Ideal Theory and Its Fairness Role

Lars J. K. Moen[1]

1 Introduction

John Rawls famously sees ideal theory as a necessary precursor to non-ideal theory. Amartya Sen, on the contrary, considers it useless for comparing feasible arrangements under non-ideal conditions.[1] Sen and other critics of ideal theory aim their arguments at its intended purposes of guiding long-term institutional reform and providing a metric for assessing the extent to which non-ideal arrangements deviate from the ideal of perfect justice. Responses from proponents of ideal theory typically also focus on these two roles.

In this paper, I show why failing to serve these two roles does not imply that ideal theory is useless in non-ideal theory. I do so by identifying a third role for ideal theory. I show how it is an essential part in a model of fairness meant to constrain choices between feasible institutional arrangements under non-ideal conditions. In ideal theory, we assume full compliance with institutions satisfying the principles of justice. Under these ideal conditions, everyone enjoys a fair share of primary goods as specified by the principles of justice.[2] When we then turn to non-ideal theory, where we work with actual conditions of partial compliance, we can use this ideal-theory model of fairness to limit institutional demands on individuals. Specifically, these demands cannot deny individuals the basic rights and liberties and opportunities to occupy advantaged positions they would enjoy under ideal conditions. Ideal

---

[1] See Amartya Sen, "What Do We Want from a Theory of Justice?" *Journal of Philosophy*, Vol. 103, No. 5, (2006), 215–238; Amartya Sen, *The Idea of Justice* (Cambridge, MA: Belknap Press of Harvard University Press, 2009).

[2] Primary goods are goods all rational persons want regardless of their plans in life. Here I focus on the social primary goods, the institutional distribution of which Rawls sees as the subject of justice. Rawls understands these goods to be basic rights and liberties, free movement and choice of occupation in a context of diverse opportunities, opportunities to seek advantaged positions in the main political and economic institutions, income and wealth, and the social bases of self-respect. See Rawls, Social unity and primary goods," in Amartya Sen and Bernard Williams, ed., *Utilitarianism and beyond* (Cambridge: Cambridge University Press, 1982), p. 162.

✉ Lars J. K. Moen
  lars.moen@univie.ac.at

[1] University of Vienna, Universitätsstraße 7, 1010 Vienna, Austria

theory is thus essential in restricting what institutions can demand in the pursuit of societal improvements.

This ideal-theory model of fairness is central in Rawlsian contractualism, which considers institutions legitimate insofar as society functions as a cooperative venture for mutual advantage. Institutional demands beyond the fairness constraint will undermine this ideal of society, as some members of society will then not benefit from the social cooperation. The fairness constraint ensures that no one is taken advantage of, and that compliance is in everyone's interest. The constraint thus preserves people's status as separate persons.

We thus see how ideal theory can, contra Sen, play a role in a comparison of feasible institutional arrangements. One arrangement, x, might appear more desirable than another, y, but y might nonetheless be the just option if x, unlike y, would restrict some individuals' share of primary goods beyond what they enjoy under ideal conditions of full compliance. We therefore cannot know which of x and y is more just without knowledge of the perfectly just society. Ideal theorizing is therefore not, as Sen suggests, merely an intellectual exercise with 'no direct relevance to the problem of choice that has to be faced'.[3] It has such relevance when it gives us a reason for choosing y rather than x.

An objection to the ideal-theory model of fairness therefore cannot be a redundancy charge, but it can seek to limit the extent to which the model ought to constrain non-ideal theory. I consider two such objections. One points out that the fair share requirement is not demanding enough, since we need some to take up the slack left by others under non-ideal conditions of partial compliance. The other objection, conversely, indicates that the fairness requirement is too demanding when compliance levels are so low that enforcing the fair-share restriction will have no beneficial consequences. Both of these objections give us good reasons for limiting ideal theory's fairness role in non-ideal theory. I show, however, how the fairness model provides grounds for evaluating non-ideal conditions even when it should not constrain non-ideal theory.

2 The Redundancy Objection

Sen distinguishes between transcendental and comparative approaches to political philosophy. He associates the former with Rawls, who takes a model of a perfectly just society formulated in ideal theory as necessary before we can consider how to actually make society more just. More precisely, Rawls's approach is "transcendental institutionalist", Sen says, since it models the basic institutions of a perfectly just society.

Sen, on the other hand, favors the comparative approach, on which we compare different feasible arrangements without concern for what the ideally just arrangement might be like. To compare two societal arrangements, Sen argues, we need no complete ranking of different arrangements going all the way up to the ideal.

---

[3] Sen, *The Idea of Justice*, p. 17.

It is important to note here that taking the transcendental approach is not to deny the importance of comparing feasible arrangements. What distinguishes the two approaches is that transcendentalists give an important role to ideals in such comparisons. The tension between transcendentalism and comparativism concerns whether or not an ideal can serve a useful function in such comparisons.

