

THE JOURNAL OF PHILOSOPHY

VOLUME CI, NUMBER 7
JULY 2004

page

331 *Justice and Personal Pursuits*
363 *Reflection without Equilibrium*

Kok-Chor Tan
Daniel Bonevac

Published by The Journal of Philosophy, Inc.

THE JOURNAL OF PHILOSOPHY

FOUNDED BY FREDERICK J. E. WOODBRIDGE AND WENDELL T. BUSH

Purpose: To publish philosophical articles of current interest and encourage the interchange of ideas, especially the exploration of the borderline between philosophy and other disciplines.

Editors: Bernard Berofsky, Akeel Bilgrami, Arthur C. Danto, Kent Greenawalt, Patricia Kitcher, Philip Kitcher, Isaac Levi, Mary Mothersill, Philip Pettit, Carol Rovane, Achille C. Varzi. *Editor Emeritus:* Sidney Morgenbesser. *Consulting Editors:* David Albert, John Collins, James T. Higinbotham, Charles D. Parsons, Wilfried Sieg. *Managing Editor:* John Smylie.

THE JOURNAL OF PHILOSOPHY is owned and published by the Journal of Philosophy, Inc. *President,* Arthur C. Danto; *Vice President,* Akeel Bilgrami; *Secretary,* Daniel Shapiro; *Treasurer,* Barbara Gimbel; *Other Trustees:* Lee Bollinger, Leigh S. Cauman, Kent Greenawalt, Michael J. Mooney, Lynn Nesbit.

All communications to the Editors and Trustees and all manuscripts may be sent to John Smylie, Managing Editor, Mail Code 4972, 1150 Amsterdam Avenue, Columbia University, New York, New York 10027. FAX: (212) 932-3721.

You may also visit our website at: www.journalofphilosophy.org

THE JOURNAL OF PHILOSOPHY

2004

SUBSCRIPTIONS (12 issues)

Individuals	\$35.00
Libraries and Institutions	\$75.00
Students, retired/unemployed philosophers	\$20.00
Postage outside the U.S.	\$15.00

Payments only in U.S. currency on a U.S. bank. All back volumes and separate issues available back to 1904. Please inquire for price lists, shipping charges, and discounts on back orders. Please inquire for advertising rates; ad space is limited, so ad reservations are required.

Published monthly as of January 1977; typeset and printed by Capital City Press, Montpelier, VT.

All communication about subscriptions and advertisements may be sent to Pamela Ward, Business Manager, Mail Code 4972, 1150 Amsterdam Avenue, Columbia University, New York, NY 10027. (212) 866-1742

The JOURNAL allows copies of its articles to be made for personal or classroom use, if the copier abides by the JOURNAL's terms for all copying beyond that permitted by Sections 107 or 108 of the U.S. Copyright Law. This consent does not extend to any other kinds of copying. More information on our terms may be obtained by consulting our January issue or by writing to us.

POSTMASTER: Periodical postage paid at New York, NY, and other mailing offices.

POSTMASTER: Send address changes to the *Journal of Philosophy* at MC 4972, Columbia University, 1150 Amsterdam Avenue, New York, NY 10027.

© Copyright 2004 by the Journal of Philosophy, Inc.

ISSN 0022-362X

+ • +

THE JOURNAL OF PHILOSOPHY

VOLUME CI, NO. 7, JULY 2004

+ • +

JUSTICE AND PERSONAL PURSUITS*

Egalitarians are committed to two potentially conflicting sets of moral ideals. On the one hand, as egalitarians, they are committed to the ideal of justice as impartiality that regards all persons with equal concern and respect. On the other hand, as moral agents with the capacity for “a conception of the good,” they are committed to various particular and personal ends, and relationships and commitments that are basic to any rewarding and meaningful human life but that need not be egalitarian in either their motivation or outcome.¹

How can the commitment to justice as impartiality be reconciled with the pursuit of personal ends and ties, or personal pursuits for short? One way of achieving a balance here is through “a moral division of labor,” as Thomas Nagel has called it, that allows for personal pursuits but only within the rules of an institutional scheme

* This paper was originally written for a session on “G.A. Cohen and Justice,” conceived and organized by Alistair Macleod, at the Canadian Philosophical Association Meeting in Quebec City in May 2001. I am grateful to members of the audience and my fellow panelists for their contribution. I benefited also from the Queen’s Political Philosophy Reading group, at Queen’s University, with special thanks to Will Kymlicka, Alistair, and Christine Synowich. For more criticisms and comments, I thank the members of the Faculty Fellows Seminar (2001–2002), The Center for Ethics and the Professions at Harvard University; and members of the audience at George Washington University. For their interest and good questions, I thank the students in Erin Kelly’s seminar on political philosophy at Tufts University (Fall 2001). For additional discussion on this paper specifically or on the topic in general, I thank Karen Detlefsen, Erin Kelly, Tim Scanlon, Dennis Thompson, and especially Samuel Freeman. For his extensive and very encouraging written comments on several drafts, I am very much indebted to Dick Miller. Finally, I wish to express my thanks and gratitude to Jerry Cohen for generously and kindly discussing some of the arguments in this paper with me one afternoon in Cambridge, MA.

¹ This dual commitment is expressed by John Rawls in terms of the two moral powers of agents: a capacity for a sense of justice and a capacity for a conception of the good—*A Theory of Justice* (Cambridge: Harvard, 1971), p. 19. See also his *Political Liberalism* (New York: Columbia, 1993), pp. 19, 81.

that meets the requirements of egalitarian impartiality.² What this means is that the ideal of justice as impartiality is to be understood to apply specifically to institutional arrangements and need not directly limit personal choices within the rules of such arrangements. To follow Brian Barry, we can say that justice as impartiality demands “2nd-order” impartiality—that is, it requires that the rules and principles of institutions be impartial with respect to individual preferences and choices. But justice as impartiality does not entail “1st-order” impartiality—that is, it does not require “impartiality as a maxim of behavior in everyday life.”³

An exemplary statement of this division of moral labor is found in the opening pages of *A Theory of Justice*, where John Rawls writes that “the primary subject of [social] justice is the basic structure of society, or more exactly, the way in which the major social institutions distribute fundamental rights and duties and determine the division of advantages from social cooperation” (*op. cit.*, p. 7).⁴ This division of labor, that takes the ideal of justice to be primarily an institutional requirement and not as a requirement of individual conduct per se, has obvious appeal. No reasonable conception of justice can demand that individuals choose and act on impartial egalitarian principles in all aspects of their day-to-day choices and interpersonal behavior. Individual life would be very much impoverished from the moral point of view if this were so. The institutional approach to justice thus

² Nagel writes: “The ideal, then, is a set of institutions within which persons can live a collective life that meets the impartial requirements of the impersonal standpoint while at the same time having to conduct themselves only in ways that it is reasonable to require of individuals with strong personal motives”—*Equality and Partiality* (New York: Oxford, 1991), p. 18.

³ Barry, *Justice as Impartiality* (New York: Oxford, 1995), p. 194. Justice does not exhaust the domain of morality and so ethical conflicts between personal pursuits and the requirements of morality may occur even with the rules of just institutions. But the goal of the institutional approach to justice is to present a social world in which such the conflict between justice and personal pursuits may be minimized as far as is possible. For an insightful study of the problem of personal concern and morality in general, see Erin Kelly, “Personal Concern,” *The Canadian Journal of Philosophy*, xxx, 1 (2000): 115–36, here see pp. 117–18.

⁴ Rawls, of course, does not describe his own idea of justice as that of justice as impartiality, choosing instead to speak of justice as reciprocity (*Political Liberalism*, p. 50). But the main point for my purpose is that Rawls accepts the division between the demands of egalitarian institutions and personal pursuits, and that the principles of justice for institutions are to be impartial in the sense that no particular person or her goals should be favored by these principles. So while the basic structure of society is to be based on terms that all individuals can reasonably accept as required by the ideal of reciprocity, “within the framework of background justice set up by the basic structure, individuals and associations may do as they wish insofar as the rules of institutions permits”—Rawls, *Justice as Fairness*, Erin Kelly, ed. (Cambridge: Harvard, 2001), p. 50.

avoids “continuous or regular interference with individuals’ plans and actions” in the name of justice. Importantly, it allows individuals to “draw up their plans” according to their legitimate expectations as defined and protected by the system of institutional rules.⁵

On this view, the aim of any conception of justice is not to impose the demands of egalitarian impartiality on individual conduct across the board, but primarily to define the limits within which personal ends can be freely and fairly realized. Or, to put it differently, justice aims to regulate social arrangements so as to protect equally individuals’ capacity to pursue their personal ends and commitments. A theory of justice that does not amply allow for, or unduly constrains, personal pursuits will not only be self-defeating, but it will serve no purpose. It is individuals’ capacity for a conception of the good that makes considerations of justice especially poignant. The good is that which gives purpose and meaning to persons’ lives, even as justice dictates the permissible bounds of the good. This point is well illustrated in Rawls’s terse remarks that “justice draws the limit...*the good shows the point*, [and] justice cannot draw the limit too narrowly.”⁶ An account of justice that does not give sufficient space for individual conception and pursuit of the good will be without a point. It certainly will not be an account of justice made for humanity.⁷

Whether such a division of moral labor can be realized is not, of course, without challenges. Nagel has pointed out that the personal point of view that motivates or informs individual pursuits may not be easily reconciled with the impersonal point of view necessary for supporting and sustaining the impartial principles and rules of an

⁵ *Justice as Fairness*, pp. 51–52. I should note that I use the term “institutional approach” in a more general sense than Thomas Pogge does. See Pogge, “An Institutional Approach to Humanitarian Intervention,” *Public Affairs Quarterly*, vi, 1 (1992): 89–103. Pogge’s “institutional approach” focuses on the rules of institutions like mine does; but his is also institutional in an additional sense that my account need not be. For Pogge, on his institutional approach, we are asked to be concerned about injustices “only insofar as they are produced by social institutions in which we are significant participants” (p. 93). That is, Pogge’s institutional approach is *institutional* in the additional and special sense in that it is concerned only with injustices that have institutional causes. His institutional approach provides both a diagnosis and prognosis with regard to injustices.

⁶ *Political Liberalism*, p. 174, my emphasis.

⁷ I paraphrase here W.K. Frankena’s remark that “Morality is made for man, not man for morality”—quoted in Peter Railton, “Alienation, Consequentialism, and the Demands of Morality” in Samuel Scheffler, ed., *Consequentialism and Its Critics* (New York: Oxford, 1988), pp. 93–133, on pp. 98–99. Frankena is here adapting a saying of Jesus in Mark 2:27: “The Sabbath is made for man, not man for the Sabbath.” I owe the Biblical reference to an anonymous reader.

egalitarian basic structure.⁸ Nonetheless, as Nagel himself notes, this division of labor between institutions and individual day-to-day choices presents “the form of a solution” for balancing the demands of justice with the pursuit of personal ends (*op. cit.*, p. 61).⁹ On this view, call this *institutional egalitarianism*, egalitarians are required as a matter of justice to be concerned with bringing about and maintaining an egalitarian basic structure; but they are not required as a matter of justice to be egalitarians in their interpersonal decisions and actions within the rules of the basic structure.¹⁰

But the institutional approach has been criticized by some for being insufficiently egalitarian. In a series of essays and lectures, many of which now form the core chapters of his book, *If You're an Egalitarian, How Come You're So Rich?*,¹¹ G.A. Cohen argues that justice should be concerned also with individual conduct *within* the rules of institutions and not just with the rules of institutions. As he writes, “both just rules and just personal choices within the framework set by just rules are necessary for distributive justice” (EG 3). A truly just society has more than just institutions; it must also have an “ethos of justice that informs individual choices” across the board (EG 3). The institutional approach, with its limited focus on institutions, is thus insufficiently egalitarian in an objectionable way because it “represents...an evasion of the burden of respecting distributive justice in the choices of everyday life” (EG 4). Arguing specifically against Rawls, Cohen says that the institutional approach allows for excessive inequalities in society because of its insensitivity to selfish individual conduct within the rules of institutions. In particular, Rawls’s difference principle—which

⁸ Nagel writes that “[h]owever powerful the impartial, egalitarian values of the impersonal standpoint may be, they have to be realized by institutions and systems of conduct that face up to the irreducibility of the individual point of view which is always present alongside the impersonal standpoint” (p. 18).

⁹ I show how the institutional approach can ground an account of global egalitarian justice that can make space for nationalist and patriotic pursuits in *Justice without Borders: Cosmopolitanism, Nationalism, and Patriotism* (New York: Cambridge, 2004 [forthcoming]).

¹⁰ People may do more within egalitarian institutions to further promote egalitarian goals, but this will not be a duty of justice as such. This is not to say that this might not be a moral duty; indeed it could be an important duty of virtue to assist others within the rules of a just scheme. The point is that this would not be a duty that would be required by a political theory of justice. Thus, personal choices within institutional rules are not necessarily *nonnormative*. It is just that these noninstitutional matters are not the proper subject of political philosophy, but of moral philosophy or, more precisely, of ethics. As Rawls says, “many different kinds of things are said to be just and unjust: not only laws, institutions, and social systems, but also particular actions of many kinds, including decisions, judgements, and imputations”—*A Theory of Justice*, p. 7. It is just that these are not direct concerns for the topic of social justice or political philosophy.

¹¹ Cambridge: Harvard, 2000; hereafter EG.

allows for inequalities that result from the extra rewards offered to the talented as long as the worst-off members of society benefit most under this arrangement compared to alternative arrangements—is taken by Cohen to be a paradigmatic example of how the institutional approach sacrifices justice to individual preferences and choices.

In this paper, I wish to defend institutional egalitarianism against this line of criticism. First, focusing specifically on Cohen's criticism against Rawls, I will argue that the inequalities that are permitted on Rawls's institutional approach properly understood need not be as excessive as Cohen believes (section II). I will go on to suggest that Cohen's rejection of the institutional approach risks compromising his own avowed concessions to personal pursuits within reasonable limits (section III). I then turn to the more general but fundamental claim that motivates Cohen's kind of criticism, namely, that justice should be concerned with how individuals conduct themselves within the rules of institutions and not merely with the rules of institutions. I will argue that there are no reasons on grounds of egalitarian *justice* to go beyond an institutional approach; thus institutional egalitarianism does not evade the demands of justice (section IV). A special focus on institutions is not only necessary but, contra the critics of the institutional approach, also sufficient for meeting the demands of egalitarian justice. Finally, I close with some conjectural remarks about the fundamental philosophical differences underlying the institutional approach to justice and its critics like Cohen (section V).

Cohen's criticism of Rawls has elicited a growing body of literature in response (mostly in defense of Rawls).¹² My discussion differs from these responses in that it begins from what I take to be a central motivation of the institutional approach, namely the need to balance the demands of justice and personal pursuits. My evaluation of Cohen's critique will thus turn on whether his alternative conception of egalitarian justice satisfactorily addresses or avoids the problem of reconciling justice and personal pursuits. I will claim that in rejecting the institutional approach, Cohen denies the moral division of labor that is necessary for preserving space for personal pursuits against the demands of justice.

To begin, let me highlight some of the relevant features of the institutional approach (section I). These features will be largely famil-

¹² For some examples, see Joshua Cohen, "Taking People as They Are?" *Philosophy and Public Affairs*, xxx, 4 (2002): 363–86; David Estlund, "Liberalism, Equality and Fraternity in Cohen's Critique of Rawls," *The Journal of Political Philosophy*, vi (1998): 99–112; Pogge, "On the Site of Distributive Justice: Reflections on Cohen and Murphy," *Philosophy and Public Affairs*, xxix, 2 (2000): 137–69; and Andrew Williams, "Incentives, Inequality, and Publicity," *Philosophy and Public Affairs*, xxvii, 3 (1998): 225–47.

iar, but I believe that clarifying some of these aspects of the institutional approach will help us anticipate and situate the responses available to the institutional egalitarian against her critics.

I

As mentioned, a chief attraction of the institutional approach is that it allows for the reconciliation of egalitarian justice with personal pursuits, and it achieves this by subordinating the exercise of (nonegalitarian) personal choice to the rules of egalitarian institutions. But it is important to get clear on what constitutes personal choice on this account. The institutions of a society, clearly, are the result of the choices that people make. Thus, in as far as some choices people make do determine the kinds of institutions that can be established and supported, institutional egalitarians would be directly concerned about these choices of individuals. That is, such politically salient choices of individuals, as we may call them, are not strictly *personal*, and institutional egalitarians can be critical of individuals who opt against institutional arrangements that are required by egalitarian principles, or if their preferences and pursuits make the establishment and maintenance of such institutions difficult. Genuinely personal choices, then, pertain only to those actions and decisions of individuals that have no direct implications for the kinds of institutions that can be established and supported in society. It is the personal choices of individuals understood in this way, that is, choices that do not directly bear on the basic structure of society and that are permitted by the rules of the basic structure, that institutional egalitarians exempt from the demands of egalitarian justice. But individual conduct or choices that have direct implications for the kinds of institutions that can be established or supported are not personal choices, and fall directly within the purview of justice on the institutional approach. A distinction has to be observed between *individual* choices in general and *personal* choices in particular.

