did so indirectly (say, by swallowing a placebo); his belief in this false causal efficacy does not alter the fact that he performed the basic action.

One might weaken the foregoing claim to require only that it must be *possible* that there be a time at which I am aware that my present belief b is a successful instance of voluntary belief acquisition, if belief at will is possible at all. This requirement might be supported by a principle of reflective consciousness which held that, in regard to some mental states that arise from actions I perform in full consciousness, it must be possible for me to see that they arose from my action without this observation affecting the observed state. However, such a principle is by no means self-evident, nor is it clear how to argue for it or for the requirement.

Thus, Williams's argument fails both because (3) is not established and because the transition from there to (7) is dubious at best. The considerations advanced against this argument indicate that the strongest claim relevant to believing at will which we have reason to accept is GCL: it is impossible for me to believe of a particular belief b that b is a present belief of mine and is sustained at will. GCL is intentionally restricted. It does not entail that I cannot believe I have acquired beliefs at will, or that I cannot regard b, which I acquired at will, as a present belief of mine, or that I must stop believing b if I come to think that b is sustained at will. GCL follows from the wider principle GC, which has consequences for other cases of believing independently of truth considerations. But these principles provide no convincing reason to accept the impossibility of believing at will.

BARBARA WINTERS

University of California, San Diego

WIDE REFLECTIVE EQUILIBRIUM AND THEORY ACCEPTANCE IN ETHICS *

THERE is a widely held view that a moral theory consists of a set of moral judgments plus a set of principles that account for or generate them. This two-tiered view of moral theories has helped make the problem of theory acceptance or jus-

tification [1] in ethics intractable, unless, that is, one is willing to grant privileged epistemological status to the moral judgments (calling them "intuitions") or to the moral principles (calling them "self-evident" or otherwise a priori). Neither alternative is attractive. Nor, given this view of moral theory, do we get very far with a simple coherence view of justification. To be sure, appeal to elementary coherence (here, consistency) constraints between principles and judgments sometimes allows us to clarify our moral views or to make progress in moral argument. But there must be more to moral justification of both judgments and principles than such simple coherence considerations, especially in the face of the many plausible bases for rejecting moral judgments; e.g., the judgments may only reflect class or cultural background, self-interest, or historical accident.

I shall argue that a version of what John Rawls has called the *method of wide reflective equilibrium* [2] reveals a greater complexity in the structure of moral theories than the traditional view. Consequently, it may render theory acceptance in ethics a more tractable problem. If it does, it may permit us to recast and resolve some traditional worries about objectivity in ethics. To make this suggestion at all plausible, I shall have to defend reflective equilibrium against various charges that it is really a disguised form of moral intuitionism and therefore "subjectivist." First, however, I must explain what wide equilibrium is and show why seeking it may increase our ability to choose among competing moral conceptions.

---

[1] Since the notion of justification is broadly used in philosophy, it is worth forestalling a confusion right at the outset. The problem I address in this paper is strictly analogous to the general and abstract problem of theory acceptance or justification posed in the philosophy of science with regard to nonmoral theories. I am not directly concerned with explaining when a particular individual is justified in, or can be held accountable for, holding a particular moral belief or performing a particular action. So, too, the philosopher of science, interested in how theory acceptance depends on the relation of one theory to another, is not directly concerned to determine whether or not a given individual is justified in believing some feature of one of the theories. Just how relevant my account of theory acceptance is to the question, Is so-and-so justified in believing P or in doing A on evidence E in conditions C, (vary P,A,E)? would require a detailed examination of particular cases. I am indebted to Miles Morgan and John Rawls for discussion of this point.

[2] The distinction between narrow and wide reflective equilibria is implicit in *A Theory of Justice* (Cambridge, Mass.: Harvard, 1971), p. 49, and is explicit in "The Independence of Moral Theory," *Proceedings and Addresses of the American Philosophical Association*, XLVII (1974/75): 5–22, p. 8.

I. WIDE REFLECTIVE EQUILIBRIUM

The method of wide reflective equilibrium is an attempt to pro-
duce coherence in an ordered triple of sets of beliefs held by a
particular person, namely, (a) a set of considered moral judgments,
(b) a set of moral principles, and (c) a set of relevant background
theories. We begin by collecting the person's initial moral judg-
ments and filter them to include only those of which he is rela-
tively confident and which have been made under conditions con-
ducive to avoiding errors of judgment. For example, the person is
calm and has adequate information about cases being judged.[3] We
then propose alternative sets of moral principles that have varying
degrees of "fit" with the moral judgments. We do *not* simply settle
for the best fit of principles with judgments, however, which would
give us only a *narrow* equilibrium.[4] Instead, we advance philosoph-
ical arguments intended to bring out the relative strengths and
weaknesses of the alternative sets of principles (or competing moral
conceptions). These arguments can be construed as inferences from
some set of relevant background theories (I use the term loosely).
Assume that some particular set of arguments wins and that the
moral agent is persuaded that some set of principles is more ac-
ceptable than the others (and, perhaps, than the conception that
might have emerged in narrow equilibrium). We can imagine the

3 Though Rawls's earlier formulations of the notion [in his "Outline for a
Decision Procedure for Ethics," *Philosophical Review*, LX, 2 (April 1951): 177–
197, esp. pp. 182/3] restricts considered judgments to moral judgments about
particular cases, his later formulations drop the restriction, so they can be of
any level of generality. Cf. Rawls, "Independence of Moral Theory," p. 8.
These "ideal" conditions may have drawbacks, as Allan Garfinkel has pointed
out to me. Sometimes anger or (moral) indignation may lead to morally better
actions and judgments than "calm"; also, the formulation fails to correct for
divergence between stated beliefs and beliefs revealed in action.

4 Narrow reflective equilibrium might be construed as the moral analogue of
solving the projection problem for syntactic competence: the principles are the
moral analogue of a grammar. This analogy is not extendable to wide equilib-
rium, as I show in "Some Methods of Ethics and Linguistics," forthcoming in
Philosophical Studies. Narrow equilibrium leaves us with the traditional two-
tiered view of moral theories and is particularly ill suited to provide a basis
for a justificational argument. It does not offer a special epistemological claim
about the considered moral judgments (other than the rather weak claim that
they are filtered to avoid some obvious sources of error), nor are there con-
straints on the acceptability of moral principles beyond their good "fit" with
the initial considered judgments. If we have reason to suspect that the initial
judgments are the product of bias, historical accident, or ideology, then these
elementary coherence considerations alone give us little basis for comfort, since
they provide inadequate pressure to correct for them. Cf. Rawls, *A Theory of
Justice*, p. 49.

agent working back and forth, making adjustments to his considered judgments, his moral principles, and his background theories. In this way he arrives at an equilibrium point that consists of the ordered triple (a), (b), (c).[5]

We need to find more structure here. The background theories in (c) should show that the moral principles in (b) are more acceptable than alternative principles on grounds to some degree independent of (b)'s match with relevant considered moral judgments in (a). If they are not in this way independently supported, then there seems to be no gain over the support the principles would have had in a corresponding narrow equilibrium, where there never was any appeal to (c). Another way to raise this point is to ask how we can be sure that the moral principles that systematize the considered moral judgments are not just "accidental generalizations" of the "moral facts," analogous to accidental generalizations which we want to distinguish from real scientific laws. In science, we have evidence that we are not dealing with accidental generalizations if we can derive the purported laws from a body of interconnected theories, provided these theories reach, in a diverse and interesting way, beyond the "facts" that the principle generalizes.