Sen rejects transcendental institutionalism for focusing on people's ideal behavior under certain institutions rather than on how they actually behave. Here Sen especially has in mind the full compliance assumption in Rawls's ideal theory: "(nearly) everyone strictly complies with … the principles of justice".[4] Knowledge about institutional arrangements designed for conditions of full compliance will not help us figure out what to do under real conditions of partial compliance, Sen argues. We must choose institutions depending on a behavioral pattern we can actually expect, he says.[5] Several theorists share Sen's view of the full compliance assumption in ideal theory. A theory based on how people ideally behave, they argue, is useless in theorizing about how to actually make people more compliant.[6]

This objection is especially directed towards what Zofia Stemplowska and Adam Swift call the "target role" of ideal theory.[7] Ideal theory serves this role by providing an ideal target to aim for in non-ideal theory. As Rawls says, "until the ideal is identified … nonideal theory lacks an objective, an aim".[8] The ideal of justice is meant to "specify the requisite structural principles and point to the overall direction of political action".[9] The criticism of the full compliance assumption implies that ideal theory cannot perform its target role because we do not know how to design incentives for making people perfectly compliant. Geoff Brennan and Philip Pettit even suggest that an ideal based on a full compliance assumption will distract us from realizing the best we actually can achieve instead of guiding our institutions towards perfect justice.[10]

The main argument in favor of ideal theory's target role is that without an ideal to aim for, we will myopically focus on small, short-term goods, and we will lose out

---

[4] See ibid., p. 61. John Rawls, *Justice as Fairness: A restatement*, ed. Erin Kelly (Cambridge, MA: Belknap Press of Harvard University Press, 2001), p. 13. See also John Rawls, *A Theory of Justice*, revised ed. (Cambridge, MA: Belknap Press of Harvard University Press, 1999), pp. 7–8.

[5] See Sen, *The Idea of Justice*, p. 69.

[6] See Geoffrey Brennan and Philip Pettit, "The Feasibility Issue," in Frank Jackson and Michael Smith, ed., *The Oxford Handbook of Contemporary Philosophy* (Oxford: Oxford University Press, 2005), 258–279; Colin Farrelly, "Justice in Ideal Theory: A Refutation," *Political Studies*, Vol. 55, No. 4, (2007), 844–864; William A. Galston, "Realism in Political Theory," *European Journal of Political Theory*, Vol. 9, No. 4, (2010), 385–411; Jacob T. Levy, "There Is No Such Thing as Ideal Theory," *Social Philosophy and Policy*, Vol. 33, No. 1–2 (2016), 312–333; David Schmidtz, "Nonideal Theory: What It Is and What It Needs to Be," *Ethics*, Vol. 121, No. 4, (2011), 772–796.

[7] See Zofia Stemplowska and Adam Swift, "Ideal and Nonideal Theory," in David Estlund, ed., *The Oxford Handbook of Political Philosophy* (New York, NY: Oxford University Press, 2012), 373–389, p. 376.

[8] John Rawls, *The Law of Peoples* (Cambridge, MA: Harvard University Press, 1999), p. 90.

[9] John Rawls, *Political Liberalism*, expanded ed. (New York, NY: Columbia University Press, 2005), p. 285.

[10] See Brennan and Pettit, op. cit., pp. 260–263.

on greater, more distant gains.[11] But as David Wiens points out, showing how ideal theory can serve this purpose requires a comprehensive analysis of the causal mechanisms at work in ensuring that prescriptions have their intended effects.[12] Ideal guidance theorists provide no such analysis, and it is hard to see how they ever will. Moving whole institutional structures in a particular direction is very complicated, perhaps even impossible. Our aim must therefore be continually revised as we learn about the ever-changing complexity of society. We then get a moving target rather than a fixed point to aim for in the long term.[13]

The other role Stemplowska and Swift attribute to ideal theory is the "urgency role".[14] Ideal theory is meant to provide a metric that enables us to compare cases of injustice and identify which is more unjust and therefore more urgently in need of a remedy. This appears to be Rawls's main reason for seeing ideal theory as a necessary precursor to non-ideal theory when he says "[t]he reason for beginning with ideal theory is that it provides, I believe, the only basis for the systematic grasp of these more pressing problems".[15] Institutional arrangements, he says, are more unjust the more they deviate from the ideal of perfect justice.[16]

A central point in Sen's critique, however, is that knowledge of the ideal is useless for making justice, and therefore urgency, comparisons. We need not know what the ideal society is like, he says, to know that hunger and starvation are problems in urgent need of a remedy. In less obvious cases, however, knowledge of the ideal seems more useful.[17] But one problem with such subtle comparisons is that there are many different ways of assessing the extent of the deviation from the ideal, as one arrangement might be closer to the ideal than another in one way but further away in a different way. To perform the urgency role, ideal theory must provide some way of assigning to institutional arrangements what Gerald Gaus calls a "justice score", which captures these different dimensions of justice.[18] But it is unclear how to weight the different dimensions and therefore unclear how to compare two arrangements when one does better on one dimension and the other better on another dimension. We need some way of determining which features of the ideal are more important and therefore worse in terms of injustice. But as Sen points out,

---

[11] See A. John Simmons, "Ideal and Nonideal Theory," *Philosophy & Public Affairs*, Vol. 38, No. 1, (2010), 5–36; Zofia Stemplowska, "What's Ideal about Ideal Theory?" *Social Theory and Practice*, Vol. 34, No. 3, (2008), 319–340, pp. 332–334.

[12] See David Wiens, "Prescribing Institutions without Ideal Theory," *Journal of Political Philosophy*, Vol. 20, No. 1, (2012), 45–70, pp. 63–64.

