

# Some Uses of Reflective Equilibrium\*

Elijah Millgram

Department of Philosophy  
University of Utah  
Salt Lake City UT 84112  
elijah.millgram@gmail.com

October 25, 2014

## 1

Over the past several decades, there has been a great deal of interest in the question of what it takes to get someone to give up an entrenched position in favor of another deeply opposed view—or even to become persuaded that it is necessary to formulate a deeply different alternative to one’s own view.<sup>1</sup> Menachem Fisch and Yitzhak Benbaji—hereafter, F&B—have recently advanced a proposal with roughly the following drift.<sup>2</sup> A systematically worked out theoretical or ethical position is normally quite difficult to dislodge on the basis of considerations that someone who occupies it will acknowledge as legitimate. However, when a differently-minded critic is sufficiently trusted, one can hear his criticisms as though they were one’s own.<sup>3</sup> Once one’s posi-

---

\*I’m grateful to Carla Bagnoli, Margaret Bowman, Svantje Guinebert, Valeria Ottonelli, Hilary Putnam, John Rawls and Aubrey Spivey for discussing the topics of this paper, to Chrisoula Andreou, Jon Bendor, Ben Crowe, Christoph Fehige and Cindy Stark for comments on earlier drafts, to the Guggenheim Foundation for support, and to the University of Washington’s Helen Riaboff Whiteley Center and the Bogliasco Foundation’s Liguria Study Center for time and space to write.

<sup>1</sup>That interest has been particularly pointed within analytic philosophy of science, where so-called paradigm shifts have captured the imaginations of scholars, but the problem is evidently more general; for instance, Alasdair MacIntyre has written extensively and thoughtfully about how one philosophical and moral tradition can emerge rightly victorious in competition with another. See MacIntyre, 1988, and MacIntyre, 1990.

<sup>2</sup>Fisch and Benbaji, 2011; freestanding page references are to this book.

<sup>3</sup>There is a closely related effect: one can come to appreciate how one’s posture looks from the outside (pp. 248–251, 290). The benefits are roughly those provided by friends

tion has been destabilized by internalizing a conflicting take on things, a new and robust view can emerge—one that presumably preserves whatever was worth saving in its predecessors. F&B’s reconstruction of the problem they are considering sensitively recapitulates and synthesizes the moves made by some of the most important previous contributors to the back and forth. And I am sympathetic enough to their suggestion to hope it is given the best possible run for its money.<sup>4</sup>

Here I want to take the occasion provided by F&B’s proposal to reconsider a notion that has become a staple of contemporary analytic philosophy. F&B’s turn to outside criticism is unlikely to be regarded as a necessary step by much of the analytic mainstream; the default view today is that the method of *reflective equilibrium* properly drives revisions to one’s theoretical position. As standardly construed, that method takes as inputs what you *already* think, no alien interventions required. What is more, the method is imagined to be a generic philosophical solvent, that is, it is the way we can approach pretty much any philosophical problem. (To be sure, that still gives it a narrower scope than the problem F&B are taking up, where the subjects one is meant to be reconsidering are not only philosophical.) Finally, F&B’s concern is that large but needed adjustments to one’s view will not happen unless the view is dislodged from outside.<sup>5</sup> However, the method of reflective equilibrium is alleged by its proponents not to be particularly conservative, and so to be able to manage those larger revisions.<sup>6</sup>

Now, F&B are clear throughout that their decision as to where to focus their efforts is in the first place a tactical choice; whatever the limits of

---

construed as “human mirrors”—a plausible partial explanation of the advantages of friendship explored by Aristotle. (This version of Aristotle’s theory is reconstructed in Cooper, 1980; I discuss the reasons it was not Aristotle’s final and considered view of the topic in Millgram, 1987.)

<sup>4</sup>The sympathy is due in part by the similarities between F&B’s view and the account in Millgram, 1997, ch. 7, of the role of trust in friendship.

<sup>5</sup>For instance, modifying a package of normative commitments on the basis of “testing for consistency, coherence, or clarity” is described as “uninteresting” (18) and “trivial” (15); “troubleshooting for coherence, fit, and clarity” does no more than “make explicit and render consistent what is essentially already there” (117f). Cf. also pp. 120, 164, 167, 194.

<sup>6</sup>See, for instance, Scanlon, 2003, pp. 150, 157, Daniels, 2013, sec. 3.1, Daniels, 1996, pp. 14, 27, 84, or again, DePaul, 1993, pp. 42f, and Copp, 1984, pp. 151, 156—although these last two are not proponents of the method. For recent discussion, see Kelly and McGrath, 2010. To be sure, those claims came in response to the complaint that the method *is* reprehensibly conservative; see note 10, below.

