   (i) X has a moral duty to ϕ.
  (ii) Y may justifiably assert that X has a moral duty to ϕ.
 (iii) Y has moral standing to demand that X ϕ.

The first two propositions do not imply the third because they do not imply anything about the moral relationship between the two people. The possession of moral standing in (iii) adds something to the facts in (i) and (ii). But there are at least two accounts of what this additional element is: a normative power or a normative liability.[8]

---

[7] Not everyone may share this intuition. Readers who believe that Bruno does have the moral standing to address this demand to Anne are invited to substitute their preferred example of a lack of standing.

[8] Here I employ a broad concept of a normative power as the ability to change an agent's normative position by intentionally performing some action. An agent has a normative liability when some other agent has a normative power with respect to her.

The first account claims that Y's moral standing in (iii) entails that he has the normative power to impose on X (or otherwise modify) a moral duty to ϕ by demanding that she ϕ. This *power theory* of moral demands explains Bruno's lack of standing in *Strangers* by claiming that his assertion ("Anne, you must not ruin Guy's romantic relationship!") would not change her moral situation. She would not become even more duty-bound to not spread deceitful rumors in virtue of Bruno's utterance. This is correct, as far as it goes, but the power theory is an inadequate theory of the standing to demand. The problem with the theory is that it places too much emphasis on the act of issuing a demand itself. A demand does sometimes involve a normative power to change its addressee's moral position. If an officer demands that a soldier retrieve some equipment, then the soldier *ipso facto* becomes duty-bound to do so. In this case, the demand is a command the officer has the authority to issue. But not all demands are commands. In many cases, demands forcefully communicate the addressee's existing duty, but do not add any force or substance to that duty. For instance, Anne has the moral duty to not punch Bruno, and he has the moral standing to demand that she comply with this duty. But no utterance of his to that effect would intensify (or otherwise change) her duty to not punch him.

The second account claims Y's standing in (iii) entails that he possesses a normative liability with respect to X and at least one of his existing duties: if X were to violate her duty to ϕ, she would thereby suspend, *ceteris paribus*, Y's duty to ψ.[9] This *liability theory* of moral demands explains Bruno's lack of standing in *Strangers* by claiming Anne's plans to ruin Guy's relationship leave Bruno's moral position with regard to her unchanged in a particular way: he does not become permitted, even *ceteris paribus*, to sanction Anne for her conduct. In broad terms, a sanction is some form of negative treatment imposed on a person to enforce compliance with a duty, negative treatment that it is ordinarily impermissible to impose. Thus, Y's duty to ψ is his ordinary or default duty to refrain from sanctioning X. Refer to the combined duties Y has with regard to X when his duty not to sanction her remains in force

---

[9] I briefly defend this account of the standing to demand in May (2013).

as his *default duty of normal cooperation* with X.[10] Bruno lacks standing to demand that Anne discontinue her meddling in Guy's relationship because her actions do not suspend his default duty of normal cooperation with her. He may think worse of her, decide not to be especially friendly toward her, or complain about her to his friends after the train journey. But these would not constitute departures from normal cooperation because they would not constitute sanctions it would ordinarily be impermissible for him to impose.

The key idea in the liability theory of moral demands is that when X is morally accountable to Y there is a link—a kind of tripwire—between one of X's moral duties and Y's default duty of normal cooperation: X's violation of her duty triggers, *ceteris paribus*, the suspension of Y's duty to not impose a sanction on her. Y's moral standing to demand that X ϕ consists in the fact that X's violation of her duty to ϕ would have this normative upshot for his moral position. A relation of mutual accountability exists when two agents are morally accountable to each other, although not necessarily in exactly the same way. Consider a revised example of the *Strangers* example, one where a relationship of moral accountability does plausibly exist:

Fellow Passengers on a Train

Anne and Bruno are fellow passengers sitting next to each other *in the quiet car* of a train. Anne is talking on her phone, describing in great detail her plans to ruin her friend Guy's romantic relationship by spreading deceitful rumors. Bruno follows her conversation for some time and forms the correct belief that Anne has a moral duty to end her phone call.

In this example, Anne has a duty to end her phone call because she is in the quiet car, where phone calls are explicitly forbidden. Talking on the phone in the quiet car is wrong, at least insofar as other passengers might be disturbed. As things go, however, it is not an especially egregious wrong. Anne's interference in Guy's romantic life is a far graver misdeed

---

[10] I assume for simplicity that different duties can be combined into a single conjunctive duty.

than her lack of consideration for her fellow passengers. Nevertheless, I assume Bruno has the moral standing to demand that she end her conversation, even as he lacks the standing to demand that she end her meddling. Whereas he would ordinarily have some duty not to bother her with an insistent request—it's the quiet car, after all—her behavior changes what he may do. Should she refuse to comply with his request, he would become permitted to take further steps, such as summoning the conductor to reproach her. None of these departures from normal cooperation on Bruno's part would be permissible if Anne were not violating a rule against phone calls. He may not disturb her or cause a fuss just in virtue of the morally objectionable content of her conversation.

The difference between *Strangers* and *Fellow Passengers* can be explained in terms of the cooperative scheme in which Anne and Bruno participate. Passengers on a train find themselves together in a common venture: they must occupy the same physical space for the duration of a journey from one station to another. The passengers cooperate with one another by complying with certain explicit and implicit rules that respond to their shared interests as passengers. These *cooperative rules* constitute the moral terms of their temporary association on the train. Each passenger has, as a participant in the cooperative pursuit of a reasonably pleasant journey, standing to expect of the others that they comply with the terms of their association. One such term, in the quiet car at least, is the rule forbidding phone calls. This rule responds to the passengers' interests insofar as they are assumed to have, in virtue of selecting the quiet car, a strong preference for a peaceful trip and no strong interest in the freedom to converse as they like. But no moral rule forbidding malicious interference in non-passengers' romantic lives would be warranted as a response to their shared interests as passengers. With respect to such extraneous moral matters, the passengers are strangers. With respect to their behavior on the train, they are fellows.