[13] See Gerald F. Gaus, *Tyranny of the Ideal: Justice in a Diverse Society* (Princeton, NJ: Princeton University Press, 2016), pp. 59–61; Alexander Rosenberg, "On the Very Idea of Ideal Theory in Political Philosophy, *Social Philosophy and Policy*, Vol. 33, No. 1–2, (2016), 55–75, pp. 63–65.

[14] See Stemplowska and Swift, op. cit., p. 376.

[15] Rawls, *A Theory of Justice*, p. 8.

[16] See ibid., p. 216; Rawls, *Justice as Fairness*, p. 13.

[17] See Laura Valentini, "A Paradigm Shift in Theorizing about Justice? A Critique of Sen." *Economics and Philosophy*, Vol. 27, No. 3, (2011), 297–315, pp. 306–309.

[18] See Gaus, op. cit., pp. 42–43.

such evaluations are not within the domain of transcendentalism, which is only concerned with the question of what justice is.[19]

But perhaps the lexical ordering of Rawls's principles of justice can serve this function.[20] A violation of the liberty principle is in greater need of a remedy than is a violation of the fair equality of opportunity principle, which in turn is in greater need of a remedy than is a violation of the difference principle. And a remedy for a violation of the fair equality of opportunity principle cannot compromise the first principle, and a remedy of a violation of the difference principle cannot compromise either of the other two principles. The problem with this way of assessing the distance to perfect justice is that the absolute priority to liberty, as Sen points out, is surely too strong.[21] And if we weaken this priority rule, it again becomes unclear how the principles are to be weighted against each other.

In light of these criticisms of ideal theory, it might be tempting to agree with Sen that the question "What is a just society?" is neither a good starting point nor a plausible end point for a useful theory, and that ideal theory is simply redundant.[22] But even if these criticisms are accurate, drawing this conclusion would be unwarranted since we have not yet considered ideal theory's fairness role.

3 Fairness

The fairness role is usually overlooked in the literature on ideal theory. To see how ideal theory performs this role, first recall that in ideal theory, everyone is assumed to strictly comply with just institutions. We also assume favorable circumstances, under which the economic means, education, skills, and other necessities for running a democratic regime are in place.[23] Under such ideal conditions, no one enjoys privileges beyond what the principles of justice allow, and everyone has a share of primary goods in accordance with these principles. That is, no one is denied basic rights or prevented from exercising basic liberties, there is fair equality of opportunity to occupy advantaged positions in society, and the distribution of income and wealth benefits everyone, even the worst off. Under these ideal conditions, everyone enjoys a fair share of primary goods.

This ideal model of society provides no complete account of a particular institutional arrangement and how individuals behave under it. It instead gives us criteria for a just institutional structure and a general view of what personal interests it allows individuals to satisfy.[24] This is particularly helpful for seeing how the

---

[19] See Sen, *The Idea of Justice*, pp. 98–101.

[20] See Rawls, *A Theory of Justice*, p. 216.

[21] See Sen, *The Idea of Justice*, p. 65. See also H.L.A. Hart, "Rawls on Liberty and Its Priority," *University of Chicago Law Review*, Vol. 40, No. 3, (1973), 534–555, pp. 534–555.

[22] See Sen, "What Do We Want from a Theory of Justice?," p. 226; Sen, *The Idea of Justice*, p. 105.

[23] See Rawls, *Justice as Fairness*, p. 47.

[24] Stone argues that some critics, and Gaus in particular, mistakenly understand ideal theory to provide a complete account of a perfectly just society instead of a set of criteria a just institutional arrangement satisfies. See Peter Stone, "Ideal Theory: True and False," *Ethical Theory and Moral Practice*, Vol. 24, No. 1, (2021), pp. 378–379.

principles of equal basic rights and liberties and fair equality of opportunity apply under non-ideal conditions. In the ideally just society, everyone enjoys basic rights and can exercise the basic liberties, and no one has opportunities to occupy advantaged positions at the expense of others because of morally arbitrary features of a person's situation, such as race, gender, or family. Individuals with similar skills and abilities have similar chances of accessing these positions irrespective of their initial social position.[25] How people actually use their opportunities is not part of the ideal. What matters is that morally arbitrary features have no impact on who occupies certain positions. And since the principles are fully satisfied in the ideal society, we also see that no one has to sacrifice basic rights or liberties for the sake of protecting others' basic rights or liberties, and no one's opportunities must be restricted for the sake of ensuring fair equality opportunity.

In non-ideal theory, which is characterized by partial compliance, the principles of justice are not fully satisfied. Non-compliant individuals might, for example, uphold a system where some lack opportunities because of morally arbitrary facts. This is so even in modern liberal societies, though the discrimination will there probably be private rather than state-sanctioned. Under such non-ideal conditions, ideal theory works as a constraint on how the principles can be satisfied more fully. Institutions cannot legitimately deny people basic rights or liberties or opportunities they possess in the ideal society, even if that would distribute these goods more equally or have some other desirable effect. The existing arrangement is then unfair to those who have less of these goods than under ideal conditions, but it would also be unfair to deny others primary goods they enjoy under ideal conditions in order to make the distribution more equal. That would mean denying compliant individuals primary goods to remedy an injustice they have not caused.[26] The problem is not these compliant individuals but the non-compliant individuals who do use their liberties and opportunities to deny others their fair shares of these goods.