Chase and Reynolds, 2010, pp. 78, 84, plausibly suggest that the move from narrow to wide reflective equilibrium, described below, was meant to forestall that complaint.

internal criticism ultimately prove to be, it is worthwhile to ask what one can turn to if it's just not enough. I will start out by adding support for their decision about where to stand and fight; appeals to reflective equilibrium, I will argue, are not going to do the work the mainstream supposes. (The problem I will focus on here is not the method's disputed conservatism; because it is almost impossible not to address that question, while it will be a motif of the coming discussion, it is meant to be only a subsidiary one.) To explain why, I will explain the method, and then survey a handful of applications for which it is well-motivated; in the course of doing so, I will remind the reader of its inventor's little-noticed view as to what an argument for the method would look like. That will put me in a position to advance the following claim: that we must proceed on the assumption that the method of reflective equilibrium is usable only for a narrow range of applications. If that is correct, then, anyway for the foreseeable future, the method of reflective equilibrium is only locally available; whether it is overly conservative or not, we cannot count on pressure toward reflective equilibrium to drive theoretical change across the board in the philosophy; *a fortiori*, we cannot count on it elsewhere.<sup>7</sup> If all that is correct, then F&B

---

<sup>7</sup>Arguing that reflective equilibrium is only locally available means prescinding from identifying the method with internal criticism of any kind whatsoever. That means I'll be leaving it an open question just how much internal criticism can achieve.

It does seem to me that F&B's skeptical take on that question may be more extreme than is really warranted. Nietzsche, you will remember, thought that *any* value ends up overcoming itself (e.g., once your training in intellectual honesty really takes, you end up questioning the value of intellectual honesty); while that is perhaps itself overstated, surely—and *pace* F&B's claim that "it is impossible to argue normatively against the norms to which one is committed" (27)—such self-overcoming happens more than occasionally.

Moreover, well-worked out, coherent and systematic intellectual or ethical positions are, in my experience, rare; shoddiness is much more typical, and when one presses on most available positions—yes, from the inside—they come apart in one's hands. Lest this sound like someone who thinks he can see through the poorly constructed views of others, let me hasten to assure you that this is my experience with my own theoretical and evaluative positions; I do not think myself somehow an exception to the general rule. Philosophy started out when Socrates began checking whether that *was* the general rule, and it is hard to believe that, if we had someone to take up his mantle, we would not be seeing more of the very same results obtained by Socrates himself.

Why do F&B have the impression that such positions cannot be dislodged from the inside? They are certainly hard to dislodge in practice, but that has to do with a mix of poor training in their adherents and a wilful refusal to consider the flaws in one's view: people *balk*. (As Nehamas, 1998, p. 41, once pointed out, Euthyphro is not convinced, he remembers that he is late for an appointment.) So I suspect that F&B are charitably but erroneously reifying the personal failings of most positions' *adherents* into presumed virtues of their *theories*, e.g., the systematicity and robustness of the positions themselves.

are right to explore further expedients.

Then I will turn to F&B's understanding of why we need criticism, and to a lacuna in their proposal. Rather than try to repair it directly, I will recommend a shift in perspective: whether or not the effects of criticism that they describe are *now* rational, they can be incorporated into a technique which would *make* them rational. That technique repurposes the notion of reflective equilibrium: instead of treating it as a form of certification, we can make it into a trigger for disruption. That is, instead of taking your views to be warranted when they are in reflective equilibrium with one another, you can come to treat reflective equilibrium as telling you that it's time to start adopting—perhaps tentatively—alien and critical views as your own.

## 2

Reflective equilibrium became part of the vocabulary of analytic philosophers when Nelson Goodman proposed that we settle on rules of inference for reasoning, both deductive and inductive, as follows: “A rule is amended if it yields an inference we are unwilling to accept; an inference is rejected if it violates a rule we are unwilling to amend.”<sup>8</sup> And it was made a byword when John Rawls appealed to it in the course of figuring out how to select basic social institutions and principles:

if... [our] principles match our considered convictions of justice, then so far well and good. But presumably there will be discrepancies. In this case we have a choice. We can either modify the account of the initial situation [the “Original Position,” an idealized negotiation in which those principles are generated] or we can revise our existing judgments, for even the judgments we take provisionally as fixed points are liable to revision. By going back and forth, sometimes altering the conditions of the contractual circumstances, at others withdrawing our judgments and conforming them to principle, 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. This state of affairs I refer to as reflective equilibrium.<sup>9</sup>

---

<sup>8</sup>Goodman, 1983, p. 64, emphasis deleted.

<sup>9</sup>Rawls, 1971, p. 20; the passage comes with a citation to Goodman.

Philosophers today appeal to reflective equilibrium under quite diverse circumstances. For instance, the method of reflective equilibrium was appropriated by many moral theorists, as roughly this recipe. First, one would collect a number of moral reactions to actual or imagined circumstances. Then one would formulate a general principle whose instances or consequences largely agreed with the reactions. Lastly, one negotiated the disagreements, either allowing the principle to override the reaction, or, where one couldn't bring oneself to do that, adjusting the principle to accommodate the recalcitrant reaction. When no disagreements remained, one adopted the revised principle.<sup>10</sup>

In the course of these discussions, it became standard to distinguish “narrow” and “wide” reflective equilibrium, the narrow sort being produced by the back and forth between general and particular judgments we saw in the passages quoted from Goodman and Rawls, and the wide sort adding into the mix background theories, values, and pretty much anything else a relevant party happens to accept.<sup>11</sup> And so the recipe I just gave is often

---

<sup>10</sup>This way of proceeding did generate protest. For one thing, it seems to enshrine most of whatever you happen to think already, and who knows how you arrived at your former opinions, and what they're worth? Thus, Stephen Stich, complaining about the use of reflective equilibrium in epistemology, labelled it “cultural chauvinism” on this account (1990, pp. 94ff). Similarly, see Hare, 1981, p. 12.