So while institutional egalitarians would reject the slogan “the personal is political” (if by this slogan it is meant literally that there are no personal choices as such, but that all individual choices are politically salient), they do not deny that *some* individual choices have direct institutional implications, and hence are political; and that these choices are subject to the critical evaluation of justice.¹³ So while the institutional approach accepts the division between justice and personal pursuits, it is one important feature of this approach that not all individual choices qualify as personal choices properly understood.

¹³ See Cohen’s discussion on, and his general endorsement of, this “feminist” slogan (EG 122–23).

What counts as personal pursuits on the institutional view is more restricted than some of its critics might think. The *primacy of justice* is therefore maintained on the institutional approach.¹⁴ Personal pursuits are defined by reference to the requirements of justice impartially defined, not the other way around.

The institutional approach is also sensitive to the effects of personal pursuits permitted by the rules of just institutions. Pursuits permissible under the current rules of an institutional scheme can have cumulative effects that may undermine the justness of that scheme. For example, an income taxation program that satisfies Rawls's difference principle at a given time (that is, a taxation program that supports a distributive arrangement that is to the greatest benefit of the worst-off compared with alternative programs) can, nonetheless, conceivably generate a distributive pattern over time, such that the same taxation program, unless revised or supplemented by, say, an estate tax, may no longer be to the benefit of the worst-off. More generally, this cumulated disparity in wealth distribution that a just taxation program can permit over time may result in a social scenario in which worst-off individuals have little reason to accept the basic structure of their society as a just one. Given this *fragility of justice*, as we may call it, institutional egalitarians will recognize that the rules of institutions will have to be adjusted or revised from time to time to ensure that the cumulated effects of permissible pursuits do not undermine justice. Rules of institutions, on the institutional approach, are not seen as fixed once and for all, but are subject to revisions and adjustments as conditions (such as distributive patterns over time) in society change due to the effects of (legitimate) personal pursuits. What this also means is that a given institutional rule may have to be supplemented by another rule, or other rules, in order to take account of the possible long-term (unjust) effects of personal pursuits that are permitted under that rule. As Rawls observes, "the tendency is...for background justice to be eroded even when individuals act fairly." Even when individuals act fairly, "the invisible hand guides things in the wrong direction and favors an oligopolistic configuration of accumulations that succeeds in maintaining unjustified inequalities and restrictions on fair opportunity."¹⁵ Institutional egalitarians, therefore, are aware of the need to regularly adjust the basic structure in order to take care of excessive and potentially unjustified inequalities that can arise through the just conduct of individuals (*ibid.*, p. 266).

This fragility of justice connects with another feature of the institu-

¹⁴ See Jeremy Waldron, "The Primacy of Justice," *Legal Theory*, ix, 4 (2003): 269–94.

¹⁵ *Political Liberalism*, p. 267.

tional approach, which we may call the *mutual dependency* of the principles of justice. Given the tendency “for background justice to be eroded even when individuals act fairly” as mentioned above, the success of any given institutional rule in meeting the requirements of justice depends on there being in place, and adequately enforced, other appropriate institutional rules to counteract this tendency. For instance, to prevent wealth accumulation over time and across generation that can possibly undermine egalitarian justice, an income taxation scheme may have to be supplemented by an estate tax. Or the offering of incentives for skilled labor that is permitted under one institutional rule may have to be supplemented by rules supporting equal access in education and employment to prevent a concentration of wealth among a small privileged minority.

The mutual dependency condition about the rules of institutions can be generalized to the principles of justice: a given distributive principle may seem insufficiently egalitarian when regarded on its own because of the personal choices it permits; but this need not be so if that principle is only one part of a larger framework of justice in which other principles are enforced, including, for example, the principle of equal opportunity. As an example, Rawls’s difference principle need not generate inequalities as excessive as some of his critics think, if the difference principle is understood to operate in conjunction with the principles of equal opportunity and basic liberties. I will elaborate on this important point in the next section.

This mutual dependency of justice is one important feature of the institutional approach that some of its critics, as we will see, tend to overlook. A proper evaluation of any institutional approach to justice has to consider the principles it advances as a package and not in isolation from each other. A single principle of justice examined in isolation may seem inadequate from an egalitarian perspective (say, because it appears to allow for excessive inequalities), though when evaluated in conjunction with other principles that ought to be operational in a just society, it need not be so. Principles of justice within a single conception, in short, are mutually reinforcing and dependent, and the strength of any institutional approach to justice can be properly appreciated only by considering the principles that it proposes as a whole. Principles of justice, we might say, stand or fall together.

To recap the relevant features and demands of the institutional approach: first, the primacy of justice suggests that the institutional approach is more restrictive of individual choice (and hence less permissive of inequalities) than some of its critics might think. Second, the fragility of justice will compel institutional egalitarians to pay attention to the effects of personal choices that may be detrimental to institutional justice, and to be prepared to revise, adjust and supple-

ment existing institutional rules in light of the demands of justice. In this way, institutional egalitarians can address the potentially unjust effects of certain personal pursuits, even if these are pursuits that are undertaken within the rules of institutions. Third, the mutual dependency of justice shows that any egalitarian principle is to be evaluated in conjunction with other principles of justice as part of a complete theory of justice. I will return to and elaborate on some of these features as I turn to the criticisms against the institutional approach.

II

Cohen worries that the institutional approach allows for excessive inequalities because it accommodates the selfish preferences of persons, in particular those of persons with the talents that can be deployed to benefit the worst-off members of society. Taking Rawls as his model institutional egalitarian, Cohen refers to the high ratio of corporate executive to wage-worker earnings in America (and hence its great corresponding inequality) as evidence of a problem in Rawls's institutional approach. He says that Rawls has to accept the American arrangement which allows the talented—corporate executives in this case—to demand very high wages when they could do the work they do for less, as evinced by the lower executive wage in Germany (to use Cohen's own example), because it is “a matter not of law but of ethos” that American executives are more demanding than German executives (EG 144–45). That is, the “selfish acquisitiveness” (EG 141) of executives in America, and the resulting great disparity in earnings between executives and workers, is a matter of personal behavior that falls outside the purview of justice on the institutional approach.

Cohen focuses on the application of Rawls's “difference principle” to support his claim. The difference principle, to recall, permits inequalities on the condition that the worst-off representative individual benefits most compared with feasible alternative arrangements. But this means, Cohen says, the difference principle, if it is a principle meant to apply only to the basic structure of society, must accommodate rather than challenge the American culture of selfish acquisitiveness, if such an accommodation is deemed necessary to best benefit the worst-off persons, the social inequality this selfish acquisitiveness engenders notwithstanding. Given the prevailing culture of rewards and incentives in a society like the United States, most qualified or capable Americans would not be motivated to become executives should executive earnings be reduced relative to worker salary, and presumably overall societal productivity would consequently fall, thus hurting the worst-off of society. But because the selfish demands of

talented individuals in the US is in conformity with the incentive requirements of the difference principle (that is, the worst-off are better off than they otherwise would be by offering the talented extra rewards), this selfishness is beyond criticism on the Rawlsian account of justice, and the inequalities that it generates are beyond rebuke.

It is worth stressing that Cohen does not intend his criticism as a rejection of the difference principle. For his purpose, Cohen grants that the difference principle is an acceptable principle of distribution. What he objects to is its limited application to the basic structure of society (EG 124). That is, Cohen wants to show that the difference principle understood as a principle limited to institutions, within which rules persons may do as they wish, allows for excessive inequalities because it submits to the personal choices of the selfish talented. Cohen's discussion of the difference principle is meant to support his more fundamental and general conclusion, that justice should be concerned not only with institutions but also with personal conduct within the rules of institutions. His point is that if the egalitarian spirit of the difference principle were extended also to personal conduct and choices within the rules of the basic structure, then the talented could not, and would not, demand the extra rewards that they are currently demanding and receiving, and, therefore, there would be no incentive-based inequalities of the sort that are permitted on the institutional approach.

But Cohen's argument seems wrongly to treat Rawlsian institutional justice as entirely at the mercy of *existing* individual preferences and tastes, and as if no institutional demands can be countenanced by Rawls if they are contrary to a prevailing culture of rewards and incentives. In fact, for Rawls, people have the natural duty of justice to establish just arrangements where none exists,¹⁶ and in Cohen's example, one could say that the executive/worker earning ratio shows that the present institutional arrangement in America is very far from approximating an ideal basic structure that would be required by Rawls's difference principle. As long as we can think of an alternative socio-economic arrangement (for example, a more progressive taxation scheme) that can maximize the situation of the worst-off individual in the long-run, existing preferences and ethos notwithstanding, Rawls's difference principle would require that we opt for such an arrangement.¹⁷

To be sure, if what justice in fact requires is to be constrained by a prevailing ethos, the appeal to the natural duties of justice would

¹⁶ *A Theory of Justice*, p. 115.

¹⁷ Pogge, "On the Site of Distributive Justice," pp. 138–40.

not deter Cohen's charge for, in this case, the institutions that we have a natural duty to establish are still those that would disproportionately reward talented but acquisitive individuals. However, the natural duties of justice are those duties that determine and define people's legitimate expectations and claims, not the other way around. As Rawls puts it, "It is within the limits of this division of responsibility that individuals and associations are expected to form and moderate their aims and wants.... Passionate convictions and zealous aspirations do not, as such, give anyone a claim upon social resources or *the design of social institutions*."¹⁸ To take responsibility for her end, a person has to exercise her moral capacity for controlling and revising her wants and desires in light of her legitimate entitlement. And she has the natural duty of justice to help establish and support institutions that reflect and enforce this entitlement. The "character and interests of individuals themselves" are not "fixed or given"¹⁹; and it is one of the aims of just institutions to inculcate in people a sense of what they are justly entitled to. The natural duty of justice tells us that we have the duty to establish the appropriate institutions that would shape our ethos in the appropriate way, rather than permit our extant interests and preferences to shape our understanding of our duties of justice. The primacy of justice is not annulled on the institutional approach.

Of course, justice cannot demand from individuals that which it cannot expect—"ought implies can," to recall this common Kantian point. So, if it is an essential aspect of human nature that the degree of people's selfish acquisitiveness (as reflected by the American executives that Cohen is critical of) is fixed and unalterable, then justice cannot demand otherwise and our institutional arrangements would have to be limited by this natural human infirmity. But Cohen's own reference to the smaller wage disparity between executives and workers in Germany (as a case of a more just arrangement) in fact shows that the selfish acquisitiveness of American executives is not an irremediable and inevitable fact about humanity, but that it is a cultural or historical product of American society. Indeed, this reference to the greater wage equality in Germany shows that there is at least one socio-economic arrangement alternative to the American one that is realistically possible in which the worst-off individuals can be better off. If it is indeed a principle of justice that the correct institutional arrangement from a range of feasible alternatives is the one in which the worst-off persons fare best, the primacy of justice means that

¹⁸ "A Kantian Conception of Equality" in *Collected Papers*, Samuel Freeman, ed. (Cambridge: Harvard, 1999), pp. 254–66, here p. 261, my emphasis.

¹⁹ *Political Liberalism*, p. 269.

individuals in that society ought to strive to bring about that arrangement and to limit their personal interests and expectations appropriately.

The recognition that institutions have a role in shaping individuals' preferences and attitudes is important, for it means that the current selfish ethos in American executive culture may have institutional roots and that, therefore, this ethos can be challenged on the institutional approach in as far as the *institutions* that have engendered the ethos are objectionable. For instance, one might say that the tax laws in America have engendered a certain culture of high expectation and grossly inflated sense of entitlement among the talented in America, and hence this ethos of excessive acquisitiveness is inculcated *institutionally* and is not something that is independent of institutions.²⁰ As Joshua Cohen reminds us, quoting Rawls, an "economic regime...is not only an institutional scheme for satisfying existing desires and aspirations but a way of fashioning desires and aspirations in the future."²¹ It is basic to the institutional view that living under just institutions can have the educative effect of altering people's unconsidered preferences, of firming their commitment to justice, of motivating them to do what they already believe to be the right thing, and so on. Instead of being constrained by human infirmities, one of the aims of institutions is to correct human infirmities that constrain justice.²² So, if it is the case that living under an unjust institutional scheme has distorted a people's sense of their responsibility and entitlement for the worse, the institutional approach would enjoin a transformation of this ethos through the creation of just institutions.²³

So even if it is true that prospective executives in contemporary America would not use their special skills in ways that would benefit the worst-off were their demands for very high wages not met, this initial shortfall could be corrected over time if these individuals come to acquire a sense of justice from living under just institutions, thereby becoming less acquisitive and hence willing to work as executives (and contribute to the benefit of the worst-off) for lower rewards.

Thus, the fact that executives in America can currently demand (and receive) very high wages could be due to the failure to secure the other Rawlsian principles of justice that secure the preconditions for the proper operation of the difference principle, rather than

²⁰ This argument has been made by Joshua Cohen in "Taking People as They Are?"

²¹ "Taking People as They Are?" p. 381; Rawls, *Political Liberalism*, p. 269.

²² Rawls, *A Theory of Justice*, pp. 490–91; 495–96.

²³ C.B. MacPherson's idea of the "possessive individual" who is the creature of a capitalist society is relevant here—see *The Political Theory of Possessive Individualism* (New York: Oxford, 1962).

because of a limitation in the difference principle itself. Recall that for Rawls, the difference principle applies only in the context of the principle of basic liberties and the principle of fair equality of opportunity. A prior institutional background that has met certain conditions is presupposed in the operation of the difference principle. That is, those with special talents may be rewarded more when this most benefits the worst-off *only when* all individuals are accorded equal liberties and equal opportunities.

In the American case, one can easily make the point that executives are able to demand the compensation that they in fact receive, or opt not to perform (thereby, we will assume, harming the worst-off as a result of a decrease in the overall production of material goods), only because a very small number of individuals have been privileged with special education and training for elite managerial positions while most others do not enjoy such access. That is, skilled executives command a high wage because they are a scarce resource. But this scarcity (if this is indeed the case) of executive skill is in turn created by unjust institutional conditions and is not a natural fact. It is due in large part to the great inequality in opportunities that characterizes American society in real life. Should the difference principle be operating in a context of real equal opportunity (as would be the case in an ideal Rawlsian society), it is arguable that the demand for talented executives would not outstrip the supply of executives so much as to allow the executive-wage inflation that is currently in effect. As Rawls stresses, the difference principle “works in tandem with the prior principles”; with greater open competition and fair equality of opportunity, “the more advantaged cannot unite as a group and then exploit their market power to force increases in their income.”²⁴

Properly understood and applied then, as a principle that follows the securing of basic liberties and equal opportunities, the difference principle need not permit as great an inequality between the talented and the worst-off as Cohen fears. If the difference principle fails in this regard, if it permits excessive inequalities by allowing selfish acquisitiveness to dictate the kinds of incentives that the talented can demand and in fact receive, it could be that institutional justice in other crucial respects has not been realized. People’s “abilities and talents cannot come to fruition apart from social conditions”; so if the social conditions that are necessary for the fruition of executive talents and skills fail to satisfy certain requirements of justice, then

²⁴ *Justice as Fairness*, p. 67.

the extra rewards granted to those with these skills and talents are also unjust.²⁵

In other words, objectionable incentive-based inequalities alone do not prove the inadequacy of the difference principle (or the institutional approach more generally); rather, these inequalities are more likely to be evidence of failures in society to secure other institutionally prior conditions alongside which the difference principle operates. This recalls the feature of institutional justice we earlier referred to as the mutual dependency of the principles of justice. The adequacy of the difference principle, as a principle meant to apply only to the basic structure, can be assessed only in conjunction with other principles in the conception of justice of which it is a part.

That preceding institutional injustice could be the cause of the difference principle's failure to limit inequalities adequately connects with my earlier point that the ethos of acquisitiveness can be accounted for by failures of institutions, in this case the failure to secure equal opportunity. Hence, because there is an *institutional explanation* for the selfish ethos that Cohen is critical of, the institutional approach need not be indifferent to this (institutionally derived) ethos of selfishness. On the contrary, it would be critical of the institutional conditions (for example, lack of equal opportunity, distorted expectations engendered by a prevailing culture of taxation, and so forth) that conduce the engendering and actual rewarding of such selfishness.²⁶

Let me sum up the above arguments. Cohen's criticism is that with respect to individual conduct within the rules of institutions, the institutional approach cannot say much; thus the selfish acquisitiveness of persons that can lead to excessive inequalities falls outside the critical scope of the institutional approach. I have tried to argue that the excessive incentive-based inequalities that worry Cohen do fall within the scope of institutional justice. First, the selfish acquisitiveness that can lead to such inequalities can be seen as a deformation of human tendencies due to the presence of unjust institutions, and so may be corrected by reforming these unacceptable institutions. Given the educative effects of just institutions on persons' character and sense of entitlement, institutional egalitarians can hope that talented individuals in a society with appropriately just institutions will not be moved to make the kinds of demands for extra rewards that Cohen finds objectionable. Second, even if such acquisitiveness persists in a society with just institutions, the basic structure of that society, if it is properly regulated by a complete set of principles of justice, will be

²⁵ Rawls, *Political Liberalism*, p. 270.