Institutions can, and indeed should, compromise someone's ability to exercise the basic liberties or opportunities to hold advantaged positions in the society only if that individual uses these goods to deny others rights, liberties, or opportunities they have under ideal conditions. With respect to the basic rights and liberties, the constraint ensures that no one can be denied a basic right or liberty she or he possesses under ideal conditions. Under ideal conditions of full compliance, no one will deny another a basic right or prevent the exercise of a basic liberty. Rawls's first principle of justice will be fully satisfied, so that everyone enjoys "the most extensive total system of equal basic liberties compatible with a similar system of liberty for all".[27] But under non-ideal conditions, some may upset this equality by using their liberties to deprive others of theirs. It might therefore be permissible to restrict someone's

---

[25] See Rawls, *A Theory of Justice*, p. 63.

[26] While the full compliance condition in ideal theory is the basis for restricting what institutions can demand of individuals, Murphy takes it to limit individuals' moral duties. Demands of beneficence cannot fall unfairly on those who comply under conditions of partial compliance. See Liam B. Murphy, *Moral Demands in Nonideal Theory* (Oxford: Oxford University Press, 2000).

[27] See ibid, p. 220.

share of basic liberties to prevent this individual from denying others basic rights and liberties. Under non-ideal conditions, therefore, just institutions can deny one person, A, a basic liberty in order to prevent A from denying B a basic right or liberty.

To illustrate how the ideal-theory constraint works, we might consider the media's freedom of expression. This freedom can be restricted only to prevent public statements that could deny someone a basic right or liberty. False or even hateful speech would rarely justify such restrictions, as the speech itself is unlikely to compromise anyone's basic rights and liberties. It could therefore take place under ideal conditions. But consider a case where a person has the right to a fair trial after being charged with breaking the law.[28] If the press declares the defendant guilty in advance, this might make the jurors biased, thus jeopardizing the fairness of the trial. To avoid this, it might be necessary to temporarily restrict media's freedom of expression. Under ideal conditions, no one would do something that could deny the defendant a fair trial. The restriction on freedom of expression in this case is therefore permissible under the ideal-theory constraint, as no one is prevented from doing something they could do under ideal conditions.

The ideal-theory constraint is often overlooked by political philosophers, but Robert Taylor applies it with respect to fair equality of opportunity in the case of affirmative action.[29] Taylor argues that a Rawlsian approach can only support a limited defense of affirmative action, since ideal theory restricts the extent to which just institutions can treat some people differently to advance the position of others. Taylor finds, specifically, that the ideal-theory constraint is compatible only with monitoring the treatment of certain groups and with training, mentoring, and funding members of disadvantaged groups. It will not allow for quotas as a way of promoting fair equality of opportunity.[30] As Rawls himself notes, fair equality of opportunity should be "secured by provisions for education and training".[31] A quota system also appears to conflict with Rawls's view that candidates should have "the same legal rights of access to all advantaged social positions".[32]

While these are examples of particular applications of the ideal-theory constraint to basic rights and liberties and opportunities to pursue advantaged positions, the

---

[28] Rawls identifies four categories of basic liberties. I understand the right to a fair trial to fall under "rights and liberties covered by the rule of law' and freedom of expression under 'freedom of thought and liberty of conscience" and perhaps "political liberties". The final category is "the rights and liberties specified by the liberty and integrity (physical and psychological) of the person". See Rawls, *A Theory of Justice*, p. 53.

[29] See Robert S. Taylor, "Rawlsian Affirmative Action," *Ethics*, Vol. 119, No. 3 (2009), 476–506, pp. 488–491. For related uses of ideal theory, see Christine M. Korsgaard, *Creating the Kingdom of Ends* (Cambridge: Cambridge University Press, 1996), 147–151; Lars J. K. Moen, "Making Sense of Full Compliance," *Social Theory and Practice*, Vol. 48, No. 2 (2022), 285–308, pp. 290–294; Tamar Schapiro, "Compliance, Complicity, and the Nature of Nonideal Conditions," *Journal of Philosophy*, Vol. 100, No. 7 (2003), 329–355, pp. 329–355.

[30] See Taylor, op. cit., pp. 491–494.

[31] Rawls, *A Theory of Justice*, p. xv.

[32] Ibid., p. 62. Andrew Mason also makes this observation in *Levelling the Playing Field: The Ideas of Equal Opportunity and Its Place in Egalitarian Thought* (Oxford: Oxford University Press, 2006), p. 78.

constraint is perhaps also applicable to other primary goods. It is unclear, however, what it would mean to say that no one should be denied the level of income and wealth they hold under ideal conditions. At least if we accept Rawls's difference principle, there appears to be no way of determining how much income and wealth each has when the distribution makes the worst off as well off as possible. As I have noted, ideal theory only provides criteria for a just institutional structure and a general view of the personal interests it allows individuals to pursue. It does not specify individuals' behavior. How much they produce and how much wealth they generate, and so how much the worst off will have, are not parts of the definition of ideal conditions. We only know that everyone will have the means to effectively exercise the basic liberties and to enjoy fair equality of opportunity.