If, as I suggested in note 7, when you check most views for coherence, they simply disintegrate, how can the method of wide reflective equilibrium, which I characterized as pressing on one's views with an eye to improving their coherence, produce the so very conservative results we see in philosophy?

John Stuart Mill once had to do spin control on his description of the Conservatives as “the stupid party,” and he glossed it by saying that he had not meant “that the Conservatives are generally stupid,” but rather “that stupid people are generally Conservative” (Reeves, 2007, p. 4). Likewise, it is not that all philosophers who deploy the method of reflective equilibrium are conservative, but rather that philosophers who are conservative anyway are likely to find the method congenial.

<sup>11</sup>The distinction was given wide currency by Daniels, 1996, esp. chs. 2, 4, who claims that it is already present in Rawls. See also Rawls, 1999, Elgin, 1996.

Once this move had been made, the method of wide reflective equilibrium was identified with what in epistemology had been called *coherentism*: revise your position to make it more coherent overall (e.g., at Daniels, 1996, pp. 2, 22). That was a mistake. Think of a political party adjusting its platform to keep groups with different agendas inside the tent; the platform is adjusted in one way to keep faction A from leaving, in another to keep faction B from leaving, and in the end, the platform is in equilibrium—but no one should expect the platform of a political party, and *a fortiori* a platform arrived at in this way, to be coherent.

Why was Goodman's and Rawls's novel concept almost instantly assimilated to a more familiar one? This sort of thing is depressingly frequent; it proves to be enormously difficult for philosophers to hold onto a new thought. Perhaps that is just what they do,

amended to make the equilibrium wide rather than narrow.

Because I think it is unsuccessful, I won't further discuss the application of reflective equilibrium in moral theory here.<sup>12</sup> But it was not just a fad in moral philosophy: here is—to take a further example almost at random—an announcement from the introduction to a book from the recent round of work on the metaphysics of vagueness: “I recommend seeking a reflective equilibrium that achieves the best balance between preserving as many as possible of our judgements or opinions of various kinds (intuitive and pre-philosophical ones among others) and meeting theoretical requirements such as simplicity.”<sup>13</sup> It is fair to say that a great many philosophers think of appeals to reflective equilibrium as a *safe* way of trying to solve just about any philosophical problem.

If the method of reflective equilibrium has substantive content, it must be appropriate only in some range of circumstances; just as there could be no such thing as a “General Problem Solver,” no genuine method works no matter what.<sup>14</sup> Recent professional practice notwithstanding, we had better ask: *when* does the method of reflective equilibrium make sense?

### 3

Reflective equilibrium has a home turf: a domain for which the method is obviously the right one, namely, linguistics.<sup>15</sup> When you are trying to construct a grammar for a language, you begin by asking native informants for their linguistic intuitions about a wide range of utterances. (Is *this* one grammatical or ungrammatical? How about *that* one?) And proceeding from this data, you formulate a set of rules which largely classify the utterances as the native speakers do. It will not necessarily agree with the native

---

but perhaps there is another force at work: if your method is reflective equilibrium, that method will itself incline you to adjust new ideas until they resemble older ideas.

<sup>12</sup>For the reasons, see Millgram, 2005, pp. 8–10, which also rehearses the justifications for using the method philosophers give in coffeeshop conversations; Millgram, 2008, discusses the problems involved in tightening up the notion into one that is sufficiently concrete to put to use in actual deliberation.

<sup>13</sup>Keefe, 2000, pp. 3, 38f, and she produces one of the coffeeshop responses mentioned in the previous note—“There is, I suggest, no possible alternative methodology” (p. 42)—thus owning up in print to what many philosophers will say only off the record. (Many but not all: for another exception, see Scanlon, 2003, p. 149.)

<sup>14</sup>For the label, see Newell and Simon, 1972, pp. 419ff.

<sup>15</sup>It's hard to believe that Goodman didn't take grammar as his model, although, as far as I know, he doesn't mention it; Rawls does mention it (1971, pp. 49), as does Daniels, 1996, ch. 4, and cf. also p. 4; the analogy is explored in Cohen, 1981.

speakers on all of them: some informants may hear “drug” as a perfectly acceptable past tense of “drag,” or, perhaps when the sentence is complicated, fail to match gender and number in the way the rule indicates. When they disagree, sometimes you rule in favor of the native speakers, and replace the proposed rule with a more contoured one. And sometimes you overrule the native speakers, dismissing one utterance or another as ungrammatical. Eventually, you arrive at a set of rules that systematize your data, along, of course, with a list of the data that fail to conform to your rules.<sup>16</sup>

This application of the method of reflective equilibrium is a sensible procedure because a language is primarily a means of communication.<sup>17</sup> All the users of a language must be able to deploy its conventions, which means the conventions must be shared. This is why it’s not possible (though I’m not going to try to decide whether the impossibility is logical, conceptual, or some other kind) for the grammar of English to be, in fact and unbeknownst to all, what you find when you open a grammar book for German.