²⁶ See Joshua Cohen, "Taking People as They Are?" p. 377.

able to mitigate the inequalities that this acquisitiveness can affect, hence limiting the actual inequalities that selfish acquisitiveness can generate. In short, institutional justice properly understood and properly secured need not succumb to excessive inequalities on account of personal selfishness. Selfish acquisitiveness can be corrected by just institutions; moreover just institutions can minimize the unjust consequences of such acquisitiveness even if this personal failing is left uncorrected. Unless the acquisitiveness of the talented is an immutable fact of human nature that is independent of social institutions, there is no reason why an institutional approach has to take a given culture of acquisitiveness as the starting point from which to shape our conceptions of justice. The institutional approach would, on the contrary, demand institutional changes that would in turn remedy, over time, this failure of ethos that is responsible for compromising the demands of justice. In particular, understood as part of an overall framework of justice, the offering of incentives that is permitted by the difference principle need not allow for inequalities as excessive or pervasive as Cohen thinks.

III

For Cohen, egalitarian justice requires not “mere politics, but a moral revolution, a revolution in the human soul” (EG 3). A society with an ethos of justice that informs individual choices will promote “a distribution more just than what the rules of the economic game by themselves can secure” (EG 128). In an earlier series of lectures on this topic, Cohen writes that in a society with an ethos of justice, “people internalize, and—in the normal case—they unreflectively live by, principles which restrain the pursuit of self-interest and whose point is that the less fortunate gain when conduct is directed by them.”²⁷

This suggests that a truly just society is not merely a society with just institutions, but also one whose members are motivated by egalitarian considerations across the board. The institutional approach, accordingly, is insufficiently just given its limited focus on institutions. In a just egalitarian society, members adopt egalitarian attitudes not only with respect to the institutions of society but also in the “thick of daily life,” in their dealings with each other *within* the rules of egalitarian institutions (EG 3). Indeed, Cohen says that on Rawls’s own terms, a well-ordered society is one in which individuals “willingly submit themselves to the standards of justice embodied in the difference principle” (EG 128). Accordingly, no extra incentives would be neces-

²⁷ G.A. Cohen, “Incentives, Inequalities, and Community” in Grethe Peterson, ed., *The Tanner Lectures on Human Values*, Volume 13 (Salt Lake City: Utah UP, 1992), pp. 261–329, p. 316.

sary in such a society to motivate the talented to contribute to the benefit of the worst-off in society, apart from the “very special cases” in which the talented “literally” could not perform (EG 127). Individuals in an ideal Rawlsian society, on Cohen’s reading, would have taken their egalitarian commitments to heart, and so would be naturally prepared to contribute in ways to benefit the worst-off (as is required by the difference principle) without the need for extra rewards. Thus a truly well-ordered society in Rawls’s sense will be a society with “(virtually) unqualified equality” (EG 124).²⁸

Leaving aside Cohen’s interpretation of a Rawlsian well-ordered society,²⁹ Cohen’s remarks about what a just society should look like suggest that even if (I am right that) Rawls’s difference principle does not condone inequalities as excessive as Cohen fears it does, Rawls’s institutional approach is still not egalitarian enough. Given its limited focus on institutions, the institutional approach permits a distributive outcome that is less egalitarian than an approach to justice that also directly governs personal conduct within the rules of institutions. So long as individuals in a society do pursue (nonegalitarian) personal ends within the rules of just institutions, that society will be

²⁸ Cohen thus thinks that there is a tension between Rawls’s account of a well-ordered society and his institutional approach that allows individuals to pursue ends that are not motivated by egalitarian considerations.

²⁹ An obvious response to Cohen’s claim—that there is a tension between Rawls’s institutional approach and his conception of a well-ordered society (in which individuals have internalized the principles of justice)—is to point out that individuals in a well-ordered society endorse the principles of justice as principles *for the basic structure* of society. Cohen, however, anticipates this response, which he calls the “basic-structure objection” (EG 129; EG 134), and counters with arguments attempting to show that the idea of a basic structure is too unclear to be workable. Cohen relies on the family as an example of how Rawls himself is inconsistent in his own understanding of the basic structure. He argues that Rawls is unsure about how the family fits in with the basic structure, seeming to allow the family to be part of the basic structure at some moments, and at other moments not. But if the family is indeed part of the basic structure, then the basic structure cannot be restricted to just those definable coercive aspects of society but must include certain noncoercive aspects of society as well, such as conventions, customs, practices and so on (EG 137–39). Ultimately, then, Cohen’s response to the basic-structure objection is that there is a “fatal ambiguity” in the idea of a basic structure, using the case of the family as the main evidence for this claim (EG 136). But Cohen’s charge that Rawls “wobbles” on the family is unfounded. Rawls treats the family as part of the basic structure at times and not at other times because he recognizes that there are aspects of the family that are part of the coercive institutions of society and so belong to the basic structure, such as marriage, the welfare of children, the equal rights of women, and so on, and other aspects of it that are not, for example the division of labor in the family. So Cohen may not use Rawls’s discussion of the family as evidence that the idea of the basic structure is “fatally ambiguous.” The basic structure is indeed the coercive institutions of society, of which some aspects of the family are a part. If this response can be made, then Cohen’s response to the basic-structure objection is disarmed. The basic structure is not as ill-defined as Cohen thinks.

less egalitarian than one whose members always organize their personal conduct so that the “less fortunate gain.”

Yet, this strong egalitarian view, if it is indeed Cohen’s position, seems rather intrusive, for it would leave very little space for meaningful personal nonegalitarian pursuits. In light of this point, it is interesting to note Cohen’s own claim that he does not want to deny a place for what Samuel Scheffler has called “agent-centered prerogatives,” namely, the prerogatives individuals have to engage in certain (non-egalitarian) personal pursuits to a reasonable extent (EG 213 n36). What he is specifically critical of, Cohen stresses, is the *selfish acquisitiveness* that drives the talented to demand more than is necessary for them to perform.³⁰ His rejection of the institutional approach is due to his belief that it permits inequalities due to personal selfishness.

But on this point, David Estlund has argued that if Cohen is indeed receptive of “agent-centered prerogatives,” he must permit more inequalities than his strong egalitarian position can allow. For instance, agents will be permitted to favor the interests of family or friends instead of acting as egalitarian principles might dictate in the name of exercising their personal prerogatives; yet such personal choices can generate certain social inequalities. At any rate, they certainly need not be to the best interest of the worst-off in society. Agents can also, on account of personal prerogatives, be permitted to give weight to other moral considerations that need not have egalitarian implications. To use Estlund’s example, one may opt to spend extra time repairing a neighbor’s garden that one has inadvertently harmed instead of using that time to better the situation of the worst-off.³¹ Accordingly, Estlund argues, the advertised difference between Cohen and the institutional egalitarians disappears, or is at least considerably diminished, if Cohen is really prepared to accommodate personal prerogatives within some limits.³² There is, in other words, a tension between Cohen’s professed receptiveness to personal prerogatives and the strong egalitarian thesis that he is advancing.

Consequently, and more directly to the point of my thesis, if Cohen

³⁰ Cohen, “Incentives, Inequalities and Community,” p. 303. On the notion of agent-centered prerogatives, see Scheffler, *The Rejection of Consequentialism* (New York: Oxford, 1982), chapter 1.

³¹ Estlund, pp. 101–03.

³² Cohen, in a footnote (EG 212–13), says that he is willing, tentatively, to concede partially to this point, though he thinks that Estlund’s arguments lead to a middle approach between his own position and Rawls’s rather than a defense of the latter against the former. Because Cohen’s remarks are, as he himself acknowledges, tentative and brief, it will not be possible to fully assess them here. The crucial point to note here is that in accepting agent-centered prerogatives, Cohen is forced to accept more inequalities than his account of a society with an ethos of justice can permit.

rejects the basic distinction between the personal and justice that underlies the institutional approach, it is unclear how he can still preserve agent-centered prerogative in a consistent and principled way. If individuals are to be bound by egalitarian principles across all of life, it seems to me that not only would selfish-acquisitive conduct be ruled out, but the more general agent-centered prerogatives (to pay special attention to one's projects, attachments, and relationships) would also be ruled out. Absent a method of demarcating selfish behavior from agent-centered prerogatives, Cohen's rejection of the distinction between the personal and the political will force him into a morally "rigoristic" position which allows little meaningful space for personal pursuits, that he himself wants to avoid. The institutional approach, by contrast, provides a principled way of demarcating selfish (that is, impermissible) conduct from merely nonegalitarian personal (that is, permissible) conduct—the rules of institutions provide such a reference point. From the point of view of institutional *justice*, personal conduct within the rules of institutions cannot by definition be deemed selfish in an objectionable sense. Individual choices that overstep the bounds of institutional justice are impermissible, while choices within the rules of just institutions are permissible. The institutional approach provides a workable division of labor that can help make space within the demands of justice for permissible personal pursuits. Thus, rejecting the institutional approach because one wants to condemn selfish acquisitive behavior inevitably "proves too much" for it will also rule out reasonable personal pursuits. In short, not only is the alleged difference between Cohen and institutional egalitarians minimized if Cohen is indeed receptive to personal prerogatives; he must also adopt something like the institutional approach if he really wants to preserve space for reasonable personal prerogatives.

Cohen can stress here, in reply, that in an ideal egalitarian society, individuals have "internalized" their egalitarian commitments to the point that their personal pursuits are always motivated by the point of view of the worst-off. Thus there is a happy convergence between personal conduct and the requirements of Cohen's strong egalitarian justice. With the revolution in the human soul Cohen speaks about, personal prerogatives will no longer include preferences that are not informed by egalitarian concerns. Individuals in an ideal egalitarian society will, therefore, not find the egalitarian demands that apply to their personal conduct to be too demanding. Being egalitarians to the core, these individuals will simply be motivated to act, even in their personal conduct, in ways that egalitarian principles would dictate. Justice and personal pursuits are brought together through this revolution of the soul.

Yet, until this convergence is achieved, the tension between the

strong egalitarian view and personal pursuits persists. There is, therefore, the question whether this revolution in the soul is actually attainable. Reconciling the demands of justice and personal pursuits within a person's soul is not a real option if this revolution is a mere fantasy. But more interestingly for our purpose, even if this individual transformation is attainable, there is the question as to whether it is a requirement of egalitarian justice. Is a society less just simply because it does not attempt to inculcate an egalitarian ethos among its individual members across all aspects of their lives? I will pursue the latter question below (section IV).

But first, to clarify my central claims above, it might be helpful to examine the place of incentives in Rawls's view. For Rawls, incentives are not meant to satisfy people's selfish preferences, as Cohen's criticism seems to imply. Such offerings of extra rewards (just because the talented could demand and receive them) will be ruled out within an overall conception of justice that defines the kinds of personal preferences that are admissible. Rather the primary use of incentives is to capture the ideal that "[w]hat kind of work people do, and how hard they do it, is up to them to decide in light of the various incentives society offers them" and not solely in terms of how much material contribution to society they can make.³³ Individuals have varying conceptions of the good and different personal interests, including preferences about lifestyles, and the priority of liberty means that individuals may not be forced into occupational roles just because they will make the greatest contribution to society in these roles. An appropriate reliance on incentives thus reflects what I take to be one of the principal motivating goals of the institutional approach, namely, to provide a means of reconciling the demands of justice and personal pursuits. The proper use of incentives in the market place can help motivate individuals to make trade-offs in their personal choices, say between available choices of occupation, in the direction of promoting social justice.

As mentioned, Cohen suggests that in a just society, the talented should not be permitted to demand rewards beyond those that without which they could not "literally" produce (and that they would not need these extra rewards were they true Rawlsian egalitarians). It is not immediately obvious what "literally" means here. But presumably, the extra rewards that are literally needed can include the following: extra rewards to make up for the costs of training without which the person could not literally do the skilled job; extra rewards to make up for a particularly demanding or taxing job to provide persons

³³ Rawls, *Justice as Fairness*, p. 64.

holding that job with extra sustenance (broadly construed) without which most people will not be able literally to continue productively in that job; and, stretching the sense of "literal" perhaps, extra rewards to make up for a particularly dull or arduous job without which persons can reasonably lack the motivation needed to literally do the job. So rewards that are literally needed can plausibly be interpreted to include those needed to sustain a person in demanding or boring jobs, or to make up for the costs of a person's extra years of training.

But what if it is the case that without such some additional incentive (beyond that which is literally required in the ways described above) a talented person who can take on a socially useful job as an economic planner (that very few people are capable of filling) chooses to remain a struggling poet? It is not clear if we would want to say that this talented person is *literally* unable to become an economic planner without the additional incentive. After all, it is just that she prefers not to, preferring instead to experiment with poetry. There is nothing literal about her not being able to become a planner in the absence of offering her extra incentives (beyond that required to make up for her extra years of training, and the high demand and stress that come with the job). It is just that being a practicing poet is her dominant goal. Accordingly, it seems that on Cohen's account of egalitarian justice, this person could be faulted for opting to remain a poet unless she is offered sufficient material incentive for her to feel that it is worth her while to give up her poetic fantasies and take on the role of an economic planner at which she is particularly gifted.³⁴ Yet this will have the implication of drawing the limit of justice on persons' conceptions of the good life too narrowly. Such an understanding of justice will not be able appropriately to make sense of persons' freedom of occupation and pursuits, and seems to fail to appreciate the use of incentives as a legitimate basis of balancing the promotion of social justice with individual conceptions of the good life. To permit incentives only up to the point in which a person cannot *literally* perform the valued role implausibly denies that which Rawls recognizes to be a basic personal prerogative, namely that the "kind of work people do, and how hard they do it, is up to them to decide in light of the various incentives society offers them" and the diverse personal pursuits they may have.

Now Cohen might say that he allows for *genuine* personal lifestyle preferences of the sort described above to dictate the kinds of incentives that can be offered to people; what he objects to is the selfish

³⁴ Here, I factor out the social costs of her training. We can assume that whatever social obligation she incurred in her training to be a skilled economist has been discharged.

acquisitiveness of the talented that moves them to demand more incentives than is presumably necessary were they strictly to limit their considerations to genuine lifestyle preferences. It is the selfish acquisitiveness that (the incentive provision of) the difference principle caters to that exercises Cohen, not personal occupational choices as such. To clarify, one might make the distinction between extra rewards as such, and “unnecessary” rewards, taking the latter to refer to the additional rewards (i) *beyond* those without which the talented literally could not contribute in the way described above, and (ii) *beyond* those needed to compensate for a person’s genuine lifestyle preference. Cohen is concerned about the inequalities that unnecessary rewards generate, and his criticism is directed at the selfish attitudes that insist on these rewards.

But this response presupposes that, in practice, lifestyle choices can be easily divorced from, or can be understood independently of, acquisitive attitudes. In reality, however, it is not clear how we can determine when a demand for extra incentives is due to a genuine lifestyle preference and when it is due to material acquisitiveness. In most cases, lifestyle preferences are influenced by the incentives attached to the different lifestyle options available to a person, and the draw or perceived value of the incentives attached to each option is determined by the strength and depth of a person’s commitments to each of these respective options. The amount of compensation that an aspiring poet must receive before she is prepared to give up on her artistic goals to become an economist will be influenced by the perceived worth of the compensation offered to her to become an economist and this will in turn be influenced by her acquisitive attitude, as well as the worth to her of each of these pursuits in their own sake. Indeed, we can imagine cases where an overly high incentive demanded is not due to selfish acquisitiveness but is in fact due to the contrary—a great indifference to material reward. Imagine someone for whom money means so little, while her aspirations to pursue a materialistically modest artistic life are so much more important to her that it will take a lot of monetary enticement before she is prepared to do a lifestyle switch. Here the high incentive that is necessary to motivate her to switch career paths is not because of selfishness but of a deep attachment to a (nonmaterialistic) preferred way of life. Contrast this with the case in which a person for whom material wealth means a lot and so would not take on the socially valued role (even though she is only weakly committed to the alternatives available to her) unless she is very well compensated beyond that which is necessary to get her to switch, because she knows she can in fact demand and get the high material compensation. In both cases, high rewards are asked for, but only in one case is it due to

selfish acquisitiveness of the sort that Cohen is concerned with. Yet there is no practical means of identifying the point at which it is selfishness as such, rather than genuine lifestyle preferences, that is determining the personal choice not to take on a socially valuable role unless amply rewarded. Rejecting the institutional approach, and in particular the incentive provision in Rawls's difference principle, will undermine people's freedom of occupational choices. There is no practically feasible way of narrowly targeting selfish motivations without also targeting genuine lifestyle preferences of individuals.