The train examples are simplistic in at least two ways. First, not all moral rules governing train travel are a matter of cooperation with one's fellow passengers. In addition, train travel is a socially coordinated activity. Thus, there are typically rules governing how the passengers

are to behave that do not constitute terms of association between them. Rather, the passengers are only accountable to the train's staff for compliance with these *coordination rules*. For instance, each passenger is obligated to purchase a valid train ticket, just as every passenger in the quiet car is obligated to refrain from phone calls. But whereas the latter obligation stems from both types of rule, the former obligation plausibly stems from a coordination rule alone. On this supposition, Bruno does not have standing to demand that ticketless Anne seek out the conductor to pay her fare.[11] Second, members of the same society are participants in a great many cooperative schemes, with complicated boundaries and frequently nebulous terms of association. Even strangers occupying the same train car are not merely fellow passengers and hence do not only have standing to make moral demands of one another in that capacity. For instance, if Anne were to start beating her infant child, Bruno would have standing to demand that she stop. If she refuses to stop, he becomes permitted to restrain her. This is not primarily because he has an interest as a passenger in not hearing the child's wails. Rather, the relevant association is something deeper: as members of the same society, people may expect of each other that they not violently abuse infants. Bruno would have the same standing if he were walking past Anne and her child on the street.

Identifying the salient terms of association is often difficult—even if there is rough agreement on what moral duties people have, it can be hard to determine the constituency to which people are accountable for performing those duties. Disagreement about the contours of account-ability relationships and the precise terms of normal cooperation gener-ates much of the heat of moral conflict in common life. My present point is not to deny (still less settle) these complexities, but to defend the claim that a relationship of moral accountability with respect to certain duties is something additional to those duties and something that itself requires justification. In particular, moral accountability requires a normative tripwire between one agent's moral duties and another agent's default duties of normal cooperation with her, a tripwire that stands in need of

---

[11] I do not deny that there may be circumstances in which train passengers do have the moral standing to demand that they each purchase a ticket.

moral justification. This point would hold even if it were true—implausibly—that every person is always morally accountable to every other person for all their moral duties.[12]

The liability theory of moral demands provides the third premise for the cooperative interpretation of ideal theory:

(3)   An individual Y has moral standing to demand that another individual X comply with a duty only when X's violation of this duty would trigger, *ceteris paribus*, the suspension of Y's default duty of normal cooperation with her.

III. Strict Compliance in the Moral Theory of Cooperation

I have thus far presented three premises in the cooperative interpretation of ideal theory:

(1) A just political society constitutes a grand scheme of social cooperation between free and equal persons for reciprocal advantage.

(2) A scheme of genuine social cooperation exists only when individual participants are morally accountable to each other in that they each have standing to demand that the other participants of the scheme comply with its rules.

(3) An individual Y has moral standing to demand that another individual X comply with a duty only when X's violation of this duty would trigger, *ceteris paribus*, the suspension of Y's default duty of normal cooperation with her.

In this section, I argue that premises (2) and (3) imply a further claim:

(4) The moral terms of genuinely cooperative schemes depend on a strict compliance assumption.

---

[12] Linda Radzik (2011) raises a similar objection to the universal scope of Stephen Darwall's (2006) account of the second-personal standing to blame.

The argument for premise (4) turns on the interdependence of the duties in genuinely cooperative schemes. It comprises two parts. The first part establishes the logical need for a strict compliance assumption at some point in the justification of the scheme's rules. The second part establishes the justificatory priority of this assumption.

1. Strict Compliance and the Logic of Moral Demands

Premise (2) asserts that the participants of a cooperative scheme have standing to demand of each other that they comply with the scheme's rules. I also assume that if some individuals have standing to demand of each other that they comply with a social rule, then that rule is cooperative and constitutes one of the terms of their association. Other social rules may also govern the individuals' behavior, but without the required relation of mutual accountability, these would not constitute rules of a cooperative scheme that comprises the individuals as participants. Instead, they would be rules of some other scheme of social interaction.

The first part of the argument for premise (4) is quite intricate. It comprises six steps. The first step of the argument posits a link between the individuals' duties and the rules of their cooperative scheme:

(i)   A participant X of a cooperative scheme S has a duty to ϕ, *qua* participant in S, only if some cooperative rule R_1 in S imposes on X a duty to ϕ.[13]

Refer to the duties X has as a member of a cooperative scheme as her cooperative duties. X may have other duties that are not imposed by any of the rules of S and which are therefore not S-cooperative duties. These extraneous duties would be ones she has as a member of another scheme of interaction. They may be related to her S-cooperative duties in complicated ways—they could even be indirectly grounded in her identity as

---

[13] I assume for simplicity that different social rules can be combined into a single conjunctive social rule. Nothing turns on whether there is only one identifiable norm in play or a number of different norms.

a participant in S and hence be duties she has only because of her S-cooperative duties. But if she is not accountable to the other members of S for compliance with these duties, then they are not duties she has *qua* participant in S.

Premise (3) incorporates the liability theory of the standing to demand. If some individual Y has the standing to demand that X comply with her S-cooperative duties, then he must have a default duty of normal cooperation with her—a default duty that is suspended, *ceteris paribus*, when X violates her S-cooperative duty to ϕ. But this default duty must itself be a cooperative duty, otherwise the rules of the scheme would be incomplete. This means there must be a cooperative rule of S that imposes on Y a default duty of normal cooperation with X. Thus, the second step of the argument posits a particular S-cooperative rule:

(ii)   Some cooperative rule R_2 imposes on Y a default duty of normal cooperation with X on the assumption that she complies with her various S-cooperative duties.