Because humans acquire the tool by being trained in it, and because the training is ongoing, it must be generally possible to correct mistakes on the part of other speakers, which means that speakers must be able to discriminate correct from incorrect usage. (Even for those aspects of the language whose acquisition is pretty much hard-wired, such as the phonemics, you can pretty much tell when someone is doing it wrong—e.g., when they have an accent or a speech impediment.) A working communication tool of this kind requires its users to know how to use it, so when we want to find out if it’s being used correctly in this or that case, it makes sense to treat the users as authorities.

But only defeasibly so: because the tool won’t work if the encoding and decoding schemes aren’t suitably coordinated, there has to be a contrast between a particular user getting it right and getting it wrong on some occasion; which is a further reason why the possibility of correction has to be built into the language. It is that possibility that gives us the leeway to rule out some grammatical intuitions as mistakes, and, incidentally, gives grammar its prescriptive character.

---

<sup>16</sup>The problem of finding the *best* grammar for a language is, so understood, an early version of Minimum Description Length. (For an overview, see Grünwald, 2007.) Accordingly, a tighter formulation of the procedure would be of interest, in order to allow us to determine whether it is computationally hard. (‘Eventually’ might be quite a long time.) If, most of the time, we have to settle for *n*th-best grammars, and likewise, *n*th-best theories of other kinds, precisely because we are using the method of reflective equilibrium, that matters more in some domains than other, for reasons laid out in Millgram, 2000.

<sup>17</sup>Of course, it is not only that; for instance, Tugendhat, 2007, pp. 22f, 43, emphasizes some of the other distinctively human uses to which language is put.

The point I want to emphasize is this. In reflective equilibrium's home domain, the legitimacy of the method turns on features local to that domain, specifically, the social function of the linguistic practice a grammatical theory describes. That function requires that grammatical correctness not be too systematically at odds with speakers' judgments of correctness.

## 4

Let's now return to the use Rawls himself made of the notion. In attempting to design basic political institutions suitable for a contemporary liberal society, Rawls had decided that this was best treated as, precisely, a political problem. At this point in history, he observed, it is evident that neither citizens nor theoreticians agree on central matters of ethical and moral theory. So successful political theory means putting ethical questions to one side, and getting on with constructing institutions that, first and foremost, *work*. Indeed, this is why liberal political institutions are necessary; if citizens agreed on what the Good Life was, there would be no obstacle to using the power of the state to promote it. Liberal political theory is occasioned by what looks to be indefinitely persisting disagreement.

What is it for a set of political institutions to work? Minimally, they must be stable: that is, they should not simply fall apart, or collapse by provoking widespread rebellion. Rawls was writing during a time in which guerilla insurrections, popular uprisings, and regimes purporting to represent a working class revolution were an important part of the political scene, and he had been impressed by the need to give everyone within a polity a stake in the political system. (His famous Difference Principle, which allowed social inequality only when it could be shown to benefit the worst off, was meant to preempt proletarian revolutionaries, by putting defenders of the status quo in a position to remind the lower classes that, under an alternative arrangement, the lower classes would do even worse.) Once upon a time, disenfranchised subjects could perhaps be kept in line by brute force; but during the 1960s and 1970s, the lesson of recent history seemed to be that an alienated populace could and would bring the most seemingly permanent regimes crashing down.

Thus stability entails that political institutions have to seem *reasonable* to the citizens who live within them—reasonable *enough* to elicit the citizenry's cooperation. The method of reflective equilibrium elicits a set of principles that are (pretty much) guaranteed to seem reasonable; after all, if a principle has consequences that seem unreasonable, that will result in

revising the principle, unless the result is a principle that seems even more unreasonable.<sup>18</sup>

In the political domain, principles guide behavior, in the sense that political institutions only function when citizens who deviate from their dictates can be corrected; political institutions are necessarily and essentially coercive. That means that deviating must make sense as a possibility; however, if *all* behavior deviates—if no one pays attention to a set of political principles or institutions—then they exist in name only, in which case, demanding compliance is moot. Political institutions, that is, have to come with their characteristic brand of prescriptivity, and reflective equilibrium seems to deliver political prescriptivity in something like the right strength.<sup>19</sup> Taking these considerations together, it turns out that reflective equilibrium has a straightforward justification within the Rawlsian project: as in linguistics, it is a method that makes sense, given what the results are *for*.

## 5

Nelson Goodman had taken up the notion of reflective equilibrium in the course of thinking about how to solve what he called the New Riddle of Induction, and he proposed to use our reasoning about deductive inference as a model:<sup>20</sup>

---

<sup>18</sup>F&B attribute to Rawls the view that “intuitions as to what is just . . . transcend all manner of particular local practice,” and that “we are endowed with a sense of justice whose existence is independent of the historical contingencies of a particular form of life” (p. 95). This is a misreading. Rawls is committed to two requirements: first, that (sufficiently many) members of a society for which his institutions are suitable (these are Western states, and only as of recent times) share the sense of what is just and reasonable that he demarcates; second, that citizens raised within the institutions he specifies will, as a matter of psychological and sociological fact, largely come to share that sense of what is reasonable—or perhaps a somewhat further refined and entrenched version of it.