This analogy suggests one way to achieve independent support for the principles in (b) and to rule out their being mere accidental generalizations of the considered judgments. We should require that the background theories in (c) be more than reformulations of the same set of considered moral judgments involved when the principles are matched to moral judgments. The background theories should have a scope reaching beyond the range of the considered moral judgments used to "test" the moral principles. Some interesting, nontrivial portions of the set of considered moral judgments that constrains the background theories and of the set that constrains the moral principles should be disjoint.

Suppose that some set of considered *moral* judgments (a') plays a role in constraining the background theories in (c). It is important to note that the acceptability of (c) may thus in part depend on some *moral* judgments, which means we are not in general assuming that (c) constitutes a reduction of the moral [in (b) and (a)]

---

[5] The fact that I describe wide equilibrium as being built up out of judgments, principles, and relevant background theories does not mean that this represents an order of epistemic priority or a natural sequence in the genesis of theories. Arthur Caplan reminded me of this point.

to the nonmoral. Then, our *independence constraint* amounts to the requirement that (a') and (a) be to some significant degree disjoint.[6] The background theories might, for example, not incorporate the same type of moral notions as are employed by the principles and those considered judgments relevant to "testing" the principles.

It will help to have an example of a wide equilibrium clearly in mind. Consider Rawls's theory of justice.[7] We are led by philosophical argument, Rawls believes, to accept the contract and its various constraints as a reasonable device for selecting between competing conceptions of justice (or right). These arguments, however, can be viewed as inferences from a number of relevant background theories, in particular, from a theory of the person, a theory of procedural justice, general social theory, and a theory of the role of morality in society (including the ideal of a well-ordered society). These *level* III theories, as I shall call them, are what persuade us to adopt the contract apparatus, with all its constraints (call it the *level* II apparatus). Principles chosen at level II are subject to two constraints: (i) they must match our considered moral judgments in (partial) reflective equilibrium; and (ii) they must yield a feasible, stable, well-ordered society. I will call *level* I the *partial* reflective equilibrium that holds between the moral principles and the relevant set of considered moral judgments. *Level* IV contains the body of social theory relevant to testing level I principles (and level III theories) for "feasibility."

The independence constraint previously defined for wide equilibrium in general applies in this way: the considered moral judgments [call them (a')] which may act to constrain level III theory acceptability must to a significant extent be disjoint from the considered moral judgments [call them (a)] which act to constrain level I partial equilibrium. I argue elsewhere that Rawls's construction appears to satisfy this independence constraint, since his central

---

[6] My formulation is not adequate as it stands, since there will even be trivial truth-functional counterexamples to it unless some specification of 'interesting' and 'nontrivial' is given, to say nothing of providing a measure for the "scope" of a theory. This is a standing problem in philosophy of science [cf. Michael Friedman's attempt to handle the related question of unifying theories in "Explanation and Scientific Understanding," this JOURNAL, LXXI, 1 (Jan. 17, 1974): 5–19, esp. 15 ff]. I will assume that this difficulty can be overcome, though doing so might require dropping the loose talk about theories. I am indebted to George Smith for helpful discussion of this point.

[7] I draw here on my "Reflective Equilibrium and Archimedean Points," forthcoming in *Canadian Journal of Philosophy*.

level III theories of the person and of the role of morality in society are probably not just recharacterizations or systematizations of level I moral judgments.[8] If I am right, then (supposing soundness of Rawls's arguments!), the detour of deriving the principles from the contract adds justificatory force to them, justification not found simply in the level I matching of principles and judgments. Notice that this advantage is exactly what would be lost if the contract and its defining conditions were "rigged" just to yield the best level I equilibrium.[9] The other side of this coin is that the level II apparatus will not be acceptable if competing theories of the person or of the role of morality in society are preferable to the theories Rawls advances. Rawls's Archimedean point is fixed only against the acceptability of particular level III theories.

This argument suggests that we abstract from the details of the Rawlsian example to find quite general features of the structure of moral theories in wide equilibrium. Alternatives to justice as fairness are likely to contain some level II device for principle selection other than the contract (say a souped-up impartial spectator). Such variation would reflect variation in the level III theories, especially the presence of alternative theories of the person or of the role of morality. Finally, developed alternatives to justice as fairness would still be likely to contain some version of the level I and level IV constraints, though the details of how these constraints function will reflect the content of component theories at the different levels.

By revealing this structural complexity, the search for wide equilibrium can benefit moral inquiry in several ways. First, philos-

[8] In particular, Rawls's level III theories rest on no considered moral judgments about rights and entitlements: all such considered judgments are segregated into level I, so that level III theories provide a foundation for our notions of rights and entitlements without themselves appealing to such notions (though they appeal to other moral notions, such as fairness and various claims about persons). Rawls seems attracted to this view of his project [cf. his "Reply to Alexander and Musgrave," *Quarterly Journal of Economics*, LXXXCII, 4 (November 1974), p. 634]. In contrast, Ronald Dworkin argues that a background right to equal respect is needed at what I call "level III." [Cf. his "The Original Position," *University of Chicago Law Review*, XL, 3 (Spring 1973): 500–533; reprinted in my anthology, *Reading Rawls* (New York: Basic Books, 1975), pp. 16–53, esp. p. 45, 50 ff.] For an argument that Dworkin is wrong to posit such a level III right, see my "Reflective Equilibrium and Archimedean Points."

[9] Rawls leaves himself open to the accusation of "rigging" when he says, e.g., "We want to define the original position so that we get the desired solution" (*A Theory of Justice*, p. 141). Critics have had a field day using this and similar remarks to show that the contract can have no justificational role.

THE JOURNAL OF PHILOSOPHY

ophers have often suggested that many apparently "moral" dis-
agreements rest on other, nonmoral disagreements. Usually these
are lumped together as the "facts" of the situation. Wide equilib-
rium may reveal a more systematic, if complex, structure to these
sources of disagreement, and, just as important, to sources of agree-
ment as well.

Second, aside from worries about universalizability and generaliz-
ability, philosophers have not helped us to understand what factors
actually do constrain the considerations people cite as reasons, or
treat as "relevant" and "important," in moral reasoning and argu-
ment. A likely suggestion is that these features of moral reasoning
depend on the *content* of underlying level III theories and level II
principle selectors, or on properties of the level I and IV constraints.
An adequate moral psychology, in other words, would have to in-
corporate features of what I am calling "wide equilibrium." Under-
standing these features of moral argument more clearly might lead
to a better grasp of what constitutes evidence for and against moral
judgments and principles. This result should not be surprising: as
in science, judgments about the plausibility and acceptability of
various claims are the complex result of the whole system of inter-
connected theories already found acceptable. My guess—I cannot
undertake to confirm it here—is that the type of coherence con-
straint that operates in the moral and nonmoral cases functions to
produce many similarities: we should find methodological conserva-
tism in both; we will find that "simplicity" judgments in both
really depend on determining how little we have to change in the
interconnected background theories already accepted (not on more
formal measures of simplicity); and we will find in both that ap-
parently "intuitive" judgments about how "interesting," "impor-
tant," and "relevant" puzzles or facts are, are really guided by
underlying theory.[10]

A third possible benefit of wide equilibrium is that level III dis-
agreements about theories may be more tractable than disagree-
ments about moral judgments and principles. Consequently, if the
moral disagreements can be traced to disagreements about theory,
greater moral agreement may result.