The third step generalizes the second:

(iii)   Some cooperative rule R_3 imposes on Y an array of default duties of normal cooperation with all other participants in S on the assumption that each of them complies with their S-cooperative duties.

Here R_3 entails R_2 as a subordinate rule, since Y's default duties with regard to X follow from his array of default duties with regard to all the scheme's participants. In its simplest form, R_3 just universalizes R_2—that is, it states that Y's default duties of cooperation are the same with regard to all other participants of the scheme. In more complicated cases, R_3 imposes on Y different default duties with regard to different members of S. For instance, the default duties of normal cooperation adults have with regard to children presumably differ from those they have with regard to other adults. Whatever complexity is required here can be accommodated in the complexity of R_3.

The fourth step narrows the scope of R_2, since it identifies the default duties of normal cooperation Y has with regard to X on the assumption

that she complies not with all her S-cooperative duties, but only with the various default duties of normal cooperation she has with regard to the other participants of S in virtue of a corresponding rule of type R_3.

(iv) Some cooperative rule R_4 imposes on Y a default duty of normal cooperation with X on the assumption that she complies with her own array of default duties of normal cooperation with all other participants of S.

The fifth step generalizes the fourth, just as the third step generalizes the second:

(v) Some cooperative rule R_5 imposes on Y an array of default duties of normal cooperation with all other participants in S on the assumption that each of them complies with their default duties of normal cooperation with all other participants of S.

The final step generalizes once again:

(vi) Some cooperative rule R_6 imposes on all participants of S an array of default duties of normal cooperation with all other participants of S on the assumption that each of them complies with their own array of default duties of normal cooperation with all other participants of S.

The generalization in the sixth step allows for similar complexity to that in the third. Just as the default duties adults have with regard to children may differ from those they have with regard to other adults, the default duties of children may differ from those of adults. The necessary complexity can be accommodated in R_6. Note, though, that however complex the rule may be, and whatever its particular content, it presupposes that all the participants of the cooperative scheme comply with its rules. This is because if all members comply with their default duties, there is no violation of a rule that could trigger the permissibility of sanctions or the applicability of rules governing the imposition of sanctions. In this case, the only applicable rules governing the participants are R_6 and the various subordinate rules it entails. This means that whatever a full

specification of the rules of S may ultimately contain, it contains at least one rule that applies to the participants when they all comply with the applicable rules of the scheme. The justification of R_6 must accordingly proceed on the basis of a strict compliance assumption. Refer to a rule of type R_6 as a *strict compliance rule*.

Note that a strict compliance rule is worked out on the assumption that all the participants of a scheme of cooperation comply with their duties. It does not follow that the rule ceases to have any normative purchase as soon as one participant violates a duty. A rule that ceased to have normative purchase as soon as it was violated would be unworkably fragile. Instead, strict compliance rules are more or less robust, depending on the scheme in question, insofar as they generate duties that hold across a more or less extensive range of circumstances, from the core case of strict compliance to various cases of partial compliance. A fully moral theory of the scheme in question will include auxiliary rules that specify how the participants' duties are affected the more non-compliance there is.

To illustrate the idea of a strict compliance rule, consider the terms of association in *Fellow Passengers*. The strict compliance rule is the rule that Anne, Bruno, and the other passengers are bound to follow on the assumption that all the occupants of the quiet car follow that same rule. It might have something like the following content:

The Quiet Car Rule
Sit quietly in your seat for the duration of your journey. Do not disturb other passengers on leaving your seat or otherwise physically annoy or pester them. Do not talk on the phone or to yourself, or otherwise make any noise that might disturb other passengers. If you wish to listen to music, use headphones and keep the music at a volume other passengers cannot hear. Do not try to engage other passengers in conversation. If for some pressing reason you must communicate with someone else on board the train, do so very briefly and in a low voice.

The quiet car rule is a strict compliance rule since it supposes that the passengers all comply with it. It does not say how Bruno should respond to Anne's refusal to end her phone call, at least after a whispered request

to her proves ineffective. Her continued violation of the quiet car rule leaves him in a somewhat awkward predicament. He does not become permitted to do just anything to get her to comply with the rule—presumably, he may not grab her phone and throw it out the window. But nor is he necessarily required to sit silently and endure her lack of consideration. Thus, a full specification of the cooperative rules of the quiet car would include some further rules providing guidance to the passengers in cases where some of their number fail to abide by the strict compliance rule.

2. The Priority of Strict Compliance

As the argument stands, every scheme of genuine social cooperation requires a strict compliance rule of type R_6. It does not follow, however, that this strict compliance rule is anything other than an artificial way to articulate the logical implications of a relationship of mutual account-ability in a cooperative scheme. What real work does the strict compli-ance rule do in the justification of the rules of S as a whole? There are two problems here. First, the strict compliance rule must have some sort of priority in the moral justification of the overall rules of the scheme. Without this priority, the strict compliance puzzle cannot be solved. Ideal theory could be a philosophically defensible endeavor, but there would be no reason to accept Rawls's claim that ideal principles of justice provide the necessary foundation for moral norms governing responses to injustice. Refer to this as the *priority problem*. Second, the strict compliance rule must be justified on the basis that the participants' overall array of default duties of normal cooperation constitute a fair and reciprocal scheme of cooperation in itself. If the values of fairness and reciprocity only governed the scheme's rules once the strict compli-ance assumption was suspended, then there would be no important substantive content in the role played by ideal theory. Refer to this as the *fairness problem*.