<sup>19</sup>For criticism of the role that widespread support plays in Rawlsian assessments of institutions, see Fehige, 1997, esp. sec. 3.1.

<sup>20</sup>Goodman, 1983, pp. 63f, emphasis deleted. The New Riddle arises with the observation that inductive inferences deploying *some* predicates seem sane, plausible and worthy of our respectful endorsement, whereas very similarly shaped inferences deploying other predicates seem just plain crazy. To take a toy example, the grass on my lawn has always been green, so I infer, inductively, that it will be green tomorrow, and that’s pretty plausible. But it is also true that the grass on my lawn has always been (and here comes a complicated predicate): green, at all times up to the present, or blue tomorrow. The analogous inductive inference, deploying *this* predicate, would tell me that my lawn is going to turn blue tomorrow, and that’s just crazy. The New Riddle asks for a way of sorting the predicates into the ones that deliver sane inferences, on the one hand, and the ones that deliver crazy inferences, on the other. This turns out to be harder than it looks,

How do we justify a deduction? Plainly, by showing that it conforms to the general rules of deductive inference. . .

Principles of deductive inference are justified by their conformity with accepted deductive practice. Their validity depends upon accordance with the particular deductive inferences we actually make and sanction. If a rule yields unacceptable inferences, we drop it as invalid. Justification of general rules thus derives from judgments rejecting or accepting particular deductive inferences.

Notice, against what seems to be a widespread but inattentive reading of his work, that Goodman had in mind an argument in favor of using the method of reflective equilibrium to settle questions in deductive and inductive logic. Specifically, by claiming that reflective equilibrium is how we *do* go about legitimizing inference rules, he gestured at a reflective-equilibrium argument for reflective equilibrium. Arguments of this form (I mean, using a logical device to argue for the validity of that very logical device) have a long and controversial history, and I don't now want to take up the question of whether the circularity is vicious. Here I mean to highlight, first, that the philosopher most responsible for introducing the method into the contemporary debate acknowledged something which many of its users have since lost track of, namely, that it requires a supporting argument if it is to be

---

and when Goodman made the New Riddle famous, it obliged by reciprocating.

Goodman's own solution to the New Riddle was *entrenchment*: roughly, that the more a predicate has been used, the less crazy it is to use it. ("Roughly": on the one hand, Goodman's account of entrenchment is somewhat more nuanced than that; on the other, it was clearly not intended as a finished product, and during a lengthy and difficult exchange which commenced with Zabłudowski, 1974, it became clear that it was not anywhere close to satisfactory. So I won't be picky about the fine points here.) That solution is a little bit hard to swallow; it seems to entail that sheer dogged stubborn stupidity will be vindicated if it is just persistent enough. So you would expect entrenchment to have been supported by an argument, and you might have expected the argument to deploy the method of reflective equilibrium: that Goodman would write a scholarly intellectual history of concept use, in an attempt to show that the predicates we take to be inductively legitimate ('projectible,' in Goodman's jargon) are the most entrenched ones. But no such history is on offer, and there is a likely explanation for that fact.

Entrenchment is entailed by the method of reflective equilibrium itself, and so does not need to be established by an application of the method. Reflective equilibrium converts our judgments, in the majority of cases, into a norm. The more cases we accumulate (the more entrenched a predicate becomes), the weightier the majority becomes. Furthermore, the more habituated we become to using the predicate: to the extent that our current judgments are the effect of habituation, our current judgments will tend to agree with those produced by application of the entrenched predicates. After a while, the judgments we find to be in reflective equilibrium *will be* the entrenched ones, and vice-versa.

legitimate; for Goodman, reflective equilibrium did not go without saying. Second, I want you to notice that any such argument would be local to a particular domain, for only within a circumscribed domain is it possible to survey whether the technique that has been put to use is in fact in reflective equilibrium.

Goodman did not give the argument at which he gestured. Actually providing it would have involved documenting instances in which the method was used and thought reasonable. However, the sort of historical treatise that you would thus be led to expect remained absent, and for my own part, I do not think it could have been filled in; the history of logic does not supply what Goodman's argument sketch requires. Logic overrides people's judgment in ways that grammars need not; people think logically, by and large, only after rigorous training, whereas they speak more or less grammatically whenever they speak their native language.<sup>21</sup> The contrast is especially vivid for recent extensions of logic, such as probability theory, and I hope to argue elsewhere that the history of means-end inference also fails to confirm the reflective-equilibrium account of logical method. For now, I want only to register that although the argument Goodman envisioned did not take the form of an appeal to the point of a practice, the only successful arguments for reflective equilibrium that we actually have *do* take that form.

As we have just observed, Goodman's would-be argument for reflective equilibrium as the preferred method of philosophical logic was rather different in shape from those we saw supporting its use in linguistics and political theory. So I want to indicate how one could put together arguments for using reflective equilibrium in philosophy of logic that conform—partially—to the model of the arguments for political theory and linguistics. Let's take induction first, and allow that the point of logic is to guide our reasoning: logic is to serve as a standard for correctness of inference. Thus, on the one hand, if it is to serve as a standard, it must be able to correct us when we go wrong; the dictates of logic have to be able to disagree with logic's users, at any rate sometimes. But on the other hand, inductive reasoning is defeasible, and defeasible reasoning can't be executed mechanically; this means in turn that it has to make sense to its users. Let me explain.