Some examples may perhaps make this claim more plausible.

[10] We know relatively little about these features of theory acceptance in
either domain, a fact traceable to the same empiricist and positivist legacy: too
narrow an account of the relation between theory and "data," be it laws-plus-
observation or principles-plus-judgments.

A traditional form of criticism against utilitarianism consists in deriving unacceptable moral judgments about punishment, desert, or distributive justice from a general utilitarian principle. Some utilitarians then may bite the bullet and reject reliance on these "pretheoretical" intuitions. Rawls has suggested an explanation for the class of examples involving distributive justice. He suggests that the utilitarian has imported into social contexts, where we distribute goods between persons, a principle acceptable only for distributing goods between life-stages of one person. Derek Parfit urges a different explanation: the utilitarian, perhaps supported by evidence from the philosophy of mind, uses a weaker criterion of personal identity than that presupposed by, say, Rawls's account of life plans. Accordingly, he treats interpersonal boundaries as metaphysically less deep and morally less important. The problem between the utilitarian and the contractarian thus becomes the (possibly) more manageable problem of determining the acceptability of competing theories of the person, and only one of many constraints on that task is the connection of the theory of the person to the resulting moral principles.[11]

A second example derives from a suggestion of Bernard Williams.[12] He argues that there may be a large discrepancy between the dictates of utilitarian theory in a particular case and what a person will be inclined to do given that he has been raised to have virtues (e.g., beneficence) that in general optimize his chances of doing utilitarian things. We may generalize Williams's point: suppose any moral conception can be paired with an *optimal* set of virtues, those which make their bearer most likely to do what is right according to the given conception. Moral conceptions may differ significantly in the degree to which acts produced by their optimal virtues tend to differ from acts they deem right. Level III and IV theories of moral psychology and development would be needed to

[11] Compare Derek Parfit, "Later Selves and Moral Principles," in Alan Montefiore, ed., *Philosophy and Personal Relations* (London: Routledge & Kegan Paul, 1973), pp. 149–160. See Rawls's reply in "Independence of Moral Theory," p. 17 ff. The degree to which a theory of the person constrains moral-theory acceptance, and conversely, the degree to which moral theory constrains theories of the person are discussed in my "Moral Theory and the Plasticity of Persons," forthcoming in *The Monist*, LXII, 3 (July 1979). The Marxist worry that there is no non-class-relative notion of the person substantial enough to found a moral theory on thus appears as an argument about level III theory acceptability.

[12] "Utilitarianism and Moral Self-Indulgence" in H. D. Lewis, *Contemporary British Philosophy*, IVth Series (New York: Humanities, 1976), pp. 306–321.

determine the facts here. Since we want to reduce such discrepancies (at least according to some level III and IV theories), we may have an important scale against which to compare moral conceptions.

More, and better developed, examples would be needed to show that the theory construction involved in seeking wide equilibrium increases our ability to choose rationally among competing moral conceptions. But there is a general difficulty that must be faced squarely: level III theories may, I have claimed, depend in part for their acceptability on some considered moral judgments, as in Rawls's level III theories. (If the independence constraint is satisfied, however, these are not primarily the level I considered judgments.) If the source of our disagreement about competing moral conceptions is disagreement on such level III considered judgments, then it is not clear just how much increase in tractability will result. The presence of these judgments clearly poses some disanalogy to scientific-theory acceptance. I take up this worry indirectly, by first considering the charge that reflective equilibrium is warmed-over moral intuitionism.

II. THE REVISABILITY OF CONSIDERED MORAL JUDGMENTS

A number of philosophers, quite diverse in other respects, have argued that the method of reflective equilibrium is really a form of moral intuitionism, indeed of subjective intuitionism.[13] If we take moral intuitionism in its standard forms, then the charge seems unfounded. Intuitionist theories have generally been foundationalist. Some set of moral beliefs is picked out as basic or self-warranting. Theories differ about the nature or basis of the self-warrant. Some claim self-evidence or incorrigibility, others innateness, others some form of causal reliability. A claim of causal reliability might take, for example, the form of a perceptual account which even leaves room for perceptual error. Some intuitionists want to treat principles as basic. Others begin with particular intuitions, and then attempt to find general principles that systematize the intuitions, perhaps revealing and reducing errors among them. Still, and this is the central point, the justification for accepting such moral principles is that they systematize the intuitions, which carry the epistemological privilege.[14]

[13] The charge is made by R. M. Hare ("Rawls' Theory of Justice," in *Reading Rawls*, p. 82 ff), by Peter Singer ["Sidgwick and Reflective Equilibrium," *Monist*, LVIII, 3 (July 1974): 490–517, p. 494], and by Richard Brandt (*A Theory of the Good and The Right*, Oxford, forthcoming, ch. I).

[14] My characterization of intuitionism emphasizes its foundationalism. For an account that deemphasizes its foundationalism, see M. B. E. Smith's excellent

No such foundationalism is part of wide reflective equilibrium as I have described it. Despite the care taken to filter initial judgments to avoid obvious sources of error, no special epistemological priority [15] is granted the considered moral judgments. We are missing the little story that gets told about why we should pay homage ultimately to those judgments and indirectly to the principles that systematize them. Without such a story, however, we have no foundationalism and so no standard form of moral intuitionism.

Nevertheless, it might be thought that reflective equilibrium involves an attempt to give us the *effect* of intuitionism without any fairy tales about epistemic priority. The effect is that a set of principles gets "tested" against a determinate and relatively fixed set of moral judgments. We have, as it were, foundationalism without foundations. Once the foundational claim is removed, however, we have nothing more than a person's moral opinion. It is a "considered" opinion, to be sure, but still only an opinion. Since such opinions are often the result of self-interest, self-deception, historical and cultural accident, hidden class bias, and so on, just systematizing some of them hardly seems a promising way to provide justification for them or for the principles that order them.

This objection really rests on two distinct complaints: (1) that reflective equilibrium merely systematizes some relatively determinate set of moral judgments; and (2) that the considered moral judgments are not a proper foundation for an ethical theory. I will return in section III to consider (2) in a version that abstracts from the issue of the revisability of considered judgments. Here I shall consider objection (1).

"Rawls and Intuitionism," in Kai Nielsen and Roger Shiner, eds., *New Essays on Contract Theory, Canadian Journal of Philosophy*, suppl. vol. III: 163–178. Smith agrees that Rawls's "revisionism" in wide equilibrium is contrary to the spirit of intuitionism. Still, he thinks the intuitionist can accept Rawls's method as a "check" on his own, *provided* it does not lead to strongly counterintuitive revisions. It is unclear to me how much of a "check" one has if such a proviso is imposed. Smith also argues that Rawls's method of wide equilibrium cannot yield principles, such as the principle governing the duty of beneficence. If he is right, then the method not only is not acceptable to the intuitionist who wants "definitive" answers, but also does not meet Rawls's requirements. Though he does not note the point, Smith's argument turns on features of the *contract* and *not* on features of *wide reflective equilibrium* as a method.

[15] The fact that these sources of error have been minimized does give considered judgments *some* modest degree of epistemic priority, as William Lycan has reminded me.