The solution to the priority problem draws on the idea that default duties of normal cooperation always include a duty to refrain from sanctioning a person for some violation. A sanction is not just any sort

of negative treatment—it is not always ordinarily impermissible to treat another participant in a cooperative scheme in a negative manner. Rather, sanctions have a particular rationale: their imposition is warranted only in response to some (threatened) violation of a duty. This means all rules governing what sanctions may be imposed, who may impose them, what must be done once they are imposed, and so forth are ultimately secondary, either directly or indirectly, to the duties participants have when the question of sanctions does not arise. But these are the duties the participants have when no one violates a duty; that is, when they all comply with the rules of the scheme. And the duties they have when all comply with the rules of the scheme are the default duties of normal cooperation stemming from the strict compliance rule. Thus, a strict compliance rule is the core element of any system of rules for a cooperative scheme.

The solution to the fairness problem turns on a particular kind of interest at stake in the justification of any fair scheme of cooperation. A cooperative scheme is fair only when it properly promotes or satisfies the various relevant interests of the participants. Not every individual interest is relevant, to be sure, but the question of fairness does not arise without some reference to the participants' interests. Some of the relevant interests are typically non-normative, in the sense that their fulfillment makes a person's life go better independently of any normative state she may be in. For instance, people have interests in being healthy and happy, but health and happiness are not in themselves normative states. But individuals also have normative interests; that is, interests in being in a certain normative state (Owens 2012). For instance, people have an interest in possessing the normative power to permit others to enter their homes. Another normative interest helps resolve the fairness problem. This is the normative interest a participant in a cooperative scheme has in being a member in good standing of the scheme. For an individual to be a member in good standing in a scheme is for all participants' default duties of normal cooperation with her to be in force, and for this fact to be generally recognized by the participants. The interest in good standing is distinct from any interest the person has in not being sanctioned by the other participants for some (perceived) violation of a duty—the value of good standing to her is not exhausted by how it

makes such negative treatment less likely. Indeed, it is very often the threat to this normative interest in itself that motivates compliance with the scheme's rules—people are frequently more perturbed by the mere prospect of being in moral trouble with their peers than of experiencing any negative treatment that may arise in consequence (May 2013, 559).

The normative interest in good standing is a natural element of the Rawlsian cooperative conception of society.[14] Rawls identifies two moral powers: the capacity for a conception of the good and the capacity for a sense of justice (1971/1999, 41). The sense of justice—the effective desire to comply with the rules of a just scheme—corresponds to the interest in good standing. This is because when citizens recognize a person is effectively motivated to act in accordance with just rules, they recognize their default duties of normal cooperation with her remain in force. The function of a sense of justice in justice as fairness is not simply to explain how citizens might freely comply with the institutions of a well-ordered society. In addition, the principles of justice are themselves justified, in part, by how well they allow for the proper development and realization of a sense of justice. Other things being equal, a conception of justice that stymied the sense of justice would be deficient. One reason for this is that the parties in the original position are motivated to ensure the social bases of self-respect, but cannot do so without providing for the interest in good standing. This is because a person's self-respect depends, in part, on the respect she receives from others, and to respect another person is to regard her "as a moral being with a sense of justice and a conception of the good" (1971/1999, 296). People are often indignant when they are unjustifiably accused of violating just terms of association primarily because such an unfair accusation is disrespectful, not because it may lead to some sanction in consequence.

More generally, the interest in good standing is always a relevant interest in the justification of the rules of a fair scheme of cooperation. A relationship of mutual respect is a corollary of a relationship of mutual accountability—the participants respect each other insofar as they are

[14] In addition, Rawls (1999b) implicitly appeals to the interest in his defense of the claim that decent hierarchical peoples should be recognized as full members in good standing of the international Society of Peoples.

250 SIMON CĂBULEA MAY

prepared to hold each other accountable for following the rules of the scheme. Thus, the fairness of a cooperative scheme requires an assessment of its fairness when each participant's interest in good standing is satisfied. But this occurs only when everyone's default duties of normal cooperation remain in force; that is, in the circumstances governed by the strict compliance rule. This means the strict compliance rule must be justified on the basis that the default duties of normal cooperation themselves constitute a fair and reciprocal cooperative system. Unless this constraint is satisfied, the scheme as a whole could only fairly balance the various other interests at stake at the cost of the interest in good standing of at least some of the participants. Fairness and mutual respect would be at loggerheads.

The fairness of a cooperative scheme's strict compliance rule has an important consequence. Violations of the rules of the scheme can be correctly criticized as unfair. For instance, in *Fellow Passengers*, Anne unfairly takes advantage of the quiet provided by the other passengers by violating the quiet car rule. But violations of the rules of the scheme must be unfair in virtue of departing from some standard of fairness. This standard of fairness cannot itself be a standard that assumes violations of the scheme's rules, since then these violations would not constitute departures from fairness. Rather, they would be preconditions of the scheme's overall fairness. Thus, criticism of Anne's behavior as unfair refers back, ultimately, to the fairness of the quiet car rule. More generally, then, the strict compliance rule possesses justificatory priority with respect to the justification of the overall scheme. The moral terms of a cooperative scheme therefore depend on the justification of its strict compliance rule. Since this rule depends on a strict compliance assumption, premise (4) follows:

(4) The moral terms of genuinely cooperative schemes depend on a strict compliance assumption.

IV. The Well-Ordered Society

I have argued that moral theories of cooperation require strict compliance assumptions because cooperative schemes involve relations of

mutual accountability between the participants. Premise (1) and premise (4) together imply that principles of justice should be justified as an exercise in ideal theory:

(1)   A just political society constitutes a grand scheme of social cooperation between free and equal persons for reciprocal advantage.

(4)   The moral terms of genuinely cooperative schemes depend on a strict compliance assumption.

Therefore:

(5)   The moral terms of a just political society—its principles of justice—depend on a strict compliance assumption.

In this section, I illustrate the cooperative interpretation of the ideal theory of justice by discussing two features of Rawls's statement of justice as fairness: two roles played by the principles of justice of a well-ordered society and the connection between ideal theory and Rawls's restriction of the principles of justice to the basic structure of society.