Cohen is of course right to be concerned about the inequalities that may result from the selfish acquisitiveness of the talented. But contra Cohen, I have earlier tried to show that the institutional approach can in fact minimize such inequalities. Inequalities due to selfish acquisitiveness, and the selfish acquisitive attitudes themselves, need not fall below the radar screen of institutional egalitarians (section II). Thus it is not necessary, for the sake of combating inequalities due to selfishness, to reject the institutional approach; on the other hand, it is necessary for the sake of preserving meaningful space for personal pursuits, including the freedom of occupational choices, to maintain the institutional approach.

IV

It is natural that a society whose members have undergone a moral revolution of the sort Cohen describes, a society in which individuals have become egalitarians through and through, would have a more egalitarian distributive outcome than a society whose members pursue nonegalitarian ends within the rules of egalitarian institutions. But is this revolution required as *a matter of justice*? While a society with an egalitarian ethos in the way Cohen describes is likely to have a more egalitarian distributive outcome than a society with only egalitarian institutions, it is quite a different matter as to whether the latter is flawed from the point of view of justice. So, the question remains as to whether the more egalitarian distribution that this society would achieve is required as a matter of egalitarian justice, and, consequently, whether the institutional approach, because of its limited focus on institutions, represents "an evasion of the burdens of...justice" (EG 4) as Cohen charges. That is, does egalitarian *justice* require that individual conduct within the rules of institutions be informed by egalitarian principles? Do the burdens of justice include the inculcation of an egalitarian ethos among individuals even in their day-to-day choices and "in the thick of daily life"?

So leaving behind Cohen's specific arguments, I want to turn to the general idea motivating his line of criticism—that egalitarian justice must adopt a more encompassing view than a narrow focus on institutions. Should this be the case? To get a grip on this question,

let us first imagine a society that has satisfied institutional egalitarian standards, and then see if egalitarians should demand more of it as a matter of justice. Using Rawls's institutional justice as our model, let us suppose that the ideal of the difference principle is approximated in a society with a taxation scheme *T* that is enforced along with the principle of basic liberty and the principle of equal opportunity.³⁵ So we will say of this society that it has met the requirements of *institutional* justice, even though some inequalities will remain as the result of the personal pursuits that are permitted by the rules of this scheme.

The issue, then, is whether egalitarians should require more of this society as a matter of justice. That is, should individuals also be egalitarians with respect to their disposable income within the rules of scheme *T*? Again, we can easily grant that this society would be more egalitarian in its distributive outcome should its members be motivated to dispose of their post-tax income along egalitarian principles, than if they were to pursue nonegalitarian personal ends with this income. To be sure, the inequalities in this society that are permitted under scheme *T* need not be as excessive as some critics of the institutional approach might fear (as I tried to argue in section II); nonetheless, there would be some residual inequalities that would likely not be present should members of this society also be thoroughgoing egalitarians in Cohen's sense. The pertinent question, however, is whether the society is less just for *this* reason.

To properly address this question, I think we need to turn to a more basic question, namely, what is the purpose of equality? Or, to put the question in a different way, what are the grounds of egalitarian justice? Broadly speaking, there are two kinds of reasons for caring about inequality. Following Charles Beitz, we can call these (a) "derivative" and (b) "direct" reasons.³⁶ Derivative reasons for equality are arguments for equality that appeal to some other value/s that (more) equality would help bring about; direct reasons for equality appeal to the importance or significance of equality *as such* for individuals. I want to suggest that neither of these two kinds of reasons for caring about equality forces one to abandon institutional egalitarianism for a more encompassing ideal of egalitarian justice. Note that my aim here is not to settle the very interesting and important question regarding the point of equality—to wit, whether we are concerned about equality ultimately on derivative or direct grounds. My claim here is

³⁵ See Rawls, *Justice as Fairness*, pp. 160–61, for some discussion on the difference principle and taxation.

³⁶ "Does Global Inequality Matter?" *Metaphilosophy*, xxxii, 1 (2001): 95–112, here see pp. 97–98; see also T.M. Scanlon, "The Diversity of Objections to Inequality" in *The Lindley Lectures* (Kansas City: Kansas UP, 1997).

a ground-neutral one, that the common arguments proposed in defense of egalitarian justice do not provide reasons for going beyond an institutional focus.

(a) Consider first, the *derivative argument* for equality. A popular derivative argument in the current literature on equality is the argument from reciprocity. As Rawls has put it, the gap between rich and poor in a democratic society “cannot be wider than the criterion of reciprocity allows.”³⁷ Under the ideal of reciprocity, citizens may impose only those social arrangements on each other that each can reasonably accept, and any arrangement that permits too great an inequality will fail this reasonable-acceptance test. Accordingly, on this derivative argument for equality, egalitarians want an egalitarian basic structure because this is one necessary condition under which a shared social arrangement meets the criterion of reciprocity, thereby gaining legitimacy in the eyes of its members.³⁸ Rawls takes his difference principle (which offers the talented extra rewards for their contribution on the condition that the worst-off benefits) to express the ideal of reciprocity, in that the inequality resulting from the talented being rewarded extra for their contribution must be seen as reasonable from the point of view of the worst-off.³⁹

But does this derivative argument for equality require citizens to do more than establish an egalitarian basic structure? That is, does the criterion of reciprocity call on egalitarians not only to impose social arrangements on each other that all can reasonably accept, but also to act as egalitarians in their personal interactions with each other?

It is not clear that it does. In a society with unjust institutions, the idea of reciprocity is clearly not realized because those unfairly disadvantaged could reasonably object to the institutional arrangements that the well-off are helping to impose on them. Thus personal pursuits that feed off or sustain this injustice should be of concern to egalitarians. But in an ideal society where the terms of justice are fully and appropriately met by the basic structure, there is no reason why personal nonegalitarian pursuits need be objectionable to egalitarians motivated by the ideal of reciprocity. A society that has fully met the requirements of *T*, to use our example, may have lingering inequalities due to personal choices within the rules of *T*, but these

³⁷ *The Law of Peoples* (Cambridge: Harvard, 1999), p. 114.

³⁸ See also Richard Miller, “Cosmopolitan Respect and Patriotic Concern,” *Philosophy and Public Affairs*, xxvii, 3 (1998): 202–24. I take the idea of reciprocity to underlie the idea of “democratic equality.” See Elizabeth Anderson, “What Is the Point of Equality?” *Ethics*, cix, 2 (1999): 287–337.

³⁹ Rawls, *Justice as Fairness*, pp. 64, 76–77.

lingering inequalities need not violate the principle of reciprocity. The ideal of reciprocity will require that our *shared social scheme*—the background rules within which we pursue personal ends—be as egalitarianly informed and structured as possible. Such a scheme, and only such a scheme, would be accepted as reasonable by participants of that scheme. But the ideal of reciprocity would allow for personal pursuits within the rules of the egalitarian scheme even if these result in some inequalities.

This last point does not mean that the ideal of reciprocity does not apply to interpersonal choices and actions but only to institutional arrangements. It is compatible with my argument that the ideal of reciprocity is an ideal that is basic to morality as such, and so applies in interpersonal moral situations as well.⁴⁰ My point is that reciprocity enjoins different kinds of commitments in different contexts. Reciprocity with respect to the kinds of institutions we may impose on each other would require that we advance and support only those institutional arrangements that give equal consideration to the interests of all. However, reciprocity at the personal level need not necessarily require an impartial egalitarianism of this sort but can in fact allow for deviations from such egalitarian considerations. As T.M. Scanlon has noted, an interpersonal morality that does not provide space for personal concern, for example the partial concern between friends, would be one that reasonable people may reasonably reject.⁴¹ Thus, all I am claiming is that the ideal of reciprocity generates an obligation to secure a certain kind of institutional arrangement, not that this ideal applies only to the question of institutions. Reciprocity as an ideal would apply to personal conduct as well, but the duties that reciprocity would generate in the sphere of personal interaction are different and need not be duties that require impartial egalitarianism as the norm of conduct. Thus, while it does not violate the ideal of reciprocity to favor one's kin or friends more than strangers in one's genuinely personal dealings within the context of just institutions, it would violate the ideal if institutions are arranged so as to favor specific individuals over others.

The derivative idea of equality, in particular one which takes the importance of equality to stem from the ideal of reciprocity, would require that our institutions, in order to be just, be as impartial as is possible with respect to individuals' conception of the good, their personal ties, preferences, and so on. But reciprocity does not require impartial conduct within the rules of just institutions. So, if the ideal

⁴⁰ For one view, see Scanlon, *What We Owe to Each Other* (Cambridge: Harvard, 1999).

⁴¹ *What We Owe to Each Other*, p. 160. See also Kelly, "Personal Concern."

of reciprocity is what grounds egalitarian justice, the institutional approach does not represent an evasion of the demands of justice.

(b) Consider, next, *direct reasons* for caring about equality. A common direct argument for equality is that people's life prospects should not be unfairly limited by circumstances beyond their control; hence the motivating goal of egalitarian justice is to mitigate the effects of such contingencies. As Cohen himself has put it elsewhere, "a large part of the fundamental egalitarian aim is to extinguish the influence of brute luck on distribution."⁴² On this direct argument for equality, we care about equality not because equality facilitates, or is a necessary condition for, a higher moral objective (for example, that of reciprocity), but because inequality *itself* is bad because of its impact on people's choices. We can call this view "luck egalitarianism" following the standard usage.⁴³

But if we are concerned about equality because of a direct concern with mitigating the effects of contingencies on people's life chances, there is no immediate reason why such a concern must take us beyond the basic structure. The belief of institutional (luck) egalitarians is that an appropriately ordered basic structure will come close to annulling the effects of chance and brute luck on people's lives *without* intruding on people's liberties to pursue their ends. If our taxation scheme *T* appropriately eliminates "disadvantage[s] for which the sufferer cannot be held accountable,"⁴⁴ our society has lived up to its commitment to equality as a direct goal. Such a society, even if lingering inequalities persist due to choices people make (but not luck), need not be troublesome from a luck egalitarian viewpoint.

The reason why institutions are the direct target on this account of equality is that, to borrow Rawls's words, the "effects [of institutions] are so profound and present from the start." The social positions of people, and their expectations, are determined to a large degree by the institutions of society.⁴⁵ Natural facts in themselves need not have any implications for justice; what matters is what institutions make of these facts. As Rawls writes: "A further essential distinction is between the unequal distribution of natural assets, which is simply a natural fact and neither just nor unjust, and the way the basic structure of society makes use of these natural differences and permits them to

⁴² "On the Currency of Egalitarian Justice," *Ethics*, xcix, 4 (1989): 906–44, p. 931.

⁴³ Anderson rejects "luck egalitarianism," in effect rejecting direct arguments for equality. But, as mentioned, my aim here is not to defend luck egalitarianism but only to show that this idea of egalitarianism does not demand an ethos-based approach to equality.

⁴⁴ Cohen, "On the Currency of Egalitarian Justice," p. 916.

⁴⁵ Rawls, *A Theory of Justice*, p. 7; see also *Political Liberalism*, p. 259.

affect the social fortune of citizens, their opportunities in life, and the actual terms of cooperation between them.”⁴⁶ Institutions are what turn natural contingencies into crucial determinants of a person’s life prospects, and hence justice would require that institutions be appropriately arranged so as to not convert natural contingencies into actual social disadvantages for individuals.

The goal of luck egalitarianism is to minimize the effects of luck on people’s life chances. And arranging social institutions in the appropriate ways provides a strategy for realizing this end. But what about the impact of luck within the rules of institutions on personal pursuits? Surely luck also affects the outcome of people’s personal choices within the rules of institutions. Should luck in the personal sphere not also be of concern to the luck egalitarian given her goal of mitigating the impact of luck on people’s life prospects?

An immediate response here might be to stress that luck egalitarians are concerned with mitigating the effects of “brute luck” but not “option luck” to use Ronald Dworkin’s distinction.⁴⁷ The outcome of most freely and fairly undertaken business ventures (within the rules of just institutions), for instance, are often luck-dependent to a degree. But this is option luck in that it pertains to choices that individuals have freely made. That is, one freely chooses to engage in a business venture, and so for this reason the outcome of such a venture is not a matter of the kind of luck that luck egalitarians need to nullify, even though how things actually do turn out in this venture is to some degree affected by luck. On the other hand, the inheritance one acquires, the ability or disability that one is born with, and so on, are due to brute luck in that they are not chosen at all. Unlike, say, business ventures, these are simply situations that one is born into. So, while an institutional approach does not nullify the effects of option luck, neither does it *aim* to nullify such effects given that these are fundamentally the results of the choices people have made. The institutional approach aims only to mitigate the effects of brute luck. Brute luck pertains to the background conditions within which people make choices, and background inequalities are the sorts of inequalities (one’s natural abilities, social class, and so forth) that institutions can address.

But more importantly, an institutional approach preserves the other twin tenet of the luck egalitarian position, namely the ideal that distributive justice ought to be sensitive to the choices of individuals. Luck egalitarianism is premised, fundamentally, on the idea that dis-

⁴⁶ Rawls, “Kantian Constructivism in Moral Theory,” in *Collected Papers*, p. 337.

⁴⁷ *Sovereign Virtue* (Cambridge: Harvard, 2000), pp. 73–74.

tributive principles ought to be luck- or circumstance-insensitive, but choice-sensitive in that people should be held responsible and accountable for their choices (though not for the unchosen situations that they happen to find themselves in). Indeed, one might say that the basic moral idea behind luck egalitarianism is that people ought to be free to choose and pursue their goals, and to take responsibility for their choices and pursuits; this is precisely why conditions that compromise people's ability to choose that are not themselves the result of their choice need to be rectified. A distributive arrangement that reflects people's unchosen circumstances is unjust; but so too is a distributive scheme that does not adequately reflect people's choices, for this would have the effect of subsidizing people who have made bad choices by penalizing those who have made good choices.

The institutional approach allows us to identify the relevant background circumstances that ought to be equalized with respect to luck, while preserving, at the same time, space for individual choice and responsibility. It identifies the basic structure of society as this background condition within which individual choices are made, and so the impact of which on people's life prospects should not be subject to luck. Individual choices within the rules of the basic structure are, however, not defined as matters of luck on this view, and luck egalitarians would not wish to mitigate the effects of these on distributive outcomes. A limited focus on institutions avoids imposing a distributive requirement on individuals that is choice-insensitive because it provides a reasonable way of demarcating choice and circumstance.

So, while it is certainly arguable that luck may affect a person's options even in the personal sphere, to nullify the effect of luck in this context by imposing egalitarian requirements on personal conduct will be at the high cost of sacrificing the commitment to choice-sensitivity. As long as the effects of luck can be sufficiently (even if not completely) mitigated by institutional means, any attempt at countering the effects of luck in personal conduct by interfering with personal pursuits within the rules of just institutions will be overly broad.⁴⁸ The residual inequalities of luck on personal life is acceptable

⁴⁸ Cohen has argued that the person who is simply unlucky to be born with expensive needs, for example, to be a photographer, is unlike another who has an expensive taste for champagne because he "schooled himself into it"—"On the Currency of Egalitarian Justice," pp. 914 and 923. On Cohen's view, the former (though not the latter) ought to be compensated somehow for her "bad" luck because she was "stuck" with her tastes, unlike the latter who freely chose to acquire his. In his reply to Cohen, Dworkin writes that Cohen's recommendation does not track "ordinary people's ethical experience," to wit, that people take responsibility for the personalities that they have (even if these are matter of brute luck) (pp. 289–90). Note also this related comment by Rawls that "it is a normal part of being human to cope with the preferences that our upbringing leaves us with"—*Political Liberalism*, p. 185n. My response is a different one, namely, that once we begin compensating

given the greater costs of attempting to eliminate these inequalities. The institutional approach, in other words, provides a practicable means of realizing the luck egalitarian core ideal, namely, that distributive justice should be circumstance-insensitive but choice-sensitive. It is better able to identify and preserve the choice/circumstance distinction than an ethos-based approach to justice (such as Cohen's).

Let me summarize the arguments in this section. On the view that equality matters because of the ideal of reciprocity, egalitarian institutions are required, but not necessarily egalitarian interpersonal conduct as such. Indeed, the ideal of reciprocity should allow for nonegalitarian personal pursuits within egalitarian rules. On the alternative view that equality matters because inequalities due to luck are unacceptable, egalitarian institutions can go a long way towards mitigating the effects of luck; but going beyond this institutional focus risks compromising the other luck egalitarian ideal, namely, that inequalities due to personal choices must be respected. An institutional focus is thus necessary *but also sufficient* for the purpose of egalitarian justice. It is necessary because establishing the right kinds of institutions is required as a matter of reciprocity (on the derivative argument for equality); or is required for mitigating the effects of luck on people's life chances (on the direct argument for equality). But it is also sufficient in that going beyond institutions and requiring egalitarian conduct across the board in society will undermine the ideal of reciprocity; or it will have the consequence of annulling the effects not just of luck but also of those due to individual choice and ambition.