Wide reflection equilibrium does not merely systematize some determinate set of judgments. Rather, it permits extensive revision of these moral judgments. There is no set of judgments that is held more or less fixed as there would be on a foundationalist approach, even one without foundations. It will be useful to see just how far from the more traditional view of a moral intuition the considered moral judgment in wide reflective equilibrium has come.

The difference does not come at the stage at which we filter *initial* moral judgments to arrive at *considered* moral judgments. Sophisticated forms of intuitionism leave room for specifying optimal conditions for avoiding errors of judgment. Nor does the difference come at the stage at which we match principles to judgments, "smooth out" irregularities, and increase the power of the principles. Again, sophisticated intuitionism is willing to trade away some slight degree of unrevisability for the reassurance that errors of judgment are further reduced. It is because *narrow* reflective equilibrium allows no further opportunities for revision than these two that it is readily assimilated to the model of a sophisticated intuitionism.

But *wide* reflective equilibrium, as I have described it, allows far more drastic *theory-based* revisions of moral judgments. Consider the additional ways in which a considered moral judgment is subject to revision in wide equilibrium. Suppose the considered judgment is about what is right or wrong, just or unjust, in particular situations, or is a maxim that governs such situations. In that case, it is a judgment relevant to establishing partial reflective equilibrium with general moral principles. Consequently, we must revise it if background theories compel us to revise our general principles or if they lead us to conclude that our moral conception is not feasible. Suppose, in contrast, the considered moral judgment plays a role in determining the acceptability of a component level III theory. Then it is also revisable for several reasons. Feasibility testing of the background theory may lead us to reject it and therefore to revise the considered judgment. The judgment may be part of one background theory that is rendered implausible because of its failure to cohere with other, more plausible background theories, and so the considered judgment may have to be changed. The considered judgment may be part of a system of background theories that would lead us to accept principles, and consequently some other level I considered judgments, which we cannot accept. If we can trace the source of our difficulty back to a level III con-

sidered judgment that we can give up more easily than we can accept the new level I judgment, then we would probably revise the level III judgment.

In seeking wide reflective equilibrium, we are constantly making plausibility judgments about which of our considered moral judgments we should revise in light of theoretical considerations at all levels. No one type of considered moral judgment is held immune to revision. No doubt, we are not inclined to give up certain considered moral judgments unless an overwhelmingly better alternative moral conception is available and substantial dissatisfaction with our own conception at other points leads us to do so (the methodological conservatism I referred to earlier). It is in this way that we provide a sense to the notion of a "provisional fixed point" among our considered judgments. Since all considered judgments are revisable, the judgment "It is wrong to inflict pain gratuitously on another person" is, too. But we can also explain why it is so hard to imagine not accepting it, so hard that some treat it as a necessary moral truth. To imagine revising such a provisional fixed point we must imagine a vastly altered wide reflective equilibrium that nevertheless is much more acceptable than our own. For example, we might have to imagine persons quite unlike the persons we know.

Wide reflective equilibrium keeps us from taking considered moral judgments at face value, however much they may be treated as starting points in our theory construction.[16] Rather, they are always subjected to exhaustive review and are "tested," as are the moral principles, against a relevant body of theory. At every point, we are forced to assess their acceptability relative to theories that incorporate them and relative to alternative theories incorporating different considered moral judgments.[17]

[16] C. F. Delaney suggests quite plausibly that the greater revisability of considered moral judgments in reflective equilibrium in *A Theory of Justice* as compared to "Outline of a Decision Procedure" corresponds to the shift from a more positivist view of the relation between facts and theory to a more coherentist, Quinean view. Cf. Delaney's "Rawls on Method," in Nielsen and Shiner, *op. cit.*, pp. 153–161. The assumption of some critics that Rawls's approach to wide equilibrium is intuitionist may itself derive from their own latent positivism.

[17] One reason philosophers have thought reflective equilibrium "intuitionist" is a failure to distinguish narrow and wide equilibria. A more obvious source lies in Rawls's remark, cited by nearly everyone who makes the charge of intuitionism, that "There is a definite if limited class of facts against which conjectured principles can be checked, namely our considered judgments in reflective equilibrium" (*A Theory of Justice*, p. 51). It is tempting to read the remark

III. COHERENCE AND JUSTIFICATION

A. *No Justification without Credibility.* Consider now the claim
(2) that wide equilibrium uses inappropriate starting points for
the development of moral theory. Here the accusation of neo-intui-
tionism seems to take the opposite tack, suggesting that considered
judgments are not foundational enough. The traditional intuition-
ist seemed to have more going for him. With some pomp and cir-
cumstance, the earlier intuitionist at least outfitted his intuitions
with the regal garb of epistemic priority, even if this later turned
out to be the emperor's clothes. The modern intuitionist, the pro-
ponent of reflective equilibrium, allows his naked opinions to
streak their way into our theories without benefit of any cover
story. Richard Brandt has raised this objection in a forceful way
which avoids the mistake about revisability noted earlier.

Brandt characterizes the method of reflective equilibrium as fol-
lows. We begin with a set of initial moral judgments or intuitions.
We assign an *initial credence level* (say from 0 to 1 on a scale from
things we believe very little to things we confidently believe). We
filter out judgments with low initial credence levels to form our
set of considered judgments. Then we propose principles and at-
tempt to bring the system of principles plus judgments into equi-
librium, allowing modifications wherever they are necessary to pro-
duce the system with the highest over-all credence level.[18] But why,
asks Brandt, should we be impressed with the results of such a
process? We should not be, he argues, unless we have some way to
show that "some of the beliefs are initially *credible*—and not
merely initially believed—for some reason other than their coher-
ence" in the set of beliefs we believe the most (*op. cit.*, chapter 1).
For example, in the nonmoral case, Brandt suggests that an ini-
tially believed judgment is also an initially credible judgment when
it states (or purports to state) a fact of observation. "In the case
of normative beliefs, no reason has been offered why we should

---

as follows: "to arrive at a reflective equilibrium, treat considered judgments as
a 'definite if limited class of facts' which is to determine the shape and content
of the rest of the theory." R. M. Hare ("Rawls' Theory of Justice," in *Reading
Rawls*, p. 83) and Peter Singer (*op. cit.*, p. 493) read Rawls's remark this way.
But the remark can and should be taken to mean that "the small but definite
class" emerges *only when* reflective equilibrium is reached, and still is revisable
in the light of further theory change.

[18] Presumably, we could use fairly standard treatments of degree of belief,
rooted in probability theory, to formalize what is sketched here. This formal-
ization might give particular content to the assumption that persons are ra-
tional, imposing certain constraints on revisability and acceptability. I am
indebted to Paul Horwich for discussion of this point.

think that initial credence levels, for a person, correspond to *credi-bilities*." [19] The result is that we have no reason to think that in-creasing the credence level for the system as a whole moves us closer to moral truth rather than away from it. Coherent fictions are still fictions, and we may only be reshuffling our prejudices.[20]

If Brandt's "no credibility" complaint has force, a question I take up shortly, it has such force against wide, and not just narrow, reflective equilibrium. In my reconstruction, considered moral judg-ments *may* play an ineliminable role constraining the acceptance of background (level III) theories in wide reflective equilibrium. (In general, level III theories do not reduce the moral to the nonmoral, and level IV constraints do not select only one feasible system.) But level III considered moral judgments seem to be as open as level I considered judgments to the objection that they have only initial credence and not initial credibility. At least it would take a special argument to show why worries that initial level I considered judg-ments about justice lack initial credibility fail to carry weight against initial level III judgments about fair procedures or about which features of persons are morally central or relevant. The problem is that all such initial judgments are still "our" judg-ments.[21] The fact that wide equilibrium provides support for the principles independent from that provided by level I partial equi-librium does not imply that this support is based on considered judgments that escape the "no credibility" criticism. The criticism does not go away just because wide reflective equilibrium permits an intra-theory gain in justificatory force not provided by narrow equilibrium.