1. Two Roles for Principles of Justice

In section 1 of *A Theory of Justice*, Rawls discusses the role of the principles of justice as follows:

> Let us assume, to fix ideas, that society is a more or less self-sufficient association of persons who in their relations to one another recognize certain rules of conduct as binding and who for the most part act in accordance with them. Suppose further that these rules specify a system of cooperation designed to advance the good of those taking part in it. Then, although society is a cooperative venture for mutual advantage, it is typically marked by a conflict as well as by an identity of interests. . . . A set of principles is required for choosing among the various social arrangements which determine [the] division of advantages and for underwriting an agreement on the proper distributive shares. These principles are the principles of social justice: they provide

a way of assigning rights and duties in the basic institutions of society
and they define the appropriate distribution of the benefits and bur-
dens of social cooperation.   (1971/1999, 4)

Rawls's statement of role of the principles of justice includes two elem-
ents. The first and most obvious role is *distributive*. The principles of
justice set out the fundamental moral criteria for the distribution of
benefits and burdens between members of society. This is a role justice
as fairness shares with utilitarianism, since the latter also includes an
account of how benefits and burdens ought to be distributed. More
generally, every theory of justice includes principles that play this
distributive role.

The second, *regulative* role is less obvious. Principles of justice deter-
mine the rules of association that the members of society "in their
relations to one another recognize . . . as binding." The binding force of
the principles of justice, that is, does not lie simply in the fact that they
govern the distribution of benefits and burdens. In addition, the prin-
ciples are binding in the moral relations between citizens. I interpret this
to mean that in a just society, the citizens are bound together by a certain
relation of mutual accountability: citizens are accountable to each other
for complying with the institutional rules implied by the principles of
justice. This is not a role for the principles that justice as fairness shares
with utilitarianism. In the utilitarian approach, the principles of justice
may simply bind a central coordinating authority in its imposition of
various social arrangements. Whether or not the members of a utilitarian
society should recognize the principle of utility as binding in their moral
relations with each other is a quite separate matter.

The two roles can be expressed in terms of two questions. In their
distributive role, the principles answer the question, "what, as members
of society, do we owe to each other?" In their regulative role, the
principles answer the question, "what, as members of society, may we
demand of each other?" A well-ordered society constitutes a grand
scheme of cooperation, then, when it is governed by principles that
provide a unified answer to these two questions. On this conception of
justice, what we owe each other is the same as what we can demand of
each other. My claim here is that it is in virtue of the regulative role of the

principles of justice that they must be worked out, in justice as fairness, on the basis of a strict compliance assumption.

2. Ideal Theory and the Basic Structure Restriction

In section 2 of *A Theory of Justice*, Rawls imposes two limitations on the subject of justice. First, the principles of justice are to be developed as an exercise in ideal theory. Second, they apply to the basic structure of society; that is, "the way in which the major social institutions distribute fundamental rights and duties and determine the division of advantages from social cooperation" (1971/1999, 6). As an exegetical matter, the combination of these two restrictions in a single section should seem slightly curious. On the telic approach to ideal theory, there is no obvious reason why principles of justice should take the basic structure of society as their subject. If the well-ordered society is the perfectly just society political decision makers should aim at, then they should aim at it whatever the scope of its principles. Consider G. A. Cohen's (2008) critique of the basic structure restriction. On Cohen's view, the individuals in a just society would make certain types of decisions that are not required by the society's rules. For instance, they would freely choose to forgo incentives that shift society away from an egalitarian distribution. For Cohen, principles of justice identify a state of affairs in which each person receives what she is owed. The basic institutions of society are important subjects of justice, as Cohen recognizes, but only because they play an important role in bringing about the appropriate distribution. Since an egalitarian ethos can play this role too, there is no good reason to single out the basic structure as special. Both institutions and free choices are important as strategic routes to the realization of the ultimate telos.

The cooperative interpretation of ideal theory, in contrast, provides at least some rationale for Rawls's basic structure restriction. A preliminary argument to this effect can be briefly sketched as follows. The principles of justice are the terms of association of the well-ordered society—in their regulative role, they define what citizens may demand of each other as citizens. But the duties citizens have standing to demand of each other

are cooperative duties, and these duties must stem from the cooperative rules of the society. Thus, there are no requirements of justice that do not stem from the cooperative rules of a well-ordered society. The basic structure of society comprises those institutions that establish the fundamental rules of society (Rawls 1971/1999, 47). All other social institutions establish social rules only in accordance with the rules of the basic structure. Whatever citizens may demand of each other as a matter of fundamental justice, they may demand of each other only in accordance with the institutional rules of the basic structure.[15] Thus, if the principles of justice are to have a regulative role, as required by the idea of relations of mutual accountability, they should apply in the first instance to the basic structure. Principles of justice for other institutions, associations, and practices are subordinate to the principles governing the basic structure. And if the principles of justice for the basic structure are the fundamental terms of association, establishing the fundamental horizontal relation of mutual accountability, then they must be justified on the basis of a strict compliance assumption.