An inference is defeasible when—even if it looks alright as it is—further information can defeat it.<sup>22</sup> For example, take the inference from

1. My lawn was green ten summers ago.

---

<sup>21</sup>Not a new point; see, e.g., Stich, 1990, pp. 83f.

<sup>22</sup>If you're from the computer science world, you're probably familiar with this concept under the name *nonmonotonic inference*.

2. My lawn was green nine summers ago.
3. My lawn was green eight summers ago.
- ⋮
- to
4. My lawn will be green next summer.

That inference is alright, as far as it goes. But it can nevertheless be defeated by further facts, such as:

- There's going to be a severe drought and water rationing next summer.
- My housesitter will forget to water the lawn.
- My landlord will asphalt it over.
- ⋮

—which is to say, since this is a typical inductive inference, that induction is defeasible.

If an inference pattern is defeasible, you have to determine when to proceed to the conclusion, and when to abort it in the face of some further bit of information. That judgment call requires that the inference make sense to you, and it will make sense only if the profile of correctly drawn inferences of this type agrees with, at any rate, *most* of your particular judgments (as to the satisfactoriness of inferences of that kind). So reflective equilibrium, which imposes this kind of overall agreement as a constraint, while allowing the standard to flex its muscles and correct your inference on occasion, is a reasonable method for tracing out *usable* rules of inductive inference.<sup>23</sup>

There may be a way to adapt this line of argument to deductive inference as well. Deductive reasoning is not defeasible, and you can check the validity of a deductive proof mechanically (whether it makes sense to you or not). But normally you can't *find* or *construct* a deductive argument without understanding what's going on.<sup>24</sup> For the most part, if a set of purportedly deductive inference patterns doesn't make intuitive sense to you, you will be unable to strategize about how to construct your deductive arguments. And that means that you need deductive rules which (for the most part) make sense to you. And that in turn means that (with a caveat I'll get to

---

<sup>23</sup>I was helped in my thinking here by Mikel Negugogor.

<sup>24</sup>There are borderline cases, e.g., the famous proof of the Four Color Theorem.

in just a second), even for deductive inference rules, reflective equilibrium is a reasonable way to arrive at them.

Two observations: First, the way of understanding logic in play in this argument takes back the early-20th-century rejection of psychologism. Logic is being understood as, roughly, the grammar of a practice of inference, a practice in which its users must feel at home. In this it follows Goodman's own line of reasoning, and contrasts strikingly with one contemporary way of thinking of logic, as the science of the consequence relation.<sup>25</sup> Second, the suitability of the method of reflective equilibrium for logic turns on a question that did not come up for linguistics and political theory. If the point of inference is to get us to conclusions we're happy with, whatever logic results from the method of reflective equilibrium will be fine. If, however, the point of inference is to be right about facts we can't control, then while reflective equilibrium must characterize any logic we can use (or anyway, use in the normal circumstances of daily life), and so is in that respect a *sine qua non*, a logic may be in reflective equilibrium, and nonetheless remain pointless, because the inferences it directs turn out not to match the facts we ultimately encounter.<sup>26</sup>

## 6

The sample size is small—there are only a handful of muscular and visibly successful uses of the method of reflective equilibrium—but I think it's not too early to adopt the following as the upshot of the survey. We had better treat the method of reflective equilibrium as philosophically appropriate only when it comes with a domain-specific explanation of *why* it is. That justification will turn on what is perhaps an only loosely-characterizable

---

<sup>25</sup>This way of phrasing it is lifted from Sorensen, 2001, p. 10; for historical and sociological treatments of the psychologism debate, see Kusch, 1995, Ringer, 1990, pp. 295–298.

<sup>26</sup>Thus Cohen, 1981, pp. 317, 320, “any normative analysis of everyday reasoning... must in the end rely for its defence on the evidence of relevant intuitions,” and so a “normative theor[y used to]... evaluate the rationality or irrationality of a naive subject's inference... is itself acceptable for the purposes only so far as it accords, at crucial points, with the evidence of untutored intuition.” But not everyone accepts the caveat; he continues:

The best that normative theorists can hope for in this field... is that the contents of all relevant intuitions... can be made to corroborate one another by being exhibited as the consequences of a consistent and relatively simple set of rules or axioms that also sanctions other intuitively acceptable... patterns of reasoning. (322)

family of features of the various domains, but nevertheless, a family whose members we cannot expect to find across the board: e.g., the need for users of a communication system, or of a manual for reasoning, or of political institutions to find their guidelines (sufficiently) intuitive to apply. This directive is itself defeasible: if someone comes up with an argument showing that reflective equilibrium is an across-the-board reliable certification, then more power to him. But in the meantime, the sort of support I've reviewed is the only support we have.