B. *Credibility and Coherence.* Much of the plausibility of the "no credibility" objection derives from the contrast between nonmoral

[19] *Loc. cit.* Brandt's discussion draws on early characterizations of justifica-tion by Nelson Goodman ["Sense and Certainty," *Philosophical Review*, LXI, 2 (April 1952): 160–167] and Israel Scheffler ["Justification and Commitment," this JOURNAL, LI, 6 (March 18, 1954): 180–190.] In Scheffler's discussion, the method is described using the notion of "initial credibility," which is not ex-plicated for us. Later in the article we are told that initial credibility is only an indication of our "initial commitment to . . . acceptance" (187). Perhaps Brandt's argument should be construed as the objection to assuming, as Scheffler is willing to do, that initial credibility and initial commitment to acceptance (Brandt's "credence level") correspond in the moral case the way they do in the nonmoral case.

[20] Hare's and Singer's complaints have a similar ring to them, once purged of the mistaken view they share about the unrevisability of considered judg-ments (see fn 17 above).

[21] Cf. Kai Nielsen, "Our Considered Judgments," *Ratio*, XIX, 1 (June 1977): 39–46.

observation reports and considered moral judgments or "intuitions." A minimal version [22] of the claim that initial credibility attaches to observation reports must do two things. It must allow for the revisability of such reports. It must also treat them as generally reliable unless we have specific reasons to think they are not. Observation reports seem to satisfy these conditions because we can tell some story, perhaps a causal story, that explains why the reports are generally reliable, though still revisable. In contrast, moral judgments are more suspect. We know that even sincerely believed moral judgments made under conditions conducive to avoiding mistakes may still be biased by self-interest, self-deception, or cultural and historical influences.[23] So, if we construe a considered moral judgment as an attempt to report a moral fact, we have no causal story to tell about reliability [24] and many reasons to suspect unreliability.

I would like to suggest three responses to this way of contrasting considered moral judgments and observation reports. First, the assumed analogy between considered moral judgments and observation reports is itself inappropriate. A considered moral judgment, even in a particular case, is in many ways far more like a "theoretical" than an "observation" statement. (I am not assuming a principled dichotomy here, at most a continuum of degree of theory-dependence). Evidence comes from the way in which we support considered moral judgments as compared to observation reports: we readily give reasons for the moral judgments, and our appeal to theoretical considerations to support them is not mainly concerned with the conditions under which the judgments are made. Further evidence for my claim would require that we carry out the programmatic suggestion made earlier: see whether we can explain the features of reason-giving by reference to features of wide equilibrium.

[22] A stronger version can be formulated. It would treat some class of observation reports as self-warranting or even incorrigible. I consider only the more plausible, weaker version above. On the strong version, the criticism of reflective equilibrium is just a foundationalist attack. On the weak version, it is an attempt to show that, foundationalism aside, coherence theories of moral justification face special problems not faced by coherence theories of nonmoral justification.

[23] The contrast is hardly complete, since observation reports may also be affected by various aspects of a person's "set."

[24] Gilbert Harman makes a similar point when he claims that p's obtaining plays no role in explaining my making the moral judgment that p, but q's obtaining does play a role in explaining my nonmoral observation that q. Cf. *The Nature of Morality* (New York: Oxford, 1977), p. 7 ff. I think Harman overdraws the contrast here, but that is a matter for another discussion.

On the other hand, some may cite other evidence to support the analogy between observations and moral judgments. They might point, for example, to language-learning contexts, in which children are taught to identify actions as wrong or unjust much as they are taught to identify nonmoral properties. Or they may point to the fact that we often judge certain acts as right or wrong with great *immediacy*—the "gut reaction," so called. But such evidence is not persuasive. One thing that distinguishes adult from childish moral reasoning is the ready appeal to theoretical considerations.[25] Similarly, we are often impatient with the person who refuses to provide moral reasons or theory to support his immediate moral judgments, much more so than we are with the person who backs up "It is red" with nothing more than "It sure looks red."

Consequently, I conclude, though I have not fully argued the point here, that the comparison of moral judgments to observation reports is misleading. Rightness and wrongness, or justice and injustice, are unlikely to be simple properties of moral situations. Consequently, they are unlikely to play a role analogous to that played by observational properties in the causal-reliability stories we tell ourselves concerning observation reports. But the "no credibility" argument gains its plausibility from the assumption that the analogy to observation reports *should* hold and then denigrates moral judgments when it is pointed out they differ from observation reports. If they *should* and *do* function differently—because they are different kinds of judgments—that is not something we should hold against the moral judgments.

Secondly, the "no credibility" criticism is at best premature. It is plausible to think that only the development of acceptable moral theory in wide reflective equilibrium will enable us to determine what kind of "fact," if any, is involved in a considered moral judgment. In the context of such a theory, and with an answer to our puzzlement about the kind of fact (if any) a moral fact is, we might be able to provide a story about the reliability of initial considered judgments. Indeed, it seems reasonable to impose this burden on the theory that emerges in wide reflective equilibrium. It should

[25] Even the access to theoretical considerations generally found in mature and sophisticated adult moral reasoning will not, of course, be as extensive and developed as what I suggest is involved in wide equilibrium, despite my earlier remark that much more structure may be present than we have recognized. Moreover, how much we expect theory to play a role in adult moral reasoning will depend on our purpose in seeking the particular moral justification. These factors affect the degree to which a coherence theory of justification based on wide equilibrium carries over into a theory of individual justification; see fn 1 above.

help us answer this sort of question. If we can provide a reasonable answer, then we may have a way of distinguishing initially credible from merely initially believed types of moral judgments.

The "no credibility" criticism gains initial plausibility because we *are* able to assign initial credibility to nonmoral observation reports, but not to moral judgments. The credibility assignment, however, draws implicitly on a broadly accepted body of theory which explains why those judgments are credible. Properly understood, the credibility story about nonmoral observation reports is itself only the product of a nonmoral wide reflective equilibrium of relatively recent vintage. In contrast, we lack that level of theory development in the moral case. What follows from this difference is that the "no credibility" argument succeeds in assigning a burden of proof. *Some* answer to the question about the reliability of moral judgments must be forthcoming. But the argument is hardly a demonstration that no plausible story is possible.

Thirdly, a more positive—though still speculative—point can be made in favor of starting from considered moral judgments in our theory construction. It is commonplace, and true, to note that there is variation and disagreement about considered moral judgments among persons and cultures. It is also commonplace, and true, to note that there is much uniformity and agreement on considered moral judgments among persons and cultures. Philosophers of all persuasions cite one or the other commonplace as convenience in argument dictates. But moral philosophy should help us to *explain both* facts.