Thus, there are no reasons on grounds of egalitarian justice to require that individuals adopt egalitarian principles in the thick of daily life. A society with an ethos of justice in Cohen's sense, a society whose individuals have internalized the requirements of justice and are naturally motivated to act in ways that have the best interests of the worst-off at heart, is ultimately not a society that is required as a matter of justice. Indeed, such a society can be described as a society "beyond justice" in that one of the circumstances of justice, that is, the conditions that make the consideration of justice relevant, no longer obtains. Recall David Hume's point that the usefulness of justice would be "suspended" were the human "mind so enlarged,

the aspiring photographer for the expensive pursuit that she is (unfortunately) stuck with, we cannot avoid also compensating the champagne lover for the habit that he has actively trained himself to have. There is no way of telling the two apart. The institutional approach is willing to permit brute luck to impact people's lives in this limited way because any interference with this would undermine the importance of respecting choice altogether.

and so replete with friendship and generosity, that every man has the utmost tenderness for every man, and feels no more concern for his own interest than for that of his fellows."⁴⁹ In Cohen's ideal society in which personal pursuits are motivated by egalitarian ideals across the board, individual's minds have become "so enlarged" that justice would be a redundant concept. Cohen's egalitarian society would be a society in which justice ceases to be a relevant concept. It would be an admirable society, a society with virtually unqualified equality as Cohen says; but it would not be a society required as a matter of egalitarian justice even if humanly attainable.⁵⁰

To summarize, Rawls's institutional approach is not as permissive (of inequalities) as Cohen's criticism suggests (section II). Moreover, not only is (Cohen's account of egalitarian justice or) an ethos-based approach too restrictive with respect to personal pursuits (section III), there is also no reason on account of egalitarian justice to go beyond the institutional approach (section IV).

V

Let me end with some reflections on the fundamental differences underlying the institutional approach and that of its critics. At bottom, the disagreement between the institutional view and its critics seems to concern the goal of political philosophy, a disagreement that divides the "ancients" (for example, Plato, Aristotle) from the "moderns" (for example, Kant).⁵¹ The critics of the institutional view seem to share with the ancients the belief that political philosophy is concerned too with how to instill in people the right virtues, or, if we may put it thus, how to make *people* just. On this view, the objectives and scope

⁴⁹ *An Enquiry Concerning the Principles of Morals* (Indianapolis: Hackett, 1977 [1777]), pp. 21–22.

⁵⁰ Given Rawls's own starting assumptions about the Humean circumstances of justice (*A Theory of Justice*, pp. 126ff.), Rawls cannot mean, when he speaks of individuals in a well-ordered society internalizing the principles of justice or taking them to heart, that individuals have become egalitarians through and through in the way Cohen describes (so as to be motivated by impartial egalitarian considerations in all arenas of their life). But I will not pursue this point further here. I thank Erin Kelly for suggesting the relevance of the notion of "beyond justice" here.

⁵¹ For the ancients (Plato's *Republic* is a good example), political philosophy is continuous with ethical philosophy in that political principles are simply ethical principles writ large. The aim of a just society is to make people virtuous. For the moderns (Kant's distinction between duties of virtue and duties of justice is, I think, a good example), there is a discontinuity between ethics and political philosophy, and the goal of political philosophy is to create appropriate institutions to regulate the interaction of citizens. This distinction between the "ancient" and "modern" approaches to political philosophy is of course stylistic and impressionistic. I follow here Rawls's reference to Benjamin Constant's distinction between "the liberties of the moderns" and "the liberties of the ancients," and my use of this distinction roughly parallels Rawls's own use—*Political Liberalism*, pp. 4–5.

of ethics and political philosophy are indistinguishable. The state of one's soul and the institutions of one's state are inseparable concerns. The moderns, however, accept a distinction between justice and ethics. While ethics is concerned with how well one's life goes, justice is concerned with the background rules within which one pursues one's ends. On this view, justice does not dictate the end; it only sets the limits on the kinds of ends that are admissible and how these ends are to be pursued.

It is important here not to conclude hastily that the divide between the institutional egalitarians and its critics stems from Rawls's division between political liberalism and comprehensive liberalism. This might be a tempting conclusion to draw because Rawls points out in *Political Liberalism* that one feature of the political conception of justice is that it takes the basic structure of society to be its main subject.⁵² But this would be a fundamental misreading of Rawls. After all, Rawls already affirms the institutional approach in *A Theory of Justice*, which, as Rawls himself admits, can be plausibly read as a comprehensive liberal theory of justice. A focus on the basic structure is not unique to political liberalism, but is a feature of any modern view of liberalism as a theory of justice.⁵³ More significantly, as noted above, the institutional approach is already evident in Kant, as is reflected by his distinction between the duties of justice and the duties of virtue. To recall a familiar point, duties of justice for Kant are duties that may be institutionalized and enforced, whereas duties of virtue can only be personally motivated.⁵⁴ Yet Kant is a paradigmatic comprehensive liberal on Rawls's own description. The difference between the institutional egalitarians and its critics, therefore, reflects a much deeper difference (both philosophically and historically speaking) than the one between comprehensive and political liberalism: it reflects the different ways of understanding the relationship between justice and virtue, and, consequently, the scope and goal of political philosophy, in the ancients and the moderns.

To be exact, most modern political philosophers (including mod-

⁵² *Political Liberalism*, pp. 11–12.

⁵³ I examine the differences between political liberalism and comprehensive liberalism more fully in "Toleration and Diversity: Two Liberal Views" in my *Toleration, Diversity, and Global Justice* (University Park, PA: Penn State, 2000), pp. 47–77—see also pp. 4–10 in "Introduction." In this work, I suggest that the common criticisms made by political liberals against comprehensive liberals are either (i) overstated or (ii) apply also to political liberalism.

⁵⁴ As Kant says, duties of justice are those duties for which an external lawgiving is possible (p. 55); and duties of virtue are those for which external lawgiving "cannot be so given" (p. 185); references from Kant, *The Metaphysics of Morals*, Mary Gregor, trans. (New York: Cambridge, 1991 [1797]).

ern liberals) recognize that there are certain virtues of citizenship and that the state has a great interest in inculcating these virtues in its citizens. But these are specifically political virtues, for example, the virtue of civility, the virtue of fair cooperation, and so on; and, to be sure, modern liberals accept that the cultivation of these political virtues may require the state to take a rather active role in the education of citizens.⁵⁵ The ancients, however, take a more expansive approach to virtue that concerns not just public life but private life as well (to concern “the state of one’s soul” so to speak), and hold that it is the business of the state not just to make good citizens but good persons. On their account, the question “how ought I to live my life” is not fundamentally distinct from the question “how ought we to live together.” Indeed, for them, the solution to the second problem necessarily follows from solving the first. Virtuous individuals are also individuals who would treat others justly. Cohen’s position, that the personal and egalitarian points of view should be reconciled within a person’s soul rather than externally through a moral division between the personal and the institutional, seems to me to reflect this aspiration.

The moderns replace this dominance of virtue with the dominance of justice. “How ought I to live” is regulated by “how ought we to live together.” The rules for how we are to treat others define the parameters for how I ought to live my life. But the modern conception recognizes that other than defining the limits to how one ought to live, justice says nothing more about the matter. Other than supporting and living by the rules of just institutions, individuals need not further explain themselves. So there is a sharper divide between virtue (in the expansive extra-political sense) and justice on the modern view. This is not to say that how one ought to live is an unimportant issue, normatively speaking, for contemporary political philosophers; but it is a question of ethics, not of political philosophy as such. The institutional approach reflects this modern division between justice and ethics. In contrast to an ethos-based view of justice, one might say that, on the institutional approach, the primary goal of political philosophy is not to make people just, but to make the right institutions within which rules people can pursue their diverse ends justly.

KOK-CHOR TAN

The University of Pennsylvania

⁵⁵ For some accounts of liberal political virtues, see William Galston, *Liberal Purposes: Goods, Virtues, and Duties in the Liberal State* (New York: Cambridge, 1991); and Eamonn Callan, *Creating Citizens: Political Education and Liberal Democracy* (New York: Oxford, 1994). See also Rawls, *Political Liberalism*, pp. 199–200.

REFLECTION WITHOUT EQUILIBRIUM*

Justice is a concept by far more subtle and indefinite than is yielded by mere obedience to a rule.—Benjamin N. Cardozo

One of John Rawls's most abiding contributions to moral and political philosophy is his idea of reflective equilibrium. Even many who dissent from Rawls's principles of justice or contractarian framework find reflective equilibrium an apt characterization of philosophical method. Rawls provides a compelling if somewhat vague account of ethical reflection as going "back and forth" between considered judgments and principles, adjusting each in light of the other.

Why, however, should we expect the process of reflection Rawls outlines to lead to equilibrium? Surprisingly, Rawls offers little argument. In defining that state, he makes it clear that equilibrium is simply an assumption. "By going back and forth," Rawls writes, "...I assume that eventually we shall find a description of the initial situation that both expresses reasonable conditions and yields principles which match our considered judgments duly pruned and adjusted."¹ Most other writers have been no less sanguine. Michael Sandel, for example, discussing the process of mutual adjustment, simply remarks, "a final product emerges."²

* I delivered a version of this paper at a symposium in memory of John Rawls at the University of Texas at Austin in March 2003. I am grateful to my fellow participants and audience members—especially James Fishkin, David Braybrooke, T. K. Seung, Jay Budziszewski, Benjamin Gregg, Paul Lyon, and Michael O'Connor—for their helpful comments and advice, and to an anonymous referee, Bryan Register, John Messerly, and especially Anthony Gillies for their comments on later drafts.

¹ *A Theory of Justice* (Cambridge: Harvard, 1971), p. 20, my emphasis; hereafter TJ.

² *Liberalism and the Limits of Justice* (New York: Cambridge, 1982), p. 48. The exception that proves the rule: Geoffrey Sayre-McCord recognizes that "actually achieving a comprehensive reflective equilibrium will almost surely remain always at most an ideal"—"Coherentist Epistemology and Moral Theory," in Walter Sinnott-Armstrong and Mark Timmons, eds., *Moral Knowledge?* (New York: Oxford, 1996), pp. 137–89, on p. 142. His concerns are primarily practical rather than theoretical. We do however share a crucial premise: that reflection does not merely cull initially given intuitive judgments and principles but also generates them. As he puts it, "new commitments will come on board thanks sometimes just to expanding experience, and other times to seeing what is implicit in, or required by, what else one believes" (p. 141). This is an important distinction. Judgments at equilibrium are not immune to all revision; they remain stable unless "expanding experience" alters them. So, new commitments arising from expanding experience do not challenge the possibility of reflective equilibrium. Those arising from responses to currently held judgments and principles, however, do.

I will argue that we have no reason to expect equilibrium to emerge from the process of reflection Rawls describes. Indeed, I shall argue that, on Rawls's own conception, the *equilibrium problem*—the question whether reflection will reach equilibrium in a finite time—is unsolvable. That has important implications for his overall view, since reflective equilibrium is “the test,” the “criterion” we use to determine which conception of justice, “so far as we can now ascertain, is the one most reasonable for us.”³ If the equilibrium problem is unsolvable, so is the problem of selecting an optimal conception of justice.

The equilibrium problem also undermines Rawls's argument against intuitionism. If the equilibrium problem is undecidable, so is the dispute between intuitionism and Rawlsian constructivism.

We can nonetheless develop a concept of reflection and an alternative to equilibrium that can play much the same methodological role as reflective equilibrium without any commitment to termination of the process after a finite time. The result, however, is to transform Rawls's Kantian constructivism into a pragmatic intuitionism.

I. RAWLS'S ASSAULT ON INTUITIONISM

“Intuitionism,” James Fishkin writes, is the “doctrine Rawls is most concerned to argue against” in *A Theory of Justice*.⁴ Certainly, in Rawls's view, intuitionism and utilitarianism are the two chief competitors to his own theory. Yet it is difficult to discern arguments against intuitionism in *A Theory of Justice* or, for that matter, in the rest of Rawls's works.⁵ This should not surprise us, however, for Rawls intends *A Theory of Justice* as a whole as one long argument against intuitionism.

³ *Political Liberalism* (New York: Columbia, 1993), p. 28.

⁴ *Beyond Subjective Morality: Ethical Reasoning and Political Philosophy* (New Haven: Yale, 1984), p. 17.

⁵ Rawls does argue against a position he calls “rational intuitionism” in *Political Liberalism*, among other places, but that is quite a different view, which will not concern us here. Intuitionism, as understood in *A Theory of Justice*, is a kind of value pluralism. Rational intuitionism, in contrast, is a kind of realism. Rawls characterizes it differently in different works. In “Kantian Constructivism in Moral Theory,” this JOURNAL, LXXVII, 9 (September 1980): 515–72, reprinted in *Collected Papers* (Cambridge: Harvard, 1999), pp. 303–58, he defines it as the twofold thesis that moral concepts do not reduce to nonmoral concepts and that basic moral judgments are self-evident (p. 343). In “Themes in Kant's Moral Philosophy,” in Eckhart Förster, ed., *Kant's Transcendental Deductions: The “Three Critiques” and the “Opus Postumum,”* (Stanford: University Press, 1989), pp. 81–113, reprinted in *Collected Papers*, pp. 497–528, Rawls adds the condition that moral first principles “are regarded as true or false in virtue of a moral order of values that is prior to and independent of our conceptions of person and society, and of the public social role of moral doctrines” (p. 511). *Political Liberalism* (pp. 91–92) defines rational intuitionism as consisting of four theses, which amount to the view that reason can discover through intuition mind-independent moral facts. Rational intuitionism, as understood in any of these senses, is plainly logically independent of pluralism.

Indeed, he thinks it offers the only possible kind of argument against that view:

The intuitionist believes...that the complexity of the moral facts defies our efforts to give a full account of our judgments and necessitates a plurality of competing principles. He contends that attempts to go beyond these principles either reduce to triviality, as when it is said that social justice is to give each man his due, or else lead to falsehood and oversimplification, as when one settles everything by the principle of utility. The only way therefore to dispute intuitionism is to set forth the recognizably ethical criteria that account for the weights which, in our considered judgments, we think appropriate to give to the plurality of principles. A refutation of intuitionism consists in presenting the sort of constructive criteria that are said not to exist (TJ 39).

Rawls conceives intuitionism as a threefold thesis:

- (1) *Pluralism*: values differ in kind.
- (2) *Conflict*: they compete with one another.
- (3) *Complexity*: there are no higher-order rules or principles for determining the outcome of these competitions in every case.

Morality is so complex that it cannot be captured by rules or principles. The problem of determining the outcomes of value conflicts or competitions Rawls terms the *priority problem*. "Intuitionism denies that there exists any explicit and useful solution to the priority problem" (TJ 40). The only way to refute the doctrine, then, is to present such a solution: a set of rules or principles that "match our considered judgments duly pruned and adjusted."

This is just what Rawls's constructivism means to do. As he understands it, it is more than the thesis that we construct rather than discover moral value. Onora O'Neill⁶ observes that Rawls's theory is constructive in the more specific sense that it can settle disputes that intuitionism cannot. It solves the priority problem. For Rawls, constructivism should be understood as comprehending the thesis that our considered judgments can be brought into reflective equilibrium with a set of universal principles.

I follow R.M. Hare in interpreting Rawls as seeking principles that are strictly universal, without *ceteris paribus* or similar clauses—principles I term *stouthearted*.⁷ In Ronald Dworkin's terms, Rawls is

⁶ *Constructions of Reason* (New York: Cambridge, 1989), p. 207.

⁷ "Rawls's Theory of Justice," *Philosophical Quarterly*, xxiii (1973): 144–55, reprinted in Norman Daniels, ed., *Reading Rawls* (Stanford: University Press, 1973), pp. 81–107, here p. 92. Rawls says relatively little about the logical form of the rules he seeks. Hare's reference to Rawls's discussion of natural duties (115–16) strikes me as unilluminating. Rawls there argues that *some* principles reached at reflective equilibrium are unconditional, not that *all* are. In any case, 'unconditional' is not synonymous with 'universal' or 'stouthearted'. Rawls's discussion of the priority problem is more

seeking *rules* rather than principles.⁸ If we could rest content with *fainthearted* principles—those with *ceteris paribus* or similar clauses, for example, “that,” as Aristotle says, “hold good only as a general rule, but not always” (*Nicomachean Ethics* 1094b22)—there would be no dispute at all between Rawls and the intuitionist. We might reach reflective equilibrium, settling on a set of principles in full harmony with our considered judgments, without solving the priority problem. If we were to settle on a view that fails to tell us how to resolve conflicts between principles, Rawls says, “the means of rational discussion have come to an end”; we would attain “but half a conception” (TJ 41).