Even within philosophy, not every occasion for working up a theory looks like this, and, outside philosophy, domains with these features are rare. If you are the sort of metaphysician who wants to find out what the mind-independent basic constituents of the universe are, and you propose to do so by arriving at a reflective equilibrium, you are—sorry, folks—*just pretending*. And of course when (to pull just one example off the top of a very large pile) Feynman, Dyson, Tomonaga and Schwinger formulated what we now know as quantum electrodynamics, they didn't attempt for a moment to bring anyone's intuitions into a reflective equilibrium, because doing so would have been a pointless distraction: it doesn't *matter* that QED comes out easy for people to wrap their hearts and minds around.<sup>27</sup>

## 7

If the standard philosophical rendering of the revision of a theoretical or evaluative position from the inside accounts for change only within the archipelago of domains in which user-friendliness matters, then turning to the possibilities opened up by criticism on the part of outsiders with systematically differing theoretical and evaluative positions is indeed a plausible followon choice.<sup>28</sup>

---

<sup>27</sup>For the history, see Schweber, 1994.

<sup>28</sup>One can benefit a great deal from taking seriously criticism that comes from someone whose views one does not share. However, I'm unsure that F&B are correctly representing what that process looks like in real life. This is just a bit of autobiography, but while I do find criticism to be invaluable, my responses to it only rarely amount to adopting a critic's views. A local illustration: I've benefitted a good deal from F&B's discussion, but not by being induced to surrender entrenched views. Rather, they've stimulated me to reconsider the role of criticism; to articulate an objection to my former views of friendship; to press the case against the indiscriminate use made of reflective equilibrium within contemporary philosophy. Briefly, considered revision of one's views in the face of criticism is not nearly always the effect of dropping one's own views in favor of the views of the critic. That reservation registered, here I am going to follow F&B in their focus on the effects of criticism in which they are primarily interested.

We are imagining someone who has a hermetically closed view: it is by hypothesis impervious to any objection whose premises would be allowable by its own lights. F&B point out that someone with a different but not too distant view can construct arguments against it that invoke premises which the hermetically closed view in principle disallows. When the proponent of the argument is trusted, the owner of the hermetically closed view may lose track of which premises are his own and which are his critics. His take on things becomes “destabilized”; out of this can emerge a new and different view, and F&B reiterate that this procedure is “effective”.<sup>29</sup> However, the attentive reader will notice that they never get around to arguing that it is *legitimate*.<sup>30</sup> When someone on the receiving end of such an argument accepts, even tentatively and ambivalently, premises which his position commits him to rejecting, why isn’t he simply making a *mistake*?

If being destabilized in this way is not legitimate, the maneuver should be effective only temporarily, and we can see the point by considering what can happen in the context of love or friendship. Couples in romantic relationships or marriages are often seen to exhibit a local and personal version of the phenomenon F&B describe, where each member of the couple adopts opinions, preferences, evaluative postures and so on from the other, and it is natural to describe the couple as deliberating jointly from a shared pool of views held in common.<sup>31</sup> However, after the divorce or breakup, it is also common to see the defunct couple’s members going through a period of readjustment, one which is naturally described as each of them remembering who they are. Gradually or suddenly, their former preferences in movies and music reappear; opinions that had been negotiated in the sort of back and forth described by F&B vanish, and former ways of seeing reemerge; evaluations that had been submerged are back in full force. Now, this does not always happen; but the explanation that forefronts itself is that precisely when the views and stances merely result from psychologically effective destabilization within the relationship—from losing track of what one really thinks—they are unlikely to outlast the need for personal accommodation. Only when the revision is understood to be legitimate should it be expected to stick, a expectation corroborated not only in the personal sphere, but, *mutatis mutandis*, when it comes to political *Weltanschauungen* and even in philos-

---

<sup>29</sup>Thus, F&B’s remedy comes under the heading of what DePaul, 1993, p. 49, calls “alchemical methods.”

<sup>30</sup>See, e.g., Fisch and Benbaji, 2011, pp. 258f, esp. the “smuggle in” at p. 259, 262, 275, and the final “we insist”—instead of, say, “as we have shown”—three lines from the very end of the book.

<sup>31</sup>Ebels-Duggan, 2008.

ophy.<sup>32</sup> Once again, legitimacy in this process, as opposed to effectiveness, is just what F&B are unequipped to explain.

## 8

The difficulty that F&B are attempting to surmount, turning on what they take to be the limits of internal criticism, presupposes a distinction between *internal* and *external*, which comes to roughly this: if the starting points of the argument in question are things that you believe, want and so on, then the criticism is internal; whereas if your interlocutor asserts that they are true, desirable or whatever, even though you don't think so, then the criticism is external.<sup>33</sup> The distinction between inside and outside evidently serves a *logical* function: if I believe that  $p$  and also that  $\sim p$ , then I am contradicting myself, I am irrational, and I should revise my beliefs; whereas if I believe that  $p$  and *you* believe that  $\sim p$ , we are merely disagreeing, and

---

<sup>32</sup>A couple of illustrations: In the aftermath of World War II and of the *Wende* of 1989–92, Nazism and Communism respectively were no longer credible positions to adopt in public policy debate. But they are now back: in Germany, for instance, parties that inherit these ideologies regularly garner representation. (That claim comes slightly hedged: for legal reasons, in Germany far right parties cannot even now advertise themselves as Nazi.) No doubt many Germans took events to have discredited the parties and their ideologies; but others were merely forced to accommodate themselves to an environment that would not tolerate their views. When those pressures faded, the previous views returned.