We can now see how ideal theory works as the basis for a constraint on non-ideal theorizing. It thus serves a function similar to that of the Kingdom of Ends in Kant's moral philosophy. In the Kingdom of Ends, everyone is treated respectfully as an end in oneself and never as a mere means to someone else's ends.[33] Everyone acts in accordance with the categorical imperative as formulated in the "formula of universal law", which demands that we never act on a maxim we would not want everyone to act on. The perfect moral behavior of the inhabitants of the Kingdom of Ends models the moral rules that apply under non-ideal circumstances, where people are not always treated as ends in themselves. Certain actions, such as lying, are therefore always wrong regardless of circumstance and consequence because they deviate from the ideal of the Kingdom or Ends. For Rawls, similarly, just institutions cannot deny individuals the basic rights and liberties or opportunities to pursue advantaged positions that they enjoy under ideal conditions, even when such a restriction would contribute to a more desirable societal arrangement.[34]

4 Contractualism

The ideal-theory model of fairness is central in Rawls's contractualism and his view of society as a "cooperative venture for mutual advantage". A just institutional arrangement promotes everyone's interests and is therefore acceptable to everyone. That is, everyone with a sense of justice—everyone desiring an institutional structure that respects everyone as free and equal members of society—will benefit from this arrangement. Under non-ideal conditions, a move to an arrangement that does not treat everyone in this way might still be an improvement on the current arrangement for everyone. But it remains an unjust arrangement since it does not fully satisfy reasonable people's desire to live in a society where everyone is treated as a free and equal member.

---

[33] Immanuel Kant, *Groundwork of the Metaphysics of Morals*, ed. Mary Gregor (Cambridge: Cambridge University Press, 1998), pp. 41–46.

[34] However, I show in Sections 5 and 6 why ideal-theory-based constraints cannot plausibly be absolute in either Rawls's or Kant's theory.

The ideal is compromised if some people are denied primary goods they would enjoy under ideal conditions. Such an arrangement fails to promote some citizens' reasonable interests and is therefore unfair to these individuals. In a cooperative venture for mutual advantage, Rawls says, "[w]e are not to gain from the cooperative labors of others without doing our fair share".[35] All contributing individuals "have a right to similar acquiescence on the part of those who have benefited from their submission".[36] No one is to take advantage of this beneficial institutional arrangement without also complying with it.[37]

Important for ensuring that the cooperative venture is mutually advantageous is that when everyone complies, no one has to make a great sacrifice; no one is left with an unfair share of the burden of establishing or maintaining the cooperative venture. The "strains of commitment" require that compliance with just institutions cannot mean abandoning one's fundamental interests.[38] Institutions violating this requirement will feel alien to the citizens, who will therefore lack motivation for compliance. By demanding that no compliant individual's fair share of basic rights and liberties and opportunities to pursue advantaged positions in society can be compromised under non-ideal conditions, legitimate institutions will deny no one the freedom to pursue reasonable personal ends. Demand more and the cooperative venture is no longer for mutual advantage. This is why the principles of justice are not "all controlling", as Rawls says.[39] Justice, he says, does not require "acts of benevolence and mercy, of heroism and self-sacrifice".[40]

Under circumstances of partial compliance, such supererogation is necessary for realizing the same public goods as under full compliance. Compliant individuals must forgo opportunities they have under ideal conditions to take up the slack left by others. But this is no requirement under the institutions of justice as fairness, as citizens have a duty to comply only insofar as they can do so "'relatively easily" without "considerable cost" to themselves.[41] Institutionalizing such requirements would not just be unfair, it would also make the institutions unstable, since we cannot expect people to have a persistent desire to fulfil such demanding institutional requirements.[42] We can thus see that while some may voluntarily choose to do more than what just institutions demand, they will not thereby make society more just.[43]

---

[35] Rawls, *A Theory of Justice*, p. 96.

[36] Ibid, p. 301.

[37] See ibid., p. 96. Here Rawls acknowledges inspiration from Hart. See H.L.A. Hart, "Are There Any Natural Rights?" *Philosophical Review*, Vol. 64, No. 2 (1955), 175–191, p. 185.

[38] See Rawls, *A Theory of Justice*, pp. 153–160.

[39] Ibid., p. 495.

[40] Ibid., p. 100.

[41] Ibid., p. 100.

[42] Ibid., p. 296.

[43] G.A. Cohen challenges this view by pointing out that the more talented, productive individuals work to the benefit of the worst-off, the greater is the satisfaction of the difference principle. See G.A. Cohen, *Rescuing Justice and Equality* Cambridge, MA: Harvard University Press, 2008). The fairness constraint based on ideal theory, however, provides a line of response to Cohen's objection. Just institutions demand only that people do their fair share and will be no more just if people choose to do more. Such charitable behavior will likely be desirable in light of other values, however.

The fairness constraint thus makes just institutional arrangements reasonably acceptable to all members of society. Rawls develops this consensus view of justice in his later work, where he takes an overlapping consensus of reasonable persons' comprehensive doctrines as basis for a shared will.[44] With an institutional structure functioning in accordance with this shared will, people enjoy political autonomy. They have an equal share in the exercise of political power, which assures their political integrity as citizens.[45] No one's private will is imposed on them; they are instead governed by a common will in which they have an equal share. This ideal of autonomy and social cooperation is the ideal to be realized in the social world.[46]

Consequentialists might agree with Sen that the ideal-theory model of perfect justice is redundant, since the good we try to maximize can always be increased, and a more desirable society will always be preferred to any less desirable society. On a Rawlsian view, however, ideal theory is important for comparing different arrangements. It might well be true that we need no ideal target to determine which of two feasible alternatives to choose under non-ideal conditions. But we need a model of a perfectly just society to say whether an alternative might be impermissible because it compromises some individuals' fair share of basic rights and liberties and opportunities for pursuing advantaged positions. A more desirable society will not be more just if it is realized by denying some individuals a fair share of these primary goods. We might compare two arrangements, x and y, and deem x more desirable than y, but nonetheless consider y more just than x insofar as x requires compromising some individual's far share of these primary goods. So, if we, like Rawls, see justice as fairness, then ideal theory is, contra Sen, necessary for comparing the justice of different societal arrangements. The fairness constraint depends on a model of perfect justice.