I shall call Rawls’s theory *stoutheartedly constructive*, then, in the sense that it provides stouthearted principles solving the priority problem. This is an advantage only if the priority problem has a solution. As Joel Feinberg points out,

If the sort of ‘explanation’ Rawls seeks is in principle achievable, then the theory that supplies it carries the day. But there is also the possibility that rigid priority rules are, in the very nature of the case, impossible to formulate. Other things being equal, simplicity is preferable to complexity, but a distorting simplicity is worse than none.⁹

Is the priority problem solvable? Can we bring our considered judgments into line with stouthearted principles that resolve conflicts in every possible circumstance? We are back to our original question: Can we expect ethical reflection to reach equilibrium?

II. REFLECTION

This brings us to the process of reflection itself. We begin with a set J_0 of considered judgments. We examine our moral intuitions, that is, and throw out those that are unstable, vague, or in which we lack confidence. We retain only those we are willing to affirm confidently after careful thought. We provisionally adopt a decidable set P_0 of principles from which, we conjecture, the considered judgments might be derived.¹⁰ We strive to articulate principles that are logically

helpful; see section III below. My term ‘stouthearted’ is meant to contrast with Michael Morreau’s ‘fainthearted’; see “Fainthearted Conditionals,” this JOURNAL, xciv, 4 (April 1997): 187–211.

⁸ See “Is Law a System of Rules?” in R.S. Summers, ed., *Essays in Legal Philosophy* (Berkeley: California UP, 1976), pp. 25–60, and *Taking Rights Seriously* (Cambridge: Harvard, 1978).

⁹ “Rawls and Intuitionism,” in Daniels, ed., pp. 108–23, on p. 111.

¹⁰ I assume that the set of principles is decidable to capture the idea that we should think of them as ethical axioms. This seems faithful to Rawls’s conception of reflection as a process of assessing principles in light of considered judgments; to *assess* our principles within a finite time, we surely must be able to decide in a finite time, at any given stage, what they *are*. (All of Rawls’s examples, including his own favored set of principles, are not only decidable but finite.) The argument to follow goes through, however, even if we weaken this to a requirement that the set of principles be enumerable or, equivalently, axiomatizable.

weak and generally shared, or are derivable from logically weak and generally shared constraints on an original position. We then proceed to reflect on our judgments and principles against a background of relevant theories—theories of persons, of society, of the place of morality in society, of moral education and development, and so on.¹¹ Call this set of background theories T_0 . Additionally, Rawls treats reflective equilibrium, in the primary instance, as justifying principles of justice only indirectly, by way of justifying conditions placed on the original position. In addition to principles, judgments, and background theories, then, we should take into account constraints C_0 placed on an ideal or actual circumstance of choice.

Reflection proceeds in stages. At any stage $n + 1$, we search for discrepancies between the sets J_n of considered judgments and P_n of principles given our set T_n of background theories and our set C_n of constraints placed on the choice situation at the previous stage. Such discrepancies might take one of several forms: we might find

- (1) A considered judgment p in J_n that cannot be derived from P_n , the negation of which can be derived from P_n . That is, our principles might contradict our considered judgments in the sense that P_n implies $\neg p$ even though p belongs to J_n .
- (2) A considered judgment p in J_n such that neither it nor its negation can be derived from P_n . That is, our principles might be too weak to yield some of our considered judgments; P_n might imply neither p nor $\neg p$ even though p belongs to J_n .
- (3) An actual inconsistency in P_n (or J_n), that is, a judgment p such that both p and its negation $\neg p$ can be derived from P_n (or J_n).
- (4) A conflict or potential inconsistency in P_n (or J_n), that is, a judgment p such that both p and its negation $\neg p$ can be derived from P_n (or J_n) together with information specifying a possible circumstance.
- (5) A judgment p derivable from P_n such that neither p nor its negation $\neg p$ are in J_n .

At any stage, there might be infinitely many such discrepancies.

We then select one or more discrepancies to address. We take steps to reconcile our principles and considered judgments, depending on

¹¹ Daniels has elaborated reflective equilibrium along these lines in "Wide Reflective Equilibrium and Theory Acceptance in Ethics," this JOURNAL, LXXVI, 5 (May 1979): 256–82, "On Some Methods of Ethics and Linguistics," *Philosophical Studies*, xxxvii (1980): 21–36, "Reflective Equilibrium and Archimedean Points," *Canadian Journal of Philosophy*, x (1980): 83–103, and *Justice and Justification: Reflective Equilibrium in Theory and Practice* (New York: Cambridge, 1996). This seems faithful to Rawls's intentions in *A Theory of Justice, Political Liberalism*, however, seems concerned to free the theory from commitments to theories of the self, and so forth.

the kind of discrepancy found and selected. For the corresponding problems above:

- (1) We either revise the considered judgment, adopting its negation instead, or revise our principles to drop or weaken one or more to prevent the negation of our considered judgment from being derivable, or both.
- (2) We add or strengthen a principle to make the considered judgment derivable.
- (3) We drop or weaken one or more principles to remove the inconsistency by making one or both judgments underivable.
- (4) We examine the judgment derivable from P_n (or J_n) to see whether we are willing to include it or its negation among our considered judgments. If so, we add it to J_n . If not, we drop or weaken principles to make one or both underivable.
- (5) We examine p and $\neg p$ to see whether either should be added to J_n .
- (6) We reconsider our (revised) principles, asking whether it is possible to reformulate them to account for our (revised) considered judgments more elegantly and efficiently. Throughout, it is important that we be able to recognize our principles when we see them; P_n must remain decidable.
- (7) We reconsider our background theories T_n and our constraints C_n , asking whether it is advisable to reformulate them in light of the outcome of the adjustments made in our judgments and principles.
- (8) We continue this process, returning to step 3 above, until we reach a fixed point at which $J_{n+1} = J_n$ and $P_{n+1} = P_n$. That point is reflective equilibrium.

I have gone through this process in detail to make several points. First, its starting points radically underdetermine the outcome of Rawls's procedure.¹² Ethical reflection comprises several kinds of belief revision. Rawls says little about how such revision is to be conducted. The literature on belief revision suggests both that this problem is quite general and that it has no simple uncontroversial solution. Without specifying the details of the revision procedures, the process of reflection as a whole is underdetermined.

The process also poses some special problems. The significance of classifying something as a principle or a judgment must be sorted out. Should we construe reflection as updating principles with considered judgments, updating judgments with principles, going back and forth between these, or updating both together? If principles and judgments play different roles, the results will not necessarily be equivalent. We

¹² See D.W. Haslett, "What Is Wrong with Reflective Equilibrium?" *Philosophical Quarterly*, xxxvii (1987): 305–11, on p. 310.

can raise similar points about background theories and constraints on the choice situation.

The process of reflection, moreover, requires the resolution of some open problems in the theory of belief revision. First, should we conceive of Rawls's procedure as an iterated process, taking discrepancies between principles and considered judgments one by one, or as a multiple revision process, taking all discrepancies into account together? Since we may face many, even infinitely many, discrepancies between principles and judgments at any stage, the latter is probably more natural.¹³ Furthermore, any given judgment may or may not be accepted when presented as a candidate for updating; the revision in question is nonprioritized.¹⁴ If we think of principles and judgments being revised together, with no differentiation of role, the problem is essentially one of consolidation.¹⁵ This is particularly difficult, since standard belief revision procedures, inspired by possible worlds semantics, treat all inconsistent sets alike, as identical to the entire language.¹⁶ In any case, there is no settled consensus on how any of these belief revision problems ought to be solved.

If we think of Rawls's procedure as an iterated process, the outcome of the procedure—the set of judgments and principles on which we settle at reflective equilibrium, if we attain it, or the pattern of variation in judgments and principles, if we do not—may depend not only on the initial and subsequent considered judgments, the initial selection of principles, and the details of the revision procedures, but also on the selection of a discrepancy to be addressed at each stage of the revision process. The order in which we address discrepancies may

¹³ For varying approaches to multiple revision, see Sven Ove Hansson, "New Operators for Theory Change," *Theoria*, LV (1989): 114–32; R. Niederée, "Multiple Contraction: A Further Case against Gärdenfors's Principle of Recovery," in André Fuhrmann and Morreau, eds., *The Logic of Theory Change* (Berlin: Springer, 1991); Fuhrmann and Hansson, "A Survey of Multiple Contraction," *Journal of Logic, Language, and Information*, III (1994): 39–76; and J.A. Li, "Note on Partial Meet Package Contraction," *Journal of Logic, Language, and Information*, VII (1998): 139–42.

¹⁴ See J.R. Gallier, "Autonomous Belief Revision and Communication," in Peter Gärdenfors, ed., *Belief Revision* (New York: Cambridge, 1992), pp. 220–46; Hansson, "Ten Philosophical Problems in Belief Revision," *Journal of Logic and Computation*, XIII, 1 (February 2003): 37–49.

¹⁵ See Hansson, "Taking Belief Bases Seriously," in Dag Prawitz and Dag Westerstahl, eds., *Logic and Philosophy of Science in Uppsala* (New York: Kluwer, 1994), pp. 13–28; "Semi-Revision," *Journal of Applied Non-Classical Logic*, VII (1997): 151–75; E.J. Olsson, "A Coherence Interpretation of Semi-Revision," *Theoria*, LXIII (1997): 105–34, and "Coherence: Studies in Epistemology and Belief Revision" (Ph.D. diss., Uppsala University, 1997).

¹⁶ An important exception is Fuhrmann—see "Theory Contraction through Base Contraction," *Journal of Philosophical Logic*, XX (May 1991): 175–203, and *An Essay on Contraction* (Stanford: CSLI, 1997).

make a significant difference to the outcome of the process. A foolish principle of selection, for example, might lead us to neglect some discrepancies entirely. Say that a principle of selection is *adequate* if it raises each discrepancy at some stage, thus enabling the process to yield reflective equilibrium if that can be attained at all. There are many possible adequate selection principles. In general, they may not yield identical outcomes. The process of reflection may therefore be path-dependent. The procedure of Peter Gärdenfors is path-independent, but at the cost of holding the degree of justification or vulnerability for each proposition constant throughout the revision procedure.¹⁷ Surely that is implausible here. Our ethical reflection should be able to change the degree of justification or vulnerability of principles and judgments; indeed, that seems to be its chief purpose. New considered judgments may make previously adopted principles more doubtful. Procedures that allow degrees of justification or vulnerability to vary dynamically, however, such as that of Wolfgang Spohn, are not path-independent.¹⁸ Thus, two people may reach different reflective equilibria, even if they start with the same considered judgments and principles, use the same (possibly deterministic) revision rules, and judge similar conflicts similarly throughout the revision process.

Second, whether we think of revision as iterated or multiple, choosing between principles and judgments, for example, requires assessing the relative degrees of justification or vulnerability of those judgments and principles. But there is no consensus about how to represent such dynamic information in a belief revision procedure.¹⁹

Third, the decisions that have to be made, especially when facing discrepancies of kinds (1) and (4), are distinctly ethical decisions, requiring intuitive judgment as well as (or as a part of) general considerations of belief revision. Intuitive judgment, that is, enters the process not only at the initial stage but also at many revision stages.

Fourth, the extent to which the process of reflection and the attainment of reflective equilibrium count as justifying a judgment or princi-

¹⁷ "Epistemic Importance and Minimal Changes of Belief," *Australasian Journal of Philosophy*, LXII (1984): 136–57. As Anthony Gillies points out—"New Foundations for Epistemic Change," *Synthese*, cxxxviii (2004), in press—Hansson-style counterexamples to AGM revision trade on having a single revision function govern an iterated procedure. If the revision function changes as the revision proceeds, all the dynamic work is done by those changes, about which the AGM approach has nothing to say.

¹⁸ "Ordinal Conditional Functions: A Dynamic Theory of Epistemic States," in W.L. Harper and B. Skyrms, eds., *Causation in Decision, Belief Change, and Statistics II* (Boston: Kluwer, 1988), pp. 105–34.

¹⁹ In addition to Spohn, see Adnan Darwiche and Judea Pearl, "On the Logic of Iterated Belief Revision," *Artificial Intelligence*, lxxxix (1997): 1–29; and John Pollock, "Defeasible Reasoning with Variable Degrees of Justification," *Artificial Intelligence*, cxxxiii (2002): 233–82.

ple depends crucially on the properties of those revision procedures. Some have doubted whether the kind of process Rawls envisions could ever have justificatory force.²⁰ Others find it “easy to see how this could be a procedure for rationalization of individual or social norms, or, to put it in more elevated terms, a procedure for the ‘construction’ of moral or ethical systems.”²¹ Crucial to any justificatory force, according to Norman Daniels, is the independent support enjoyed by the background theories.²² Surely, whatever justificatory role Rawls’s procedure has depends on the status of the intuitive judgments with which it begins, the intuitive judgments made during the revision process, and further features of that process.

Fifth, and crucially, the revision procedure is not monotone; judgments and principles may be added or dropped. We may delete some intuitive judgments that contradict or fail to find support in our principles; we may drop principles that contradict or fail to find support in our considered judgments. In fact, we *must* do so. As Hans Rott has shown, every revision forces us to surrender some truths, unless we can leap to the truth in a single step.²³ If we could do that, of course, reflection would be unnecessary. So, if we are revising principles, we must trade in some true principles to secure the revision. If we are revising with respect to considered judgments, we must trade in some true judgments to do so.

We may also add principles and considered judgments as we reflect. Indeed, a large part of ethical reflection seems to involve imagination: of being in another’s place, of the consequences of policies, of the outcomes of events, of situations that might test principles, of the intuitive reactions we would have in those situations, and so on.²⁴ Reflecting on the principle of utility, for instance, has led us to consider and formulate judgments concerning various applications of the principle—Bernard Williams’s summary execution puzzle, to take just one example—that had never occurred to anyone before.²⁵ Reflecting on such problems, moreover, has led to formulations of heretofore unenvisioned principles such as rule utilitarianism or coopera-

²⁰ See, for example, David Lyons, “Nature and Soundness of the Contract and Coherence Arguments,” in Daniels, ed., pp. 141–68.

²¹ Richard Boyd, “How To Be a Moral Realist,” in Sayre-McCord, ed., *Essays on Moral Realism* (Ithaca: Cornell, 1988), pp. 181–228, here see p. 185.

²² *Justice and Justification: Reflective Equilibrium in Theory and Practice* (New York: Cambridge, 1996).

²³ “Two Dogmas of Belief Revision,” this JOURNAL, xcvi, 9 (September 2000): 503–22.

²⁴ See Simon Blackburn, *Ruling Passions* (New York: Cambridge, 1996); Sayre-McCord, “Coherentist Epistemology and Moral Theory.”

²⁵ For this and various other cases, see J.C.C. Smart and Williams, *Utilitarianism: For and Against* (New York: Cambridge, 1973).

tive utilitarianism. Reflection consists in more than paring down sets of judgments and principles; it frequently involves the formulation of new judgments and principles.

But proofs that a process reaches a fixed point typically depend on the monotonicity of the process. So, apart from a detailed specification of the revision procedures—and probably not even then—there is no way to construct a general argument that the process must terminate. We have no reason to expect a fixed point. We in particular have no reason to expect a fixed point at a finite stage of reflection. But that is what Rawls's concept of reflective equilibrium would require, at least if it is to have the justificatory force he envisions in the way he envisions.

To be sure, an agent engaged in Rawls's process of reflection has an advantage over a typical inquirer responding to potentially recalcitrant empirical evidence: Rawls's process is *a priori*.²⁶ The revision process starts from a set of considered judgments and a set of principles, *inter alia*, recognizes logical discrepancies between them, and revises one or both to remove those discrepancies. Does the *a priori* character of the process secure a fixed point? While neither is conclusive, I see two reasons to expect a negative answer. First, reflection on discrepancies must still proceed in a temporal sequence under only partial control of the reflecting agent. In practice there seems to be no way, even *a priori*, to survey the entire field of possible discrepancies and arrange them for consideration in some optimal order. Nor can an agent survey the outcome of the process in advance. Path dependence, the representation of dynamic information, the role of intuitive judgment, nonmonotonicity, and the expansion of the sets of general principles and considered judgments remain serious issues. All but the first, in fact, would remain serious issues even if we could arrange discrepancies in an optimal order. Second, for Rawls, Daniels, Richard Boyd, and others, the analogy between the *a priori* process of reflection and the *a posteriori* process of empirical inquiry constitutes the chief argument for reflective equilibrium's epistemic efficacy. The mutual adjustment of general principles and particular considered judgments, they contend, confirms a moral theory in the way that the mutual adjustment of theory and evidence confirms a scientific theory. It seems likely that, if the analogy between *a priori* and *a posteriori*

²⁶ I owe this point to an anonymous referee. But see the caution raised in note 2 above: the task of reaching equilibrium on the basis of our current experience and that of regaining equilibrium once our experience has been expanded may each be *a priori* in the sense of being independent of any further experience, but the overall process, which includes expansions of experience that bear on our ethical judgments, is *a posteriori*.

adjustment procedures is strong enough to support the epistemic efficacy of reflective equilibrium, it is also strong enough to raise the belief revision issues I have discussed.