What wide equilibrium shows us about the structure of moral theories may help us explain the extensive agreement we do find. Such agreement on judgments may reflect an underlying agreement on features of the component background theories. Indeed, people may be more in agreement about the nature of persons, the role of morality in society, and so on, than is often assumed. Of course, these other points of agreement might be discounted by pointing to the influence of culture or ideology in shaping level iii theories. But it may also be that the agreement is found because some of the background theories are, roughly speaking, true—at least with regard to certain important features. Moreover, widely different people may have come to learn these truths despite their culturally different experiences. The point is that moral *agreement*—at levels iii and i—may not be just the result of historical accident, at least not in the way that some moral *disagreements* are. Consequently, it would be shortsighted to deny credibility to considered judgments just because there is widespread disagreement on many of

them: there is also agreement on many. Here moral anthropology *is* relevant to answering questions in moral theory.

I conclude that the "no credibility" objection reduces either to a burden-of-proof argument, which is plausible but hardly conclusive, or to a general foundationalist objection to coherence accounts of theory acceptance (or justification). It becomes a burden-of-proof argument as soon as one notices that the credibility we assign to observation reports is itself based on an inference from a nonmoral reflective equilibrium. We do not yet have such an account of credibility for the moral case, but we also have no good reason to think it impossible or improbable that we can develop such an account once we know more about moral theory. On the other hand, the "no credibility" argument becomes a foundationalist objection if it is insisted that observation reports are credible independently of such coherence stories.

My reply to the "no credibility" criticism points again to a strong similarity in the way coherence constraints on theory acceptance (or justification) operate in the two domains, despite the disanalogy between observation reports and considered moral judgments. The accounts of initial credibility we accept for observation reports (say, some causal story about reliable detection) are based on inferences from various component sciences constrained by coherence considerations. Observation reports are neither self-warranting nor unrevisable, and our willingness to grant them initial credibility depends on our acceptance of various other relevant theories and beliefs. Such an account is also owed for some set of moral judgments, but it too will derive from component theories in wide equilibrium. Similarly, in rejecting the view that wide equilibrium merely systematizes a determinate set of moral judgments, and arguing instead for the revisability of these inputs, I suggest that wide equilibrium closely resembles scientific practice. Neither in science nor in ethics do we merely "test" our theories against a predetermined, relatively fixed body of data. Rather, we continually reassess and reevaluate both the plausibility and the relevance of these data against theories we are inclined to accept. The possibility thus arises that these pressures for revision will free considered moral judgments from their vulnerability to many of the *specific* objections about bias and unreliability usually directed against them.

IV. OBJECTIVITY AND CONVERGENCE

I would like to consider what implications, if any, the method of wide equilibrium may have for some traditional worries about objectivity in ethics. Of course, objectivity is a multiply ambiguous

notion. Still, two senses stand out as central. First, in a given area of inquiry, claims are thought to be objective if there is some significant degree of intersubjective agreement on them. Second, claims are also said to be objective if they express truths relevant to the area of inquiry. Other important senses of "objectivity" reduce to one or both of the central uses [e.g., "free from bias" (said of methods or claims) and "reliability" or "replicability" (said of methods or procedures of inquiry)]. The two central senses are not unrelated. The typical realist, for example, hopes that methods or procedures of inquiry that tend to produce intersubjective agreement do so because they are methods that give us access to relevant truths. In contrast, there are also eliminative approaches which try to show that one or the other notion of objectivity is either confused, reducible to the other, or irrelevant in a given area of inquiry. Thus some have suggested that knowledge of moral truths is unattainable (perhaps because there are no moral truths) and we should settle for the objectivity of intersubjective agreement (based on rational inquiry) if we can achieve it. Does the method of wide reflective equilibrium commit us to one or another of these approaches to objectivity in ethics?

One traditional worry, that moral judgments are not objective because there is insufficient agreement about them, may be laid to rest by seeking wide equilibrium. I have suggested that seeking wide equilibrium may render problems of theory acceptance in ethics more tractable and may thus produce greater moral agreement. Specifically, it may lead us to understand better the sources of moral agreement and disagreement and the constraints on what we count as relevant and important to the revision of moral judgments. It may allow us to reduce moral disagreements (about principles or judgments) to more resoluble disagreements in the relevant background (level III and IV) theories. None of these possibilities guarantees increased agreement. How much convergence results remains an empirical question. But I think I have made it at least plausible that wide equilibrium could increase agreement and do so in a *nonarbitrary* way. At least, it could provide us with a clearer picture of how much agreement we already have (I return to this point later). And if it does, then there are implications for how objective, at least in the minimal sense of intersubjective agreement, ethics is.

To be sure, many who point to the lack of intersubjective agreement on many moral issues do so to raise a more robust worry about lack of objectivity in ethics. They point to moral disagree-

ment as if it were strong *evidence* for the deeper claim about objectivity, that there are no moral truths for us to agree about. The inference from lack of agreement to the absence of truths to be acquired is generally unpersuasive, however. Sometimes there is the buried assumption that *if* there were such truths, we would probably have enough access to them to produce more agreement than we have. I see no way, however, to formulate this assumption so that it does not rule out the existence of truths in most areas of scientific inquiry, at least at some time in their history. Sometimes there is the qualification that it is not the disagreement about moral claims that is important, but the "fact" that we cannot agree about what would produce resolution of the disagreement. This is likely to be more true in science and less true in ethics than is usually claimed. Still, there is a kernel of truth behind the inference, though it is insufficient to warrant it: agreement, *when it is produced by methods we deem appropriate in a given area of inquiry,* does appear to have some evidential relation to what is agreed on.

What has troubled critics of reflective equilibrium, however, is an opposite worry. Anyone who believes that there *are* objective moral truths will want to leave room for the possibility that there may be consensus on moral falsehoods. The worry is clearly reasonable when we suspect that the factors that led to consensus have little, if anything, to do with rational inquiry (and we need not have in mind anything so drastic as the Inquisition). And if one thought the method of wide equilibrium fell far short of rational inquiry, the worry would again be reasonable. Moreover, it is not obviously unreasonable even if one takes wide equilibrium to be the best method available but wants to acknowledge the possibility that it may lead to justified acceptance of moral falsehoods. The fear here is that intersubjective agreement will be taken as *constitutive* of moral truth or as eliminative of any full-blown (realist) notion of objective moral truth.[26]

The worry might be put this way. Suppose that when diverse

---

[26] Peter Singer argues that the proponent of reflective equilibrium leaves no room for a notion of the "validity" of moral principles that goes beyond intersubjective agreement. Consequently, the "validity" of moral principles will have to be relativized: it depends on whose considered judgments they are tested against (cf. *op. cit.,* p. 439 ff). I do not see why he thinks so, except that it may be connected to his underestimate of the revisability of considered moral judgments (cf. sec. II above). In any case, Rawls is quite right to deny any straightforward connection between convergence in wide equilibrium and the knowledge of objective moral truths. Cf. Rawls, "Independence of Moral Theory," p. 9.

people are induced to seek the principles they would accept in wide reflective equilibrium, only one shared equilibrium point emerges. Can we still ask, Are these principles objective moral truths? Is the proponent of wide equilibrium committed to the view that such intersubjective agreement *constitutes* the principles and judgments as moral truths? Or is it at best *evidence* that we have discovered objective moral truths? Or is it any evidence at all that we have found some? I shall suggest that though convergence in wide equilibrium is neither a necessary nor a sufficient condition for claiming we have found objective moral truths, such convergence may constitute *evidence* we have found some.