We saw in Section 2 how critics condemn the full compliance assumption as an easy way of avoiding difficulties of actually attracting compliance. We now see, however, that the assumption plays quite a different and more plausible role. This role is overlooked also by some commentators with a more sympathetic view of ideal theory. Alan Hamlin and Zofia Stemplowska consider it "a mistake to identify ideal theory with at least formal full compliance".[47] By assuming full compliance, we ignore important problems of regulating non-ideal behavior, they argue, and the costs involved in motivating full compliance. When we recognize the role of ideal theory in modelling fairness, however, we see that this criticism misses an important point. Assuming full compliance is crucial for making ideal theory a fundamental component in a model of fairness. We can therefore understand why Rawls refers to

---

[44] A comprehensive doctrine is a person's set of convictions about how to live, which includes a conception of the good, how we ought to treat others, and "much else that is to inform our conduct, and in the limit to our life as a whole". See Rawls, *Political Liberalism*, p. 13. And a person's "conception of the good" consists of the ends and purposes the person considers worthy of her or his pursuit over a complete life. See Rawls, *Political Liberalism*, p. 104.

[45] See ibid., p. xlii.

[46] See ibid., p. 306.

[47] Alan Hamlin and Zofia Stemplowska, "Theory, Ideal Theory and the Theory of Ideals," *Political Studies Review*, Vol. 10, No. 1, (2012), 48–62, p. 50.

ideal theory as "strict compliance theory" and to non-ideal theory as "partial compliance theory".

Rawls combines the fairness role with the controversial target role when he understands non-ideal theory to look for "morally permissible" ways of promoting the long-term goal developed in ideal theory.[48] But the fairness constraint can operate independently of such ideal guidance. We saw in Section 2 that the target role depends on a way of approaching the ideal and the urgency role on a metric for assessing the justice of different institutional arrangements. We also saw that it remains unclear how ideal theory can provide either of these things. The fairness role, however, depends on no such complex device, as we can stipulate the share of basic rights and liberties and opportunities individuals have under ideal conditions without knowing how to realize this desirable state of affairs and without a metric for comparing the justice of different arrangements. And we need no such metric and no ideal target to rule out some ways of pursuing societal improvements as unfair and impermissible because they compromise some individuals' fair share of these primary goods.

5 Picking Up the Slack

Showing how ideal theory can serve its fairness role in a Rawlsian contractualist theory is sufficient for defeating the redundancy charge against ideal theory. Critics of ideal theory may continue to aim their attacks at the target and urgency roles, but they cannot argue that ideal theory is completely useless. But we can obviously still argue about whether, or to what extent, ideal theory should play its fairness role in non-ideal theory. In this section and the next, I consider two reasons for limiting the extent to which ideal theory ought to be used in this way. The first points out that the fairness requirement is too lenient under non-ideal conditions. Under partial compliance, people have an obligation not just to respect each other's fair share of primary goods but also to pick up the slack left by non-compliers. The second objection, conversely, takes the fairness requirement to be too demanding when compliance will have no desirable effects, or even be counterproductive, due to a low overall level of compliance.

We find a version of the first concern in Elizabeth Ashford's objection to the view that individuals cannot be required to do more than fairness requires.[49] In Ashford's example, each of two boats has enough life jackets to rescue a whole group of struggling swimmers. Under ideal circumstances, the boats distribute the same number of life jackets to save all the swimmers. Fairness therefore requires each boat to rescue half of the swimmers. But suppose that under non-ideal circumstances of partial compliance, one boat leaves the scene without saving a single swimmer. On the basis of fairness, we cannot demand that the remaining boat distributes enough life

---

[48] Rawls, *Political Liberalism*, p. 89.
[49] See Elizabeth Ashford, "Utilitarianism, Integrity, and Partiality," *Journal of Philosophy*, Vol. 97, No. 8 (2000), 421–439, pp. 421–439.

jackets to save all the swimmers; it is permitted to leave half of the swimmers to drown.