V. An Anti-Utopian Objection to Ideal Theory

In justice as fairness, the idea of a well-ordered society has two main features. First, it presents a substantive ideal of fair and reciprocal social cooperation between free and equal persons for mutual advantage. In particular, it provides an account of a society in which citizens have the moral standing to hold each other accountable for acting in accordance with the principles of justice. Second, the well-ordered society is a realistic utopia: it stands as a political telos that should orient decision making in the actual world. Once the end of a well-ordered society has been specified, the appropriate means to achieve it can be determined. On the cooperative interpretation of the ideal theory of justice, the strict compliance assumption is grounded in the first of these features. All moral theories of cooperation have a strict compliance rule at their core,

---

[15] For a related response to Cohen's objection to the basic structure restriction, see Williams (1998).

and, in the Rawlsian approach, a theory of justice is just a theory of social cooperation writ large. On the telic interpretation, in contrast, the strict compliance assumption is grounded in the second feature. If the well-ordered society is to properly fulfill its function as a realistic utopia, it must be assumed to be perfectly just. The substantive nature of the well-ordered society—in particular, the idea of mutual accountability it incorporates—forms no part of the rationale for ideal theory. The key difference between the cooperative and telic approaches therefore turns on which feature of the well-ordered society grounds the strict compliance assumption and which feature is, in consequence, inessential. This difference can be expressed in terms of two competing claims:

(i) Since the telic function of a realistic utopia grounds the strict compliance assumption in justice as fairness, the substantive nature of a well-ordered society is inessential to ideal theory.

(ii) Since the substantive nature of a well-ordered society grounds the strict compliance assumption in justice as fairness, the telic function of a realistic utopia is inessential to ideal theory.

In this section, I discuss an anti-utopian objection to ideal theory. Because this objection assumes the truth of the first claim, it does not threaten the cooperative interpretation. In the final section, I discuss the second claim and defend the possibility of non-utopian ideal theory.

Amartya Sen (2006) presents an objection to Rawls's ideal theoretic approach to justice. Sen claims political decisions are made as choices between some set of available options: a, b, c, ... etc. The decision makers should compare these options and select the best one. To do this, they do not need to identify some perfect state of affairs, z, that is not presently available to them. Since that option is unavailable to them, it is irrelevant to the assessment of the merits of a, b, c, etc. In Sen's analogy, it is unnecessary, and quite pointless, to identify the tallest mountain in the world before we determine the relative heights of Kanchenjunga and Mont Blanc (2006, 22). For Sen, justice is a matter of comparative assessments: decision makers should make improvements in the actual world by moving from one state of affairs to a better state of affairs. A conception of a perfectly just society is useless in that respect.

Note that Sen's objection to ideal theory does not depend on any substantive disagreement with Rawls's principles of justice. A comparative approach could rank the available options on the basis of three (lexically ordered) criteria: which option best guarantees equal basic liberties?; which option best promotes equality of fair opportunity?; and which option most improves the life prospects of the worst-off group? The disagreement between Rawls and Sen about ideal theory appears to be generic, in the sense that it does not turn on their substantive disagreements about the appropriate content of principles of justice. Sen's objection therefore assumes that the substantive nature of a well-ordered society forms no part of Rawls's rationale for ideal theory. Rather, he objects to the idea that the theory of justice requires the identification of an ideal Paradise Island.

The appearance of a purely generic debate between proponents and opponents of ideal theory is bolstered by John Simmons's (2010) defense of Rawls against Sen's objection. Simmons rejects the principles of justice as fairness, but not Rawls's distinction between ideal and non-ideal theory: "it is not Rawls's employment of the distinction itself…that should be the target of those who find unconvincing Rawls's (or Rawlsian) political philosophy" (2010, 6). Simmons offers two responses to Sen. The first response claims that ideal theory is necessary because political decision makers require some account of the perfectly just society z to orient their selection between a, b, c, etc. For instance, a may appear to be more just on its own merits than b, but it does not follow that justice requires that a be preferred to b. This is because a may be a cul-de-sac. If society were to go down that road, it could get stuck in a state of significant residual injustice. In contrast, b, for all its immediate faults, might put society on a much surer road to a fuller realization of justice. Decision makers need to keep their eyes on the prize, and to do that, they need a vision of a perfectly just society. Simmons's second response is that the strict compliance assumption is necessary so that different conceptions of justice can be compared in themselves—complications arising from non-compliance muddy the philosophical waters.

Since my present goal is not to criticize the telic interpretation of ideal theory, I do not argue here that Simmons's responses to Sen fail. But

there are at least some reasons for skepticism. With respect to the first response, a comparative approach can easily recognize short-term gains can come at the cost of long-term goals. Only a very short-sighted comparative analysis would lead into moral culs-de-sac. Moreover, even if some distant lodestar is needed to orient current political decision making, there is no straightforward reason why that lodestar must be a perfectly just society, governed by principles developed on the basis of a strict compliance assumption. It would plausibly be enough for it to be a society with institutions that should be established as permanent fixtures, taking the frailties of human virtue and the expected vicissitudes of compliance into account. The reasonably just society could act as a telos for political decision making just as well as a perfectly just society. With respect to the second response, an estimate of expected levels of compliance with the institutions required by proposed principles of justice may well be integral to an evaluation of those principles. If the conception requires institutional arrangements that are very likely to be flouted, then the conception may well be, to that extent, deficient as a moral guide to action. On this view, there is little practical point in being guided by a vision of a perfect society with impracticable institutions.[16]

However well Sen's objections may tell against ideal theory on the telic interpretation, they do not threaten the cooperative alternative. The key point here is that the cooperative interpretation provides an account of ideal theory that is grounded in the substantive content of justice as fairness. Even if there is no generic need for political decision making to be guided by some or other vision of a perfectly just realistic utopia, the strict compliance assumption can be defended as part of the idea of genuine social cooperation. Since principles of justice have a regulative role, justice is not simply a matter of determining which of the available options is best according to some distributive metric. A just society is one in which citizens have the moral standing to hold each other accountable for acting as just institutions require. But if the principles of justice for such a scheme of cooperation must be developed on the basis of a strict compliance assumption, as I have argued, then an account of the

---

[16] For an important argument against the claim that unrealistic theories of justice are deficient, see Estlund (2019).

perfectly just society follows in short order. Once the well-ordered society is identified as an ideal scheme of fair and reciprocal social cooperation, it should then act as a lodestar for political decision making. This defense of ideal theory therefore rejects the first of the two competing claims outlined above—the idea of a realistic utopia may include a strict compliance assumption, but only in virtue of a substantive appeal to the horizontal relations of mutual accountability in a genuinely cooperative well-ordered society.