III. COMPLEXITY AND CONSTRUCTION

Intuitionism maintains that the moral realm is so complex that ethical reflection will not reach a fixed point at any finite stage of reflection—at least if the principles are conceived as stouthearted rather than fainthearted. Traditionally, intuitionists in this sense have embraced moral and epistemological realism, contending that we have immediate access to a mind-independent realm of moral value. But nothing about the complexity thesis forces that commitment. We can see moral value as complex *and* constructed.

Dworkin unwittingly offers a model of how this might be done. He contrasts the *natural model*—according to which we discover objective truths about a mind-independent moral reality through a faculty of intuition—with the *constructive model*. It assumes that we “have a responsibility to fit the particular judgments on which [we] act into a coherent program of action, or, at least, that officials who exercise power over other men have that sort of responsibility.”²⁷ It assumes nothing about the mind-independent existence of moral value or our epistemic access to it. His idea is that intuitionism adopts the natural model, while Rawls adopts the constructive model.

The constructive model, nevertheless, may be highly attractive to an intuitionist, particularly one who seeks to avoid metaphysical commitments and attendant epistemological difficulties. The model is constructive in a weaker sense than that of Rawls; it treats moral value as constructed, but does not require that principles must be stouthearted. Or so I shall argue. Agents may “fit particular judgments into a coherent program of action” on the basis of fainthearted principles.

The best argument in favor of an intuitionist constructivism is Dworkin’s chief illustration of the model: common law adjudication. A judge must infer from previous cases general principles to use in judging further cases. Sometimes, those general principles may fit the cases nicely and permit clear judgments on further cases. Sometimes, however, there are tensions in the pattern of previous cases, so that it is difficult to find a set of principles that covers them all. Sometimes the principles that emerge from consideration of previous cases lack intuitive appeal or lead to counterintuitive consequences. Sometimes the principles conflict. Sometimes they are vague. Sometimes they

²⁷ “The Original Position,” *University of Chicago Law Review*, XL, 3 (Spring 1973): 500–33; reprinted in Daniels, ed., pp. 16–52, see p. 28; hereafter OP.

fail to anticipate new issues that arise in further cases. On this conception, a “judge tries to reach an accommodation between these precedents and a set of principles that might justify them and also justify further decisions that go beyond them” (OP 28).

The previous judgments, of course, are analogous to considered judgments in Rawls’s scheme. The process that a judge undertakes in applying common law, and, even better, that a society undertakes in adopting a system of common law, is analogous to Rawls’s due reflection. The respect for precedent shown by the judge is analogous to our respect for considered judgments in the process of reflection. The responsibility of the judge to articulate principles available for public examination is analogous to our responsibility to act on the basis of principle.

Here, however, the analogy with Rawls’s scheme breaks down. Theorists of common law adjudication typically do *not* think that judges, now or even in the long run, can articulate a set of stouthearted principles that will cover all possible cases, or even all possible cases of kinds considered up to now. If that were so, we could see the common law as a temporary stand-in for explicit legislation. It *might* be conceived that way; one might think of common law as a path toward the articulation of an ideal set of laws—a more reliable path, perhaps, than the enactments of a legislature. But one might also think of the common law as articulating gradually a body of *ceteris paribus* principles, constructing, extending, revising, and resolving conflicts between them as cases arise, with a view to past, current, and possible future cases.²⁸ There is no firm stouthearted rule for how these constructions, extensions, revisions, and resolutions are to be performed. The complexity of the world and the range of possible cases it offers prevent that. They also provide the chief argument for the common law. On that conception, then, common law adjudication and the constructive model it illustrates support an intuitionist conception of moral reflection.

The responsibility of judges to articulate principles of adjudication available for public examination is fundamental to the common law. So, too, our responsibility to articulate principles upon which we act is a fundamental part of ethical reflection and reasoning. Dworkin seems to find that responsibility incompatible with intuitionism. He is not alone. Stuart Hampshire writes, “the force of the word ‘intuition’ is to suggest that the conclusion is not established by any recognized form of argument, by any ratiocinative process involving a succession

²⁸ See, for example, Benjamin N. Cardozo, *The Nature of the Judicial Process* (New Haven: Yale, 1921), and *The Growth of the Law* (New Haven: Yale, 1924).

of steps which are logically criticisable...."²⁹ Richard Brandt concurs: "What does not offer a novel way of organizing our views [about justice and the right] is pluralistic intuitionism."³⁰ Torbjörn Tännsjö contends that, on the intuitionist's account, there is no general structure to moral thinking; moral facts are capricious and moral conclusions are groundless.³¹

But the intuitionist can agree with Dworkin that "decisions made in the name of justice must never outstrip an official's ability to account for those decisions in a theory of justice," that "we [must] act on principle rather than on faith" (OP 30). One can act on the basis of principle while acting on the basis of *fainthearted* principle. As Hampshire suggests, "All argument is not deduction, and giving reasons in support of a judgment or statement is not necessarily, or even generally, giving logically conclusive reasons" (OP 473).

Imagine, for example, a judge finding someone liable for an injury due to negligence and requiring compensation. The judge might base the decision on a well-supported principle, for example, that someone whose negligence causes injury to others ought to compensate them. But this is best understood as a *fainthearted* principle. Surely the judge is not thereby committed to holding that every possible act of negligence that harms others requires payment of compensation. The judge may consistently decline to mandate compensation in another case so long as the judge can articulate the reasons for the difference in treatment. Suppose, for example, that in another case the only injury is intangible (that Jones's negligence made Smith nervous, for example) or insignificant (that Jones's negligence consisted of his dropping Smith's penny into a sewer grate). The judge may consistently refuse to require that Smith be compensated by explaining that intangible or insignificant injuries need not be compensated. Indeed, the judge may embrace a zig-zag series of principles, such as

If *A*'s negligence injures *B*, *A* owes *B* compensation.

If *A*'s negligence injures *B* only intangibly, *A* does not owe *B* compensation.

If *A*'s negligence injures *B* only intangibly, but *A* foresaw and intended to so injure *B*, then *A* owes *B* compensation.

If *A*'s negligence injures *B* only intangibly, *A* foresaw and intended to

²⁹ "Fallacies in Moral Philosophy," *Mind*, LVIII (1949): 466–82, p. 470.

³⁰ "The Science of Man and Wide Reflective Equilibrium," *Ethics*, c (1990): 259–78, see p. 273.

³¹ "In Defense of Theory in Ethics," *Canadian Journal of Philosophy*, xxv (1995): 571–94.

so injure *B*, but the injury is insignificant, then *A* does not owe *B* compensation.

In short, we can meet our responsibility to act on the basis of principle by articulating fainthearted principles, as judges and ethical agents in practice seem to do. To justify our considered judgment that a particular injury ought to be compensated, for example, we might argue in one of two ways:

Stouthearted Justification:

Jones's negligence injured Smith.

If *A*'s negligence injures *B*, *A* (invariably) owes *B* compensation.

So, Jones owes Smith compensation.

Fainthearted Justification:

Jones's negligence injured Smith.

If *A*'s negligence injures *B*, *A* (normally) owes *B* compensation.

So, Jones owes Smith compensation.

The first argument is deductively valid. Adding premises to it would never lead us to retract the conclusion. Its second premise, however, lies open to attack by counterexample. The second argument, in contrast, is valid, not deductively, but defeasibly. (It is, in Michael Morreau's terms, *allowed*.) Additional information could lead us to withdraw the conclusion by indicating that this negligence or injury was in some way abnormal or atypical—by indicating, for instance, the relevance of other moral considerations. The argument is nevertheless acceptable; in the absence of other information, the premises do make it reasonable to believe the conclusion.

Rawls and Dworkin assume that the kind of principled justification for which judges and other moral agents are responsible is the stout-hearted, deductive kind of justification typified by the first argument. Fainthearted constructivists, in contrast, contend that arguments such as the second, fainthearted one provide an acceptable sort of principled justification, one that meets our responsibilities as moral agents. They in short embrace defeasible means of justification: arguments that are acceptable but whose conclusions might have to be withdrawn in the face of further information. Such justifications may be defeated. When they are, however, they are defeated for reasons. One moral consideration (such as that articulated in a premise such as 'Injuries caused by negligence ought to be compensated') may be undercut or overridden by another moral consideration.

It is not clear that Rawls and Dworkin can resist the acceptability of defeasible justification, given that the process of reflective equilibrium

itself is in a sense defeasible. Equilibria, Rawls stresses, are only temporary. New circumstances can present us with issues not considered in earlier reflection or face us with unanticipated consequences. They can mobilize new considered judgments or lead us to lose confidence in judgments we had earlier affirmed under due reflection. Dworkin similarly portrays the constructive model as requiring us “to proceed on the best program [we] can now fashion” (OP 36), recognizing that we may have to revise our conclusions as further developments occur. That is true for the intuitionist as well. The intuitionist, quite reasonably, contends that if the conclusions yielded by the process of reflection are in any case defeasible, there is no reason to balk at defeasible principles and defeasible justifications within the process.

IV. AN INTUITIONIST MODEL OF REFLECTION

I have been arguing that we have no reason, in general, to expect the process of ethical reflection to reach equilibrium. Are there circumstances in which equilibrium is impossible? If so, what are they? This is one form of Williams’s broader question about contractual ethical theory: Under what conditions is it appropriate?³² In this section I distinguish conditions under which reflective equilibrium is possible from those in which it remains inaccessible. I provide an intuitionist ethical model in which equilibrium proves elusive. Given a plausible representation of the ethical domain, there are uncountably many such models, but only countably many in which equilibrium can be attained at any finite stage.

Think of ethical truths as expressed in a language—English, if you like, supplemented with whatever technical, mathematical, or other terms are required to express such truths adequately. I assume that such a language would have countably many grammatical sentences, and that the truths of ethics (or, more neutrally, the sentences that would be reached at some stage and not be overturned by further reflection) would comprise a subset E of those. Under what conditions will reflection reach equilibrium on all and only the members of E at some finite stage?

To put this question more precisely: at reflective equilibrium, principles and considered judgments harmonize. If none of the kinds of discrepancy I have discussed are to occur, our considered judgments must match the consequences of our principles exactly. (I set aside background theories and choice constraints for the sake of simplicity, but, of course, we must harmonize them as well.) Let J and P represent our judgments and principles at reflective equilibrium, and let $Cn(P)$ represent the set of logical consequences of our principles after due

³² *Ethics and the Limits of Philosophy* (Cambridge: Harvard, 1985), p. 104.

reflection. (J and P depend on our initial J_0 , P_0 , T_0 , and C_0 , as well as our revision procedure and, perhaps, choices made in the order of addressing discrepancies. In the interests of elegance I shall suppress that in the notation.) Then principles and judgments must harmonize in the sense that $\text{Cn}(P) = J$. P must axiomatize J . As a first pass, then, we might reasonably expect to reach reflective equilibrium only if J is axiomatizable. As a general formulation, however, this will not do; if we do not reach reflective equilibrium, there is no set J of judgments at equilibrium to which to refer.

So, say that a judgment p is *stable* under reflection (started from J_0 , P_0 , T_0 , and C_0) if and only if, no matter in what order discrepancies are addressed, there is some stage n of reflection such that, for any subsequent stage $m \geq n$, p belongs to J_m . Stable judgments, in other words, at some stage of reflection enter our set of considered judgments and never leave. No further reflection undermines them. We can meaningfully speak of the set J^* of stable judgments under a process of reflection whether or not the process terminates in equilibrium. In general, we can expect a process to reach equilibrium only if J^* is axiomatizable. And that will be true only if J^* is enumerable.

To craft intuitionist models of the reflection process, then, identify, in a Peircean spirit, the set E of ethical truths with J^* —the set of considered ethical judgments stable under reflection—and assume that it is not enumerable. There are two cases to consider. J^* might be *partially enumerable*: it might fail to be enumerable but have infinite subsets that are enumerable. The set of stable ethical judgments in such a case might lie beyond characterization by stouthearted rules or principles. But subsets of it might be so characterizable. Alternatively, J^* might be *unruly* in the sense that neither it nor any of its infinite subsets is enumerable. In such a case, there would be no sound stouthearted principles; every principle would be prey to counterexamples and exceptions. Given a countable language, there are only countably many enumerable candidates for E . There are uncountably many partially enumerable candidates, and uncountably many others that are unruly.³³

Consider first models in which J^* is partially enumerable. There is no decidable set P of principles such that $\text{Cn}(P) = J^*$. At no finite

³³ Assuming Church's thesis, unruliness is equivalent to J.C.E. Dekker's concept of immunity. See "Two Notes on Recursively Enumerable Sets," *Proceedings of the American Mathematical Society*, iv (1953): 495–501, on p. 496, and "Productive Sets," *Transactions of the American Mathematical Society*, lxxviii (1955): 129–49, on p. 130. See Emil Post, "Recursively Enumerable Sets of Positive Integers and Their Decision Problems," *Bulletin of the American Mathematical Society*, l (1944): 284–316, on p. 298; and Hartley Rogers, *Theory of Recursive Functions and Effective Computability* (New York: McGraw-Hill, 1967), pp. 107–09, 120–26, for cardinality and other relevant results.

stage, then, will a set of principles axiomatize the ethical truths. If we define P^* as the set of stable principles, then $P^* \neq P_n$ for any finite stage n . We never reach equilibrium. Nevertheless, if J^* is partially enumerable, reflection successively approximates an axiomatization of the ethical truths without ever attaining it. That is, it is possible to reach a stage of reflection after which all remaining discrepancies between principles and judgments are of type (2), consisting of judgments such that neither they nor their negations follow from the principles. After such a stage, stable principles, facing no further counterexamples, may safely be used as premises of deductively sound arguments. We can obtain decidable sets P of principles such that $\text{Cn}(P)$ is a subset of J^* . But, for each decidable set P of principles, we may generate stable judgments not entailed by those principles. At no stage, then, do we reach reflective equilibrium.

Models in which J^* is unruly share many of these features. Again, there is no decidable set P of principles such that $\text{Cn}(P) = J^*$. Again, $P^* \neq P_n$ for any finite stage n . We never attain equilibrium. In these models, moreover, we may not understand reflection as successive approximation. We never reach a stage after which all discrepancies are of type (2). At every stage, in other words, our principles face actual or possible counterexamples. They may not safely be used as premises in deductively sound arguments; further stable considered judgments could be adduced to contradict them. There is no decidable set P of principles such that $\text{Cn}(P)$ is even a proper subset of J^* . P has infinitely many logical consequences. So, $\text{Cn}(P)$ is both infinite and enumerable. But J^* has no infinite enumerable subsets.

Models in which J^* is unruly thus ground the idea that reasoning appropriate to ethics is defeasible. Any decidable set P of principles has consequences that fail to find affirmation in our considered judgments after due reflection. That is true even for principles taken individually. So, in such models, stouthearted principles invariably imply too much. Replacing deductive with nonmonotonic consequence by itself does not help, for the latter is supraclassical. Stouthearted principles, therefore, are never stable. We must reinterpret principles as fainthearted.

Can we always do so? By Dekker's theorem, any unruly set is Turing-equivalent to an enumerable set. That means that we could reason about unruly sets as we do enumerable sets—by deriving consequences from sets of axioms, for example—if we had an oracle enabling us to complete certain infinite tasks. We could apply a principle such as 'If A 's negligence injures B , A owes B compensation', legitimately deriving conclusions such as 'Jones owes Smith compensation', if we could complete an infinite procedure certifying the safety of the inference. The role of *ceteris paribus* clauses or the equivalent, from

this perspective, is precisely to stand in for that infinite task. It is to signal that conclusions drawn from the fainthearted principles they inhabit are not completely safe. They would be if we were able to complete an infinite procedure—namely, surveying all possible interfering factors and certifying that none undercut or overrode the inference. So, in place of stouthearted principles such as ‘If A ’s negligence injures B , A invariably owes B compensation’, we can reason with principles such as ‘If A ’s negligence injures B , then, in the absence of competing considerations, A owes B compensation’, where the set of competing considerations is potentially infinite. In short, we can reason with fainthearted principles. Dekker’s theorem guarantees an appropriate set of fainthearted principles for any unruly set.

I have been arguing that models in which J^* , the set of our considered judgments stable under reflection, is unruly provide models that are at once intuitionist and constructive, supporting the construal of ethical principles as fainthearted. In such models, we never reach reflective equilibrium, but we can nevertheless meet our obligation to make ethical decisions on the basis of principle and articulate our justifications for those decisions in terms of principles. We construe ethical truth as stability under due reflection. In such models, moreover, common law adjudication serves as a model for ethical reasoning in general. We articulate fainthearted principles that may undergo refinement or even rejection as we consider further cases. In deciding cases, we appeal to principles and commit ourselves to deciding similar cases similarly, while recognizing that potentially infinitely many factors might interfere and make other *prima facie* similar cases morally dissimilar.