Ashford actually thinks a principle will not meet the contractualist publicity condition unless it requires the remaining boat to rescue all the swimmers, at least as long as that involves little cost to the crew members.[50] This condition, which Rawls understands to have a natural place in contractualism, requires that principles be publicly justified to every member of society.[51] And Rawls shares Ashford's view when he says that everyone has a natural duty to give aid to others "provided that one can do so without excessive risk or loss to oneself".[52] And a natural duty is valid regardless of whether others reciprocate or not. Natural duties can be enforced by just institutions because of the great benefit to those in need and low cost to those who fulfil these duties.[53]

A minor and perhaps temporary restriction on an individual's share of primary goods beyond what fairness demands therefore seems justifiable insofar as it ensures considerable benefits to others. In such cases, just institutions can demand more from individuals than fairness requires. To return to the affirmative action example, Taylor also allows that under special circumstances, a quota system that temporarily denies individuals of a certain group opportunities they would have under ideal conditions might be justifiable.[54] And Rawls allows that basic liberties be temporarily denied should that be essential for ensuring that everyone can exercise them effectively.[55] We see, then, that the fairness constraint is not absolute, and that Rawls is not entirely opposed to teleological concerns. As he says, "[a]ll ethical doctrines worth our attention take consequences into account in judging rightness. One which did not would be simply irrational, crazy".[56]

But in cases where the costs and benefits of suspending fairness considerations are more balanced, ideal theory can play its fairness role in non-ideal theory. When a quota system is not expected to clearly benefit some disadvantaged group, or is expected to significantly compromise others' fair share of opportunities, it is not justifiable. And unless restricting free speech, as in the example above, is very likely to be necessary to protect someone's right to a fair trial, it will not be justified.

And while fairness may be overridden by other concerns when we determine just institutional requirements, it invariably offers us grounds for condemning those not complying with fairness requirements. The boat that did not rescue a single swimmer left the other boat with an unfairly large share of the burden of rescuing the swimmers.[57] We might consider the non-compliance of one agent, A, to be worse if the compliant agent, B, does not taken up the slack, at least if A had no reasonable

---

[50] See ibid., pp. 432–434.

[51] Rawls, *A Theory of Justice*, p. 115.

[52] Ibid., p. 98.

[53] See ibid., pp. 297–298.

[54] See Taylor, op. cit., pp. 501–502.

[55] See Rawls, *A Theory of Justice*, p. 132.

[56] Ibid., p. 26.

[57] In addition, it may have put some swimmers at a greater risk than necessary by not contributing to the efficiency of the rescue operation.

expectation of B picking up the slack.[58] The captain and crew on board boat A's decision not to rescue any swimmer might be less blameworthy if they reasonably expect boat B to save all the swimmers. But we can still see their disregard for fair play as impermissible, while granting that it is worse if the consequences are bad.

With respect to legal institutions, we might say that only the captain or crew members of the boat that left without saving a single swimmer should be legally penalized, while those on board the boat that saved only half of the swimmers should not be punished. Or if we accept the legal enforcement of natural duties, we can still think the former should be given a harsher penalty than the latter because of the violation of fair play. Even with respect to such extreme cases, then, where the right course of action is obvious to most of us, the ideal-theory model of fairness can serve a function.

6 Minimal Compliance

The second reason for limiting ideal theory's fairness role in non-ideal theory is that making people act as if they were in a society of full compliance might be ineffective under conditions of low compliance. Only at a certain minimal level of compliance will complying with the fairness requirements actually serve its intended purpose. And since this minimal compliance condition does not always hold under non-ideal circumstances, ideal theory can only play its fairness role to a limited extent in non-ideal theory.

Laura Valentini identifies this limitation in her recent assessment of fairness under non-ideal circumstances.[59] For Rawls, people have a natural duty to comply with just institutions.[60] Valentini adds that "the more unjust the status quo is due to non-compliance with the demands of justice, the less demanding the natural duty of justice becomes".[61] When it becomes foreseeably pointless to do one's fair share because of others' expected non-compliance, she argues, "one is permitted to do nothing".[62] Jeremy Waldron, similarly, notes that a natural duty to comply only comes into play when institutions pass a "test of effectiveness".[63] To have such a duty to comply with the requirements of fairness, a person must be assured that a sufficient number of others will reciprocate, as the person's compliance will otherwise be inconsequential.[64]

---

[58] See Zofia Stemplowska, "Doing More Than One's Fair Share," *Critical Review of International Social and Political Philosophy*, Vol. 19, No. 5, (2016), 591–608, p. 601.

[59] Laura Valentini, "The Natural Duty of Justice in Non-Ideal Circumstances: On the Moral Demands of Institution Building and Reform," *European Journal of Political Theory*, Vol. 20, No. 1, (2021), 45–66.

[60] See Rawls, *A Theory of Justice*, p. 99.

[61] Valentini, "The Natural Duty of Justice in Non-Ideal Circumstances," p. 46.

[62] Ibid., p. 50.

[63] Jeremy Waldron, "Special Ties and Natural Duties," *Philosophy & Public Affairs*, Vol. 22, No. 1 (1993), 3–30.

[64] See ibid., p. 21.

In response, we might again turn to Rawls's point about the irrational craziness of not taking consequences into account. If fair play is unlikely to have any desirable consequences, it is not a legitimately institutionalized requirement. It is unjust to demand that some comply if compliance is highly unlikely to have good consequences due to others' non-compliance. Indeed, if following these requirements will not have good consequences, or perhaps even bad consequences, they do not exist at all. One is not required to do *x* if doing *x* is expected to do no good, or perhaps even bad. The existence of fairness requirements therefore depends on conditions of minimal compliance.