VI. Non-Utopian Ideal Theory

Rawls's well-ordered society is utopian not just because it serves to orient political decision making. In addition, it is utopian because it is valuable as an end, rather than as a means to some other end. Since a just society is non-instrumentally valuable, perfectly just institutional arrangements should be permanent fixtures, with whatever adaptations to changing circumstances are required over time. On the telic interpretation of ideal theory, the only rationale for the strict compliance assumption is to identify the nature of a realistic utopia that stands as the ultimate end of political decision making. There is no place for the assumption except within this destination-setting business. But not every scheme of social interaction is valuable as an end. Some schemes are valuable only as means to some other end and may accordingly cease to exist once that end has been achieved. If these schemes are to be genuinely cooperative, however, they should still be governed by fair terms and horizontal relations of mutual accountability. On the cooperative interpretation of ideal theory, the moral terms of every genuine scheme of social cooperation must be grounded in a strict compliance assumption, however temporary or transitional the scheme may be. This supports the second of the two competing claims introduced in the previous section: since the substantive nature of a well-ordered society—in particular its horizontal relations of moral accountability—grounds the strict compliance assumption in justice as fairness, the idea of a realistic utopia is inessential to ideal theory. Ideal theory is not essentially utopian.

In *Fellow Passengers*, the quiet car rule is a perfectly ordinary example of an ideal theory. It specifies how passengers in the quiet car are to behave on the assumption that all behave as they should. In itself, this rule says nothing about how the passengers should behave if some of their number violate the rule. Thus, a complete moral theory of cooperation would include norms governing how the passengers should respond to non-compliance. The quiet car rule is presumably robust in the sense that it is still a relevant norm if one or two passengers make a noise. In this case, the other passengers have standing to demand that they comply with the rule and return the compartment to a state of reasonable silence. Here, a non-ideal moral theory of conduct in the quiet car is needed to guide the passengers in how they respond to their inconsiderate peers. At some point, however, extensive non-compliance with the quiet car rule will cause the scheme to collapse. If no one else in the quiet car appears to be paying any attention to the rule, it becomes defunct. In this case, Bruno loses the moral standing to demand that Anne end her phone call. The cooperative terms governing their temporary association are now presumably just the same as in all other compartments.

Consider a second everyday example:

Lunch Counter
Charlize is one of a large number of workers who all wish to order lunch at a window counter staffed by a single person. Each worker has an interest in placing their order as quickly and as smoothly as possible. Charlize is unsure about which moral principles govern how she might go about getting her lunch, taking the interests of the other workers into account.

The obvious response to Charlize's moral predicament is that the workers should cooperate with each other so that they may all get their lunches in a fair and expeditious manner. They ought therefore to constitute a cooperative scheme, governed by a fair rule, and Charlize ought to play her part by complying with this rule. For instance, the workers could cooperate with each other by adhering to the following simple rule:

The Lunch Queue Rule
Form a queue by standing in a single file behind the last person to arrive before you and allow the people ahead of you in the queue to be served before you.

The cooperative moral solution to the problem of *Lunch Counter* is an ideal theory, since it specifies what Charlize should do on the assumption that the other workers also comply with the rule. The lunch queue rule is robust in much the same way as the quiet car rule. It is still a relevant norm if some of the workers start pushing in line or circumventing it entirely. In this case, the other workers have standing to demand that they get to the back of the queue like everyone else and wait their turn. At some point, though, widespread non-compliance will cause the scheme to collapse, and Charlize is left with only non-cooperative moral solutions to her predicament. For instance, a non-ideal moral theory might allow her to squeeze her way through the crowd, as long as she did not elbow or shove any of the other workers out of her way.

Neither the quiet car rule nor the lunch queue rule is utopian in any sense. Both should serve to orient the decision making of Anne, Bruno, Charlize, and their fellows, but not as any kind of ultimate end. The association of the passengers on the train and the workers in the cafeteria is temporary. Much the same point holds in the political domain. Not every political scheme of cooperation should be established as a permanent fixture of society. Some have value only as temporary arrangements. Moreover, some temporary political arrangements have value only as transitional mechanisms to a better future. Even as transitional arrangements, however, these schemes raise important questions of fairness and reciprocity. In addition, there can be good reason for them to include horizontal relations of mutual accountability. If so, the moral terms for such transitional schemes should be worked out on the basis of a strict compliance assumption.

To illustrate the possibility of a non-utopian ideal political theory, consider the predicament that a society might find itself in the day a reviled dictator is forced from office. Utopian principles of political legitimacy define the nature of the (presumably democratic) political

system that should be established as an enduring set of institutions. But these principles do not immediately dictate how the transition to democracy ought to take place. Democratic legitimacy requires that citizens be allowed to select their representatives in free and fair elections, but citizens cannot just vote in an election the day after the dictator flees. So some morally justifiable transitional mechanism must be implemented, and this mechanism cannot be full democratic government. But the morally relevant considerations are not exhausted by questions of efficiency in transitioning to democracy and side constraints imposed by basic human rights. There is also an important question of fairness in how the transition should be effected. It would be unfair, *ceteris paribus*, for one political party to seize power in a coup and unilaterally impose its preferred conception of democratic arrangements, even if this conception were the best. A coup would circumvent appropriate deliberation and negotiation about various important issues: the structure of the electoral system to be used in a democratic election, the authority of the newly elected assembly to alter the constitution, the implementation of much-needed legislation in the interim, and so forth. Thus, the political parties find themselves in a predicament analogous to that faced by Charlize and her fellow workers.