All one needs to motivate fainthearted constructivism is the *possibility* of unruliness. Perhaps our stable judgments under due reflection are not unruly. Perhaps they are at least partially axiomatizable. Perhaps they are enumerable or even decidable. But there are uncountably many models in which they are unruly. Even if our stable judgments are partially axiomatizable, furthermore, there may be important subsets that are unruly. Apart from an argument that no interesting class of stable judgments is unruly, then, we must recognize unruliness as a possibility. We must consequently recognize a possibility that deductive reasoning in ethics will lead us astray. Defeasible reasoning, in contrast, succeeds whether our stable judgments are unruly or not. Intuitionism is thus safer than stouthearted constructivism. If its underlying conception of ethical truth is inaccurate, it does not get us into trouble. Stouthearted constructivism does.

V. A POSSIBLE SHORTCUT

Are there any arguments that no interesting class of stable ethical judgments is unruly? Are there arguments, in other words, that every

interesting class of stable judgments is axiomatizable, in full or in part? Rawls himself offers one such argument. I have been speaking of the set J^* of judgments stable under due reflection. But its properties, Rawls argues, are not independent of the reflection process. It has the properties we want it to have.

...the priority problem is not that of how to cope with the complexity of already given moral facts which cannot be altered. Instead, it is the problem of formulating reasonable and generally acceptable proposals for bringing about the desired agreement in judgments. On a contract doctrine the moral facts are determined by the principles which would be chosen in the original position. These principles specify which considerations are relevant from the standpoint of social justice. Since it is up to the persons in the original position to choose these principles, it is for them to decide how simple or complex they want the moral facts to be (TJ 45).

Rawls in effect argues that intuitionism presupposes moral realism. We need to worry about the complexity of the moral facts only if those facts and their properties are independent of us. For the constructivist, Rawls contends, that is not so; we determine the facts and their properties. They will be too complex to be captured by rules only if we want them to be. Our process of reflection will thus reach equilibrium if we want it to. Rawls plainly thinks we would.

Rawls's argument directly assails my project of developing a faint-hearted constructivism. If he is right, *wanting* the priority problem to have a solution guarantees that it *does* have a solution. Desires to solve the priority problem, reach reflective equilibrium, and state stouthearted principles of justice are, if not self-fulfilling, then self-grounding in the sense that they suffice for the possibility of their own satisfaction. The quest for reflective equilibrium is, as such, capable of success. It follows that the set of stable judgments and, thus, the set of ethical truths are axiomatizable, simply because we want them to be.

So understood, Rawls's argument has many ancestors in the history of antirealism. Here is one analysis of it:

- (1) Choices in the original position determine the principles of justice.
- (2) The principles of justice chosen in the original position determine the moral facts.
- (3) People in the original position would choose to make the moral facts simple.
- (4) So, moral facts are simple.

An argument of similar form might help to arouse and focus suspicion:

- (5) Choices the batter makes determine the trajectory of the bat.
- (6) The trajectory of the bat determines whether the ball goes over the fence.

- (7) Every batter would choose to hit the ball over the fence on every swing.
- (8) So, every batter hits the ball over the fence on every swing.

What has gone wrong?

Choices the batter makes determine the trajectory of the bat—*given* the trajectory of the ball, *in the absence of* interfering factors (such as catcher's interference), and *provided that* the batter succeeds in implementing his choices successfully. Just so, choices made in the original position determine principles of justice—*given* a set of considered judgments, a set of conjectured principles, a set of background theories, a plausible procedure for revision, the intuitive judgments required for implementing that procedure, and perhaps (if the procedure is iterated) the order in which candidates for revision are considered. Since the original position is idealized and hypothetical, we need not consider interfering factors and problems of implementation. In practical applications of Rawls's procedure, however, those are real concerns. It might be more accurate, therefore, to say that choices made in the original position, under ideal conditions, *contribute to determining* the principles of justice.

The trajectory of the bat determines whether the ball goes over the fence—again, given the trajectory of the pitched ball, wind conditions, and the absence of interfering factors (such as a fielder's glove). Similarly, the principles of justice chosen in the original position determine the moral facts—given considered judgments, background theories, and so on, and in the absence of interfering factors such as new experiences, exercises of imagination, or persistent intuitive judgments that resist assimilation to theory. Once again, it would be more accurate to say that principles of justice contribute to determining the moral facts.

On to the third premise: Would every batter choose to hit a home run on every swing? Given enough idealizations, perhaps. But in reality batters choose to foul balls off, bunt, hit and run, and so on. Batters who frequently swing for the fences frequently strike out. Similarly, given enough idealization, perhaps people in the original position would choose to make the moral facts simple enough to be captured by rules. But there are other considerations here as well. Recall Feinberg's observation: "Other things being equal, simplicity is preferable to complexity, but a distorting simplicity is worse than none" (*op. cit.*, p. 111).

Why would people in the original position not choose to make the moral facts simple? For one thing, they might be concerned about inflexibility. A set of fainthearted rules allows for individual discretion when rules conflict. A set of stouthearted rules settles all such conflicts in advance, permitting no exceptions. People in the original position might reasonably hope that exceptions can be made when the facts

of a particular situation make the applicability of a rule problematic. Ample illustrations of the problems associated with inflexibility stem from “zero-tolerance” policies concerning drugs and weapons in schools. Children have been suspended from school under such policies for possession of plastic kitchen utensils, palette knives, squirt guns, asthma medication, and even lemon drops. Fainthearted principles allow for and indeed rely on common sense; stouthearted principles notoriously exclude it. Fainthearted principles invite people to develop and use good judgment; stouthearted principles make judgment irrelevant, making the application of principles, and thus moral and political decisions, a purely deductive exercise. People in the original position might reasonably choose to allow for the exercise of common sense in the application of principles. If so, however, they would opt against the sort of simplicity that Rawls needs.

People in the original position might also worry about bureaucracy. Legal regulation on the stoutheartedly constructive model—in the United States and other countries attempting to devise stouthearted rules that can cover every eventuality while applying the same rules to all—has become increasingly complex. Far from settling on two straightforward principles, as Rawls suggests, the process tends toward the proliferation of more and more specific principles. Administrative and regulatory law, not to mention the tax code, have become so enormously complex and detailed that the law seems simply unknowable. At best, a few experts can become knowledgeable about a narrow area of the rules; others are forced to rely on their expertise. Those experts, even if they can master the relevant rules, will have incentives to become corrupt, exercise arbitrary power, and create “loopholes” in the rules.³⁴ Ordinary people, meanwhile, are bound to lose respect for the rules as they produce bizarre results and as people realize that, due to their intricacy, everyone is in violation of something. It is easy to imagine people in the original position preferring a few simple rules, applied with common sense, to a vast body of detailed prescriptions that permit no exercise of judgment.

Let us summarize the premises in a more plausible form, then:

- (1') Choices in the original position contribute to determining the principles of justice.
- (2') The principles of justice chosen in the original position contribute to determining the moral facts.
- (3') People in the original position would value simplicity among other things in constructing the moral facts.

³⁴ For many examples of this and related phenomena, see Philip K. Howard, *The Death of Common Sense* (New York: Random House, 1994).

These premises do not support anything like Rawls's conclusion. At best they imply that people in the original position would take simplicity into account as one factor among many to be considered in selecting principles of justice.

Rawls's argument fails for another reason as well. Batters *do not know* what swing will, in given circumstances, enable them to hit home runs. The determination of which the premises speak is not transparent. Similarly, people—even those behind the veil of ignorance, idealized in various ways—do not know which constraints adopted in the original position will yield an enumerable set of considered judgments after due reflection. Given some initial constraints, the process of reflection may terminate in equilibrium after a finite time. But it may not. Even those in the original position cannot predict their intuitive judgments at later stages of the process. Even someone who reaches equilibrium, moreover, may not be in a position to recognize it as such.³⁵ This brings us back to our original question: How can we tell whether the process of reflection will reach equilibrium?

VI. THE UNSOLVABILITY OF THE PRIORITY PROBLEM

On Rawls's conception, I shall argue, the matter is undecidable.

Here is one idea: seek coherence between our intuitive judgments and a set of fainthearted or stouthearted principles, and then examine the result to see what kind of principles it confirms. Perhaps we will settle on Rawls's principles of justice. Perhaps we will settle on competing stouthearted principles. Perhaps we will settle on fainthearted principles.

Here is another idea: seek equilibrium. If we find it, we confirm Rawls's principles or some others. Reflective equilibrium is by definition a matching of intuitions to stouthearted principles. If intuitionism is correct, the process of adjusting intuitions and stouthearted principles never leads to equilibrium, because no set of stouthearted principles can match our considered judgments, even when "duly pruned and adjusted." No stouthearted principles, no equilibrium.

These ideas are similar, but have, at first glance, different implications for settling the debate between intuitionists and constructivists. On the first conception, intuitionism and constructivism seem to be on a par. There is no guarantee that we can reach reflective equilibrium. As long as we have not reached it, the debate cannot be resolved. Upon reaching it, however, it appears that it can be; we confirm either intuitionism or some version of constructivism. On the second conception, the debate seems to lack such symmetry. If we reach reflective equilibrium, we confirm constructivism. It seems that noth-

³⁵ I owe this point to an anonymous referee.

ing we can do, however, would confirm intuitionism. If we have not reached equilibrium, how can we tell whether we might do so in the future, upon further reflection and experience?

If we could answer that question—if we could tell, in other words, whether our process of reflection can lead to equilibrium—we could presumably apply our answer now. We could settle the debate between intuitionism and constructivism without an extended process of seeking coherence. In that case, we could address the priority problem's solvability directly, without the method of reflective equilibrium. Rawls, as we have seen, sees little promise in such a direct approach, despite the argument considered in the previous section. The only way to refute the intuitionist, he maintains, is to present the rules alleged not to exist. If we set direct approaches aside, however, it seems that, apart from general inductive considerations, we could never be in a position to confirm intuitionism. We might grow tired of seeking equilibrium and despair of ever finding it. But intuitionism could never be confirmed in the way that constructivism might be. On the second conception, then, if intuitionism is correct, it cannot be confirmed.

Despite appearances, things are much the same on the first conception. The stouthearted principles the constructivist seeks are general, universal principles, applying to everyone in situations of certain kinds. The intuitionist denies the possibility of such principles. At best, our principles can be general, but not universal. Tradeoffs among principles are both inevitable and too complex to be settled by any abstract rule. We can articulate fainthearted moral principles, but they hold not universally but generally, normally, in the absence of competing considerations, all other things being equal. Now, to tell whether our reflective equilibrium confirms constructivism or intuitionism, we might examine our set of principles to see whether they are truly universal, without any such clauses. If they are, we confirm constructivism. If not, do we confirm intuitionism? The constructivist can always say that the *ceteris paribus* principles on hand are just an approximation of the truly fundamental principles, which apply universally. The intuitionist, of course, can deny it. But how do we settle *that* dispute? Reflective equilibrium on a set of stouthearted principles would again confirm constructivism. Lack of reflective equilibrium could confirm intuitionism only if we had some independent and more direct way of determining whether reflective equilibrium on a set of stouthearted principles were possible.

The situation is familiar from the theory of computation. Consider any enumerable but undecidable set. There is an effective positive test for membership in it: list the members of the set and stop when you reach the object in question. If the object belongs to the set, the

procedure terminates in a finite time. If not, however, the procedure never terminates. A yes-no question is decidable if there is an effective procedure for answering it correctly within a finite time, but only positively (or negatively) semi-decidable if the procedure answers correctly after a finite time if the answer is yes (or no) but might run on infinitely if the answer is no (or yes).

This, it seems, is what we face in the method of reflective equilibrium. If stouthearted constructivism is correct, the procedure of reflecting on our principles and considered judgments, adjusting each to the other, eventually reaches an equilibrium in which stouthearted principles and judgments cohere. If not, the procedure does not terminate. The question of the correctness of constructivism is, on Rawls's picture, positively semi-decidable. The question of the correctness of intuitionism is negatively semi-decidable. Neither question is decidable.

Actually, the situation is more complex than that formulation suggests. The process of identifying considered judgments, working out the consequences of principles, applying them to considered judgments, adjusting one or the other, formulating new principles, and so on—in short, the process of reflection itself—is neither infallible nor mechanical. It is itself a highly complex process relying on intuitive moral judgment. At best, then, we can say that the question of the correctness of constructivism is positively semi-decidable relative to the complexity of the process of reflection.

I have predicated this discussion on Rawls's idea of the dispute between his theory of justice and intuitionism. But perhaps Rawls is wrong. Can we say anything more general about the issue? Without a precise characterization of the Rawlsian revision process, it is impossible to achieve any results. But there is an obvious analogy between the question whether the process of reflection reaches equilibrium and the halting problem, the question whether a program terminates in a finite time. If that analogy could be made formally precise, it would be possible to prove that the equilibrium problem, like the halting problem, is unsolvable.

VII. REFLECTION WITHOUT EQUILIBRIUM

I have been arguing that we have no reason in general to expect Rawls's process of reflection to terminate in equilibrium. On Rawls's own terms, the equilibrium problem is unsolvable. So, therefore, is the debate between Rawls and the intuitionist, and more generally, the problem of selecting an optimal theory of justice.

But my conclusions are not all negative. Rawls's account of reflection points the way toward a revitalized, pragmatic intuitionism.

First, as I have argued, it is possible to combine intuitionism and constructivism. We can advocate all three intuitionist theses—pluralism, conflict, and complexity—while maintaining that moral value is constructed. We can retain a Rawls-inspired reflection process without insisting on stouthearted rules.

Second, even if we cannot expect equilibrium, we can see a Rawlsian process of revising principles and judgments in light of each other as providing a coherentist justification of both.³⁶ We can treat judgments stable under reflection as true. If the set of stable judgments is unruly, no stouthearted principles will be stable. But there will always be a set of fainthearted principles that can serve to justify moral judgments. Even in an unruly universe, some fainthearted principles will be stable. We can appeal to them to offer principled justifications of moral judgments. The intuitionist can thus use Rawls's methodology to justify considered judgments and fainthearted principles with greater confidence than Rawls himself can use it to justify judgments and principles.

Third, the view of ethics that emerges from fainthearted constructivism is essentially dynamic. Just as we might think of epistemology as the science of belief revision, we might think of ethics as the science of *attitude* revision. What conveys justification on our judgments and principles in a reflective procedure, if anything does, is not only the independent justification that some elements of the process may enjoy but also the rational nature of the process itself. As I have pointed out, Rawls has little to say about the revision process. Once we see it as central, however, we can elaborate its character, recognizing that it may contain uniquely ethical elements as well as elements shared by any rational revision procedure.

Finally, fainthearted constructivism permits a view of moral conflict quite different from that assumed in Rawls's discussion of the priority problem. Rawls looks at conflicts as problems that an adequate theory must solve; anything that fails to resolve them is "at most half a conception." That violates the intuitionist's sense that we perpetually operate in the face of unresolved conflicts. Solving the priority problem is not an adequacy condition for a conception of justice or ethics. It is an ongoing task that forms one of the chief enterprises of such a conception in practice. As Isaac Levi has stressed, sometimes our decisions presuppose the resolutions of conflicts, but sometimes they

³⁶ For an elaboration of such a view, see Robert Audi, "Intuitionism, Pluralism, and the Foundations of Ethics," in Sinnott-Armstrong and Timmons, eds., pp. 101–36. For a more formal treatment, see John L. Pollock, "Evaluative Cognition," *Nous*, xxxv (2001): 325–64.

constitute such resolutions, and often they do neither.³⁷ Conflicts are spurs to further reflection and inquiry. We have no reason to believe that we can eliminate them all at any finite stage of reflection. But neither do we have reason to see any given conflict as irresolvable. Our task in facing a conflict is the one that Levi and, before him, John Dewey emphasize, namely, to exercise intelligence and common sense:

There are conflicting desires and alternative apparent goods. What is needed is to find the right course of action, the right good. Hence, inquiry is exacted: observation of the detailed makeup of the situation; analysis into its diverse factors; clarification of what is obscure; discounting of the more insistent and vivid traits; tracing the consequences of the various modes of action that suggest themselves; regarding the decision reached as hypothetical and tentative until the anticipated or supposed consequences which led to its adoption have been squared with the actual consequences. This inquiry is intelligence.³⁸

The task is to devise a theory of attitude revision: to specify precisely what moral intelligence and common sense *are*.

DANIEL BONEVAC

University of Texas/Austin

³⁷ See "Conflict and Social Agency," this JOURNAL, LXXIX, 5 (May 1982): 231–47; *Hard Choices: Decision Making under Unresolved Conflict* (New York: Cambridge, 1986), and *The Covenant of Reason* (New York: Cambridge, 1997).

³⁸ *Reconstruction in Philosophy* (New York: Henry Holt, 1920), p. 173.