

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. ???–???, 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,

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 3 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. ???–???, 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

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

be enumerable or, equivalently, axiomatizable.

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," forthcoming, pp. 235–38 (au: what do these page numbers refer to?).

¹⁵ See Hansson, "Taking Belief Bases Seriously," in Dag Prawitz and Dag Westerståhl, eds., *Logic and Philosophy of Science in Uppsala* (New York: Kluwer, 1994), pp. ???–???; "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*, ?? (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, ed., *Causation in Decision, Belief Change, and Statistics II* (New York: Kluwer, 1988), pp. ???–???

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, xcvii, 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 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*, xI, 3 (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. ??).

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, has 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 constitute such resolutions, and often they do neither.³⁷ Conflicts are

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

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).

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 emphasizes, 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

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