Democratic Transition

The Party of Justice and Democracy (PJD) is one of several political parties that all wish to see a transition from dictatorship to some form of democratic government. Each party has an interest in effecting a transition to its preferred form of democracy as quickly and as smoothly as possible. The PJD is unsure about which moral principles govern how it might go about this transition, taking the legitimate interests of the other parties and their constituencies into account.

As in *Lunch Counter*, the obvious solution to the PJD's predicament is that the parties should cooperate with each other, on fair terms, in making decisions about the various preconditions of a successful transition. For instance, suppose they should cooperate with each other in accordance with the following rule:

The Transitional Government Rule
Allow all parties to be represented in a supreme decision-making forum.
Accept that all decisions about the forthcoming election are to be made
by a super-majority of parties in the forum, and that no party may
enforce any decision not made by the forum. Accept that an executive
committee of non-aligned individuals selected by the forum should
govern the country prior to the election. Do not threaten campaigns of
violence or intimidation if the forum does not reach certain decisions.

A rule of this sort constitutes a moral theory developed on the assump-
tion that all parties behave as they should. This means it constitutes an
ideal theory of transitional government. The transitional government
rule is robust in the sense that if some recalcitrant party threatens to
violate it, the other parties have the moral standing to demand that it fall
back in line. If extensive non-compliance with the rule causes the forum
to collapse, however, the PJD must seek out a quite different non-ideal
moral solution to its predicament. At this point, it may even become
morally permissible for the party to seize power in a bloodless coup, if
there is no feasible way to establish a cooperative scheme of political
decision making prior to a democratic election. The transitional govern-
ment rule is accordingly a political ideal theory that has some moral
purchase in transitional circumstances. But it very clearly cannot be a
utopian ideal theory, since no conception of transitional government can
serve as an ultimate end to orient decision making.

Conclusion

Rawls defines ideal theory in terms of a strict compliance assumption
and claims that it takes priority over the non-ideal theory governing
responses to injustice. The standard interpretation of Rawlsian ideal
theory is telic: it supposes that the function of the strict compliance
assumption is to define a conception of a perfect society—a realistic
utopia—that should serve as the ultimate telos of political decision
making. I have defended a different account of the strict compliance
assumption. On the cooperative interpretation of ideal theory, the

assumption is grounded in the fundamental idea of society as a fair scheme of cooperation between free and equal citizen. The well-ordered society provides a vision of society as a grand scheme of cooperation in which citizens possess the moral standing to press demands of justice against one another. The well-ordered society may also act as a realistic utopia, but its moral justification depends on a strict compliance assumption in just the same way as the justification of all other genuine cooperative schemes—whether these be permanent or temporary, ultimate or transitional, grand or petty, paradisaic or mundane.[17]

References

Cohen, G. A. 2008. *Rescuing Justice and Equality*. Cambridge, MA: Harvard University Press.

Darwall, Stephen. 2006. *The Second-Person Standpoint: Morality, Respect, and Accountability*. Cambridge, MA: Harvard University Press.

Estlund, David. 2019. *Utopophobia: On the Limits (If Any) of Political Philosophy*. Princeton, NJ: Princeton University Press.

Feinberg, Joel. 1970. "The Nature and Value of Rights." *Journal of Value Inquiry* 4 (4): 243–57.

Laden, Anthony Simon. 2013. "Ideals of Justice: Goals vs. Constraints." *Critical Review of International Social and Political Philosophy* 16 (2): 205–19.

May, Simon Căbulea. 2009. "Religious Democracy and the Liberal Principle of Legitimacy." *Philosophy and Public Affairs* 37 (2): 136–70.

May, Simon Căbulea. 2013. "What We May Demand of Each Other." *Analysis* 73 (3): 554–63.

May, Simon Căbulea. 2015. "Directed Duties." *Philosophy Compass* 10 (8): 523–32.

Neufeld, Blain. 2017. "Why Public Reasoning Involves Ideal Theorizing." In *Political Utopias: Contemporary Debates*, edited by Kevin Vallier and Michael Weber, 73–93. Oxford: Oxford University Press.

Owens, David. 2012. *Shaping the Normative Landscape*. Oxford: Oxford University Press.

Radzik, Linda. 2011. "On Minding Your Own Business: Differentiating Accountability Relations within the Moral Community." *Social Theory and Practice* 37 (4): 574–98.

Rawls, John. 1958. "Justice as Fairness." *Philosophical Review* 67 (2): 164–94.

Rawls, John. 1971/1999. *A Theory of Justice*. Revised edition. Cambridge, MA: Harvard University Press.

Rawls, John. 1999a. *Collected Papers*. Edited by Samuel Freeman. Cambridge, MA: Harvard University Press.

Rawls, John. 1999b. *The Law of Peoples*. Cambridge, MA: Harvard University Press.

Rawls, John. 2001. *Justice as Fairness: A Restatement*. Cambridge, MA: Harvard University Press.

Rawls, John. 2005. *Political Liberalism*. Expanded edition. New York: Columbia University Press.

Robeyns, Ingrid. 2008. "Ideal Theory in Theory and Practice." *Social Theory and Practice* 34 (3): 341–62.

Sen, Amartya. 2006. "What Do We Want from a Theory of Justice?" *Journal of Philosophy* 103 (5): 215–38.

Simmons, A. John. 2010. "Ideal and Nonideal Theory." *Philosophy and Public Affairs* 38 (1): 5–36.

Stemplowska, Zofia, and Adam Swift. 2014. "Rawls on Ideal and Nonideal Theory." In *A Companion to Rawls*, edited by Jon Mandle and David Reidy, 112–27. Hoboken, NJ: Wiley.

Valentini, Laura. 2012. "Ideal vs. Non-Ideal Theory: A Conceptual Map." *Philosophy Compass* 7 (9): 654–64.

Williams, Andrew. 1998. "Incentives, Inequality, and Publicity." *Philosophy and Public Affairs* 27 (3): 225–47.