To see that convergence in wide equilibrium is not a sufficient condition for claiming we have found objective moral truths, suppose we actually produced such convergence among diverse persons. Whether or not the principles and judgments they accept would count as such truths would depend on *how* we come to explain the convergence. Suppose, for example, we find that we can *explain* the convergence by pointing to a psychological feature of human beings that plays a *causal* role in producing their agreement. Suppose, to be specific, that, under widespread conditions of child-rearing in diverse cultures, people tend to group others into "in groups" and "out groups" and that the effect of this mechanism is that moral judgments and principles in wide equilibrium turn out to be inegalitarian in certain ways.[27] Suppose we discover, further, that these child-rearing practices are themselves changeable and not the product of any deep features of human biology and psychology. We might begin to feel that the convergence we had found in wide equilibrium was only a fortuitous result of a provincial feature of human social psychology. Convergence would thus not by itself be sufficient grounds for constituting the principles as moral truths.

We can turn the example around to question the necessity of convergence for constituting the principles as objective moral truths. Suppose we find, after attempting to produce wide equilibria among diverse persons, that there is no actual convergence in wide equilibrium. Different families of equilibria emerge. Suppose also that we can *explain* the failure of convergence by pointing again to a provincial feature of human psychology or biology. But suppose further that we can abstract from this source of divergence. We can construct a modified and *idealized* "agreement" on principles. Such an idealization might, depending on other factors, be

27 I think the account is implausible, but cf. Gordon W. Allport, *The Nature of Prejudice* (New York: Addison-Wesley, 1954), chs. III, IV.

a good candidate for containing objective moral truths, even though it is *not* accepted in any actual wide equilibrium.

Which way we should go in either of these cases will have something to do with how fundamental we think the source of divergence or convergence is. But what we count as "fundamental" is itself determined by the view of the nature of moral judgments and principles which emerges in wide equilibrium. For example, if the convergence-producing feature of human psychology turned out to be a central fact about the emotions or motivations, say, some fact about the nature of (Humean) sympathy, which proved invariant to all but the strangest (pathological) child-rearing practices, then we might think we had reached a fundamental fact (related at least to the *feasibility* of moral conceptions). Still, even here, I do not want to assume that metaethical considerations embedded in the background theories would force us to reject a more Kantian stance. To follow up our earlier discussion, we are here concerned with factors that may affect the "credibility" of initial considered judgments, leading us to discount some and favor others; how we weigh these factors will depend on complex features of our background theories.

In short, divergence among wide reflective equilibria does not imply that there are no such things as objective moral truths; [28] nor does convergence imply that we have found them; nor need 'moral truth' be replaced by 'adopted in wide equilibrium'. How we will be motivated, or warranted, in treating the facts of divergence or convergence depends on the kinds of divergence or convergence we encounter and the kinds of explanation we can give for it. This result should not surprise us: wide reflective equilibrium embodies coherence constraints on theory acceptance or justification, not on truth.[29]

Actually, it is necessary to qualify my conclusion that wide reflective equilibrium need not be viewed as constitutive of moral truth.[30] My argument that convergence is neither necessary nor

---

[28] As Singer seems to think it does; cf. *op. cit.*, p. 494/5; but see also his n. 5, p. 494.

[29] If we construe wide reflective equilibrium as providing us with the basis for a full-blown coherence theory of moral justification, then my argument suggests that it faces the same difficulties and advantages as coherence theories of nonmoral justification. I cannot here defend my view that a coherence theory of justification can be made compatible with a noncoherency account of truth.

[30] There is some evidence that Rawls is attracted to a view resembling the eliminitive view in his portrayal of wide equilibrium as a "constructive" method. Cf. "The Independence of Moral Theory," *op. cit.*, and also a recent unpublished lecture with the same title as the Presidential Address.

sufficient to establish the discovery of moral truths depends on bringing theoretical considerations to bear which seem sufficient to destabilize the actual equilibrium in some way. Suppose we now throw back into the ring these destabilizing considerations and seek a new wide equilibrium. If we can soup up wide equilibrium in this way, so that it adds up to something like "total rational considerations," then perhaps we can revive in a strengthened form the constitutive view. We would have here, perhaps, the analogue of Putnam's "empirical realist" rejection of objective (metaphysical) moral truths.[31] This version of the eliminative view is not open to the most reasonable worries of those who feel simple moral agreement should not be taken to constitute moral truth. In any case, on its form of verificationism, ethics may be no worse off than science!

A more modest way of putting the same objection is this. My reply to the eliminative view is compatible with the following claim: there is a sense in which the question, Do we really have moral truth, given convergence in wide reflective equilibrium? is an *idle* worry in the absence of any *specific* research capable of destabilizing the equilibrium. In the absence of some particular, plausible way to challenge the convergence, the question is tantamount to strong and unfruitful skepticism.

Despite these qualifications, my inclination is not to treat wide reflective equilibrium as constitutive of moral truth (assuming convergence) and to leave room instead for a weaker *evidential* relation holding between agreement in wide equilibrium and moral truth. What we would need to support this possibility is reason to think that the methods of inquiry in ethics that tend to produce convergence do so *because* they bring us close to moral truth. I can offer only a highly qualified and indirect argument to this conclusion.

Consider for a moment a general argument of this form: [32] (1) In a given area of inquiry, the methods used are successful in the sense that they produce convergence and a growth of knowledge; (2) the only plausible account of the success of these methods is that they lead us to better and better approximations to truths of the kind relevant to the inquiry; (3) therefore, we should adopt a realist account of the relevant objects of inquiry. Arguments of this form

---

[31] Cf. Hilary Putnam, *Meaning and the Moral Sciences* (London: Routledge & Kegan Paul, 1978), part IV.

[32] I am grateful to Richard Boyd and George Smith for discussion of this argument and of section IV in general.

have been advanced to defend platonism with regard to mathematical objects and realism with regard to the referents of theoretical terms in the empirical sciences. To establish the second premise of such an argument, one must not only show that alternatives to the realist account (say intuitionist accounts in mathematics and verificationist or positivist accounts in science) will not explain the success of the methods used, but that the realist account has some independent plausibility of its own. Otherwise, it may simply seem to be a residual, *ad hoc* account. In mathematics, proponents of platonism, whatever the merits of their refutations of other accounts, have not provided accounts, aside from perceptual metaphors, which make it plausible that we can come to know anything about mathematical objects. In contrast, however, there are some interesting and promising arguments of this form in defense of scientific realism.[33] In these, a version of a causal theory of knowledge and reference is used to satisfy the requirement that we lend plausibility to the realist account of methodology independent of the refutation of alternative accounts.