Kant, similarly, appears to make right actions conditional on minimal compliance when he says that "since men are malicious, it is true that we often court danger by punctilious observance of the truth, and hence has arisen the concept of the necessary lie".[65] Acts such as stealing, cheating, and even killing, he says, might be necessary in an emergency, and he adds that "it rests upon everyone to judge whether he deems it an emergency or not; and since the ground here is not determined, as to where emergency arises, the moral rules are not certain". And even though there is peaceful cooperation in the Kingdom of Ends, Kant does not see war as never permissible. A commitment to never fight a war would leave nations vulnerable to attack from their enemies. Kant therefore sketches laws of war with perpetual peace as the target.[66]

For Tamar Schapiro, the minimal compliance condition does not imply a limitation to an ideal-theory constraint.[67] The two are fully compatible, Schapiro argues, as a permission to deviate from ideal behavior need not be granted on the basis of consequentialist efficiency considerations. If non-compliers will take advantage of your compliance to serve their own ends, your compliance will serve an alien will. The problem is not so much that compliance might not effectively serve a public will as that it might serve an alien will—that is, someone's private will.[68] When the law is turned into a tool for a private will, so are those who comply with it. Others' non-compliance does not free you from your obligation to comply insofar as it only reduces the efficiency of your compliance. If their non-compliance bears on the integrity of your compliance, however, it does weaken your obligation to comply, since it will conflict with your own will and the general will of your society. Schapiro even argues that compliance under such conditions could be impermissible as an act of complicity.

---

[65] Immanuel Kant, *Lectures on Ethics*, ed. Peter Heath and J.B. Schneewind (Cambridge: Cambridge University Press, 1997), p. 204.

[66] See Immanuel Kant, *Metaphysical Elements of Justice*, second ed., ed. John Ladd (Indianapolis, IN: Hackett, 1999), pp. 151–158.

[67] See Schapiro, op. cit.; Tamar Schapiro, "Kantian Rigorism and Mitigating Circumstances," *Ethics*, Vol. 117, No. 1, (2006), 32–57.

[68] See Schapiro, "Compliance, Complicity, and the Nature of Nonideal Conditions," pp. 346–347.

We can fit this view within the Rawlsian framework by first recalling that a just societal arrangement is a cooperative venture for mutual advantage that serves a shared will; it operates in accordance with principles that are publicly justifiable because they benefit all reasonable members of society. Under particularly bad conditions of partial compliance, however, no such arrangement exists. Compliance will then not serve a shared will but rather the private will of the non-compliers benefitting from the compliance. An obligation to comply, then, is conditional on compliance being expected to contribute to a cooperative venture for mutual advantage.

This means the obligation is conditional on the efficacy of compliance, and we may conjecture that it must pass something like Waldron's effectiveness test. Schapiro nonetheless considers this a non-consequentialist concern since it is based on the fitness to constitute a shared will to which the end pursued is attributed.[69] Demanding compliance is only compatible with respect for personal integrity insofar as people unite behind the ideal of a cooperative venture for mutual advantage. The problem of compliance under especially bad non-ideal conditions is not that you promote a bad or evil end, but that you serve a will that is alien to you and your society.[70]

This second objection to ideal theory's fairness role nonetheless reveals another way in which it is limited for working out just institutional requirements in non-ideal theory. The minimal compliance condition entails that one has no duty to behave as everyone does under ideal conditions when society does not function as a cooperative venture for mutual advantage. On the other hand, it seems likely that this minimal level of compliance will be met in many non-ideal cases, since the principles of justice can be adjusted to track popular interests, as we saw in Section 2. A high level of compliance is therefore a reasonable expectation. Waldron also notes that members of modern liberal societies enjoy institutions that generally attract enough compliance to pass the effectiveness test.[71] At least under normal non-ideal conditions, then, we can expect ideal theory to have a role to play in the comparison of feasible arrangements.

7 Conclusion

By aiming their arguments against ideal theory at its target and urgency roles, Sen and other critics have cast a significant shade of doubt on any purpose for ideal theory in comparisons of feasible alternatives under non-ideal conditions. By illuminating the fairness role, however, this paper has shown how ideal theory can serve a function in non-ideal theorizing that is left untouched by these critics. Ideal theory plays a key role in modelling fairness by stipulating what it means for individuals to comply with just institutions under conditions of full compliance. This fairness model works as a constraint in non-ideal theory by restricting how institutions can treat individuals under conditions of partial compliance. Specifically, they cannot deny individuals basic rights and liberties and opportunities to seek advantaged

---

[69] See ibid., p. 344.

[70] See ibid., pp. 344–347.

[71] See Waldron, op. cit., p. 21.

positioned in their society that they would enjoy under ideal conditions of full compliance. Ideal theory therefore has a role to play in comparisons of feasible institutional arrangements under non-ideal conditions to the extent that we give special weight to this fairness concern.

The degree to which ideal theory ought to play this role is obviously debatable. I have considered two limitations to its fairness role in non-ideal theory. The first points out that people should be required to do more than fairness demands under non-ideal conditions because of the importance of picking up the slack left by those who do not play fair. While this objection reveals limitations to the role fairness ought to play in non-ideal theory, it does not show that it should play no role at all. In particular, we can still use fairness as a basis for condemnation, as well as for stricter punishment, of non-compliers. The second limitation becomes apparent when we see how fairness is too demanding when the level of overall compliance is too low for compliance to have a desirable effect. But again, this shortcoming does not entail that ideal theory and its fairness role should be ignored. In fact, since ideal theory can be partly adjusted to suit the particular society to which it applies, we can quite reasonably expect adequate compliance. Ideal theory can therefore play a significant role in non-ideal theory, and there are good reasons for thinking it ought to.

Declarations

Competing interests The author declares none.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.