Suppose a version of such an argument for scientific realism is sound—a supposition I shall not defend here at all. Then we would be justified in claiming that certain central methodological features of science, including its coherence and other theory-laden constraints on theory acceptance (e.g., parsimony, simplicity, etc.), are consensus-producing *because* they are *evidential* and lead us to better approximations to the truth. I have been defending the view that coherence constraints in wide equilibrium function very much like those in science. If I am right, this suggests that we may be able to piggy-back a claim about objectivity in ethics onto the analogous claim we are assuming can be made for science. Suppose then that coherence constraints in wide equilibrium turn out to be consensus-producing. Then, since these constraints are similar to their analogues in science in other respects, they may also be *evi-*

[33] The most persuasive versions of the argument are found in Boyd, "Determinism, Laws and Predictability in Principle," *Philosophy of Science*, xxxix, 4 (December 1972): 431–450; "Realism, Underdetermination and a Causal Theory of Evidence," *Noûs*, vii, 1 (1973): 1–12; *Realism and Scientific Epistemology* (New York: Cambridge, forthcoming); "What Physicalism Does Not Entail," in Ned Block, ed., *Readings in Philosophy of Psychology* (forthcoming from Harvard University Press); and "Metaphor and Theory Change: What Is 'Metaphor' a Metaphor for?" (unpublished). See also H. C. Byerly and V. A. Luzara, "Realist Foundations of Measurement," *Philosophy of Science*, xl, 1 (1973): 1–27; and the classic K. MacCorquodale and P. E. Meehl, "On a Distinction between Hypothetical Constructs and Intervening Variables," *Psychological Review*, clv, (1948): 95–107.

dential. That is, we have some reason to think that wide equilib-
rium involves methods that will lead us to objective moral truths
if there are any. Notice that this conclusion does not presuppose
there are such moral truths, nor does it give an account of what
kind of truth such a truth would be.

My suggestion is obviously a highly tentative and programmatic
route to an account of objectivity in ethics. Nor can I really defend
it here. Some qualifying remarks are definitely in order, however.

(A) Developed versions of the arguments for scientific realism do
not simply talk about "convergence," but point to a variety of
effects indicative of the cumulative nature or progress of scientific
knowledge. For example, they may try to account for "take-off"
effects indicative of the maturation of an area of inquiry, or they
may point to the absence of "schools" or "sects." My supposition
that convergence may emerge in wide equilibrium falls far short
of specifying this sort of evidence for growth in moral knowledge.
There is a related point: I am not sure we know what to count as
evidence for convergence in ethics. For example, we do have moral
disagreement on numerous issues; but is the level of disagreement
compatible with enough other agreement for it to count as con-
vergence, or not? Does existing disagreement merely represent hard
or novel problems at the "frontier"? Or is it the result of special
social forces which systematically distort our views in areas of po-
litical or religious sensitivity? Some of the difficulty may stem from
paucity of work in the history of ethics and in moral anthropol-
ogy adequate to informing us whether we have experienced moral
progress.

(B) The piggy-back argument seems to rest on the assumption
that, if a feature of method (a coherence constraint) is similar in
one respect (it produces consensus) in two areas of inquiry, then it
holds in both areas for the same reason (it leads to relevant truths).
I do not think the assumption is obviously or even generally true;
that is why my suggestion is only programmatic.

(C) The arguments for scientific realism depend on some causal
account of knowledge—e.g., perceptual knowledge depends on reli-
able detection mechanisms. We are reminded, therefore, of the
burden of proof assumed in section III to provide *some* reliability
account of moral judgments (at some level). Suppose we could pro-
vide no analogue in the moral case to the causal story we may be
persuaded of for perceptual knowledge. If we still wanted to talk
about "objective moral truths," we might *retreat* to the view that
the objects of moral knowledge were "abstract," that is, more like

mathematical objects than the things we can know about through the natural sciences. But our moral realism, then, is open to the worry I earlier expressed about mathematical platonism.[34] To be sure, if our causal accounts of knowledge turn out to be unpersuasive, then the argument for scientific realism may be no better off than this in any case.

(D) My account of wide reflective equilibrium has not provided (not explicitly at least) an obvious analogue to the role of experimentation in science. Some story about moral *practice* and what we can learn from it, and not just about moral thought experiments, seems to be needed. That is, we would need to examine the sense in which moral theories guide moral practice and result in social experimentation. But this account must be left for another project.[35]

A final remark is directed not just at my suggestion about the implications of wide equilibrium for objectivity in ethics, but at my account of wide equilibrium itself. The account I have sketched defines a wide equilibrium for a given individual at a given time. The "convergence" I have been discussing is the (at least approximate) sharing of the same wide equilibrium by different persons; the ordered triples of sets of beliefs are the same for these persons. But there would seem to be another approach.

Suppose we begin by admitting into the set of initial considered moral judgments only those judgments on which there is substantial consensus.[36] There seem to be two immediate advantages. First, ethics looks more like science in that the initial considered moral judgments share with observation reports the fact that there is substantial initial agreement on them. The starting point is more "objective," at least in the sense of intersubjective agreement. One *may* gain a slight edge in respect to the problem of initial credibility discussed earlier. (Revisability is, nevertheless, presumed.) Second, the approach makes the wide equilibrium that emerges (if one does)

---

[34] For a similar suggestion, see Jane English, "Ethics and Science," *Proceedings of the XVI Congress of Philosophy*, forthcoming. For some remarks on the analogy between choosing among alternative logics and choosing among moral theories, see my "Some Methods of Ethics and Linguistics," *Philosophical Studies*, forthcoming.

[35] There are some interesting suggestions along these lines in Ruth Anna Putnam, "Rights of Persons and the Liberal Tradition," in Ted Honderich, *Social Ends and Political Means* (London: Routledge & Kegan Paul, 1976).

[36] Just before her tragic death, Jane English reminded me of the importance of this alternative, in comments on an earlier draft of this paper. She argues that ethics should be constructed on such a basis in a brilliant short paper, cited above.

much more a collective or social product from the start than does my approach, which is a quite unnatural idealization in this regard.

Though I think this alternative merits further examination, which I cannot undertake here, I am not persuaded that it offers real advantages. For one thing, it builds into its procedure the assumption that considered judgments *ought* to function like observation reports in science, a question, I have argued, there is good reason to leave open. Its apparent advantage in making ultimate convergence seem more likely might, consequently, be based on the assumption that we ought to have *initial* convergence where there is no good reason to expect it (given all the things that make *initial* considered moral judgments *un*reliable). For another thing, I have assumed that extensive consideration of alternative background theories and sets of principles will produce reasonable pressures to revise and eliminate divergent considered judgments that there are good reasons to eliminate. The alternative method may shift, in too crude a fashion (losing too many possibilities), the intermediate conclusions of my procedure into the position of methodologically warranted starting points. A less important consideration is historical: reflective equilibrium is advanced by Rawls as a model for the process of justification in ethics. Part of what he wanted to capture is a model for how we may make progress in moral argument—where we have to accommodate initial disagreement on some moral judgments. My approach retains this attractive feature, though it sheds some of the other motivations for Rawls's version.[37]

My remarks on objectivity are admittedly quite speculative; indeed, I think it a virtue of the method of reflective equilibrium that it leaves open metaethical considerations of this kind. Still, I think enough has been said about wide equilibrium, these speculations aside, to make its implications for theory acceptance in ethics worthy of closer study.

<div style="text-align: right">NORMAN DANIELS</div>

Tufts University

NOTES AND NEWS

The editors report with sorrow the death, on March 14, 1979, of Charles Leslie Stevenson, Emeritus Professor of Philosophy at the University of Michigan and Professor of Philosophy at Bennington College. Professor Stevenson had taught at Yale from 1939 to 1946, and at Michigan from then until 1977. He died, suddenly, at Bennington, at the age of 71.

[37] In particular, the analogy to descriptive syntactics (see fn 5).