Within philosophy: The British Empiricist tradition took it for granted that sensations were identifiable entities playing a foundational role in epistemology, theory of meaning and moral theory. During the middle decades of the twentieth century, the claim that there were such things as sensations was subjected to devastating criticism, most importantly by Wittgenstein and Austin. Remarkably, as of the turn of the millennium, sense-data are back, and serve as the subject matter for “consciousness studies.” The arguments against them were not refuted, but rather ceased to be taught or invoked. One noticed a palpable sense of relief as it became clear that it was now socially acceptable to ignore them; while some philosophers understood and appreciated the point of the criticisms of so-called ordinary language philosophy, other philosophers had merely accommodated themselves to a social reality by accepting postures and constraints on theorizing, which they abandoned as soon as it was feasible to do so.

<sup>33</sup>F&B couch much of their discussion in terms of communities, but here I think it will make things easier to consider the individuals making up a community.

This contrast is related to one made famous by Bernard Williams: he introduced the notion of an internal reason for action as one that could be arrived at deliberately starting out from your “subjective motivational set”; an argument for doing something made out on the basis of someone *else's* motivations, or his claims about what is important or valuable, would count as external. (See Williams, 2001, and for discussion, Millgram, 1996.)

as far as rationality is concerned, revision is not required. The laws of logic are applied only to persons and amongst the attitudes of a single person—or to items that we personify, such as merely legal persons and, when speaking loosely, theories, positions, views and the like.<sup>34</sup>

It is instructive to see this way of delimiting the scope of the laws of logic as a solution to an engineering problem, and so let me walk you through a multistage thought experiment.

Imagine a *mad metascientist*: impatient with doing research himself, he decides to automate his research processes, which he will then only have to supervise. At the outset, he builds a supercomputer; he equips it with a library of past observations, and a fleet of drones to gather new observations; he loads it with content from all the best science textbooks; finally, he gets to work on algorithms meant to troubleshoot for—as F&B put it a number of times—consistency, coherence, explicitness and clarity. All these are preliminaries, since his intent is to develop and run code that will output novel and exciting results; but he never gets beyond the preliminaries, because he discovers that even the initial cleanup is impossible.<sup>35</sup>

Our mad metascientist goes back to his drawing board. He dispenses with his supercomputer, and upgrades his formerly remote-controlled drones: they are now free-ranging, semi-autonomous *roboscientists*. A roboscientist stores in memory representations of facts, of objectives, of standards and guidelines of various sorts, and perhaps of other items also. The mad metascientist's trick is to assign each of his roboscientists a much smaller field of representations than to the defunct supercomputer, and to make sure that—by virtue of only incorporating one textbook at a time—they start off exhibiting some semblance of consistency and coherence.

As the alert reader has already figured out, the roboscientists are meant as simplified models of persons. Imagine a population of roboscientists,

---

<sup>34</sup>See Millgram, 2009, sec. 5.1, and here is a warmup objection to internalism in the accent of Jonathan Lear. (For the approach, see Lear, 1990, and for a critical overview, Millgram, 2001.) The distinction between internal and external must be a psychological achievement. Given what we know of human psychological development, that sort of demarcation will almost never be complete: rather, what we should expect to find are only occasional boundaries to the self. So the contrast presupposed as uniformly available by, for instance, Williams's argument against external reasons can only rarely be made out.

<sup>35</sup>That's an assessment of the overall difficulty of the task, given the hodgepodge of science textbooks that have to be put in order; but actually it's also provable. Consistency involves at least truth-functional consistency, and on some philosophers' views, Bayesian consistency; however, SAT and Bayesian updating are both computationally intractable. (See Garey and Johnson, 1979, pp. 38ff; Cooper, 1990.)

each of which implements the following representation management technique. When it is initialized, the field of representations within the scope of its logical responsibility normally contains occasional contradictions and other violations of relevant rules. It surveys its field of representations, moving a spotlight of attention over it, and picking out violations within the spotlight.<sup>36</sup> Each time it locates a contradiction or other violation, it initiates an investigation; it adjusts its representations of the outer world to one another (sometimes by seeking and bringing to bear further empirical data); it likewise adjusts its more general standards and its highly specific goals to one another. When the violation has been addressed, it moves on to the next. When there are no more violations to identify, it halts.

Assigning to persons the responsibility for keeping smallish fields of representation sufficiently consistent and coherent makes the use of representations to control activity and track the state of the world a manageable enterprise. Looking at that last point from the other end: for many philosophical purposes, persons just *are* such scope assignments. (However, we have to understand that point cautiously: it is not as though the representations are there first, and the persons are lines drawn around sets of them; rather, persons and the representations they manage are jointly introduced as implementations of this engineering technique.) Inference is one way of eliminating inconsistency, and so scope-delimited fields of representation are an inference-enabling architecture.<sup>37</sup>

Let me introduce a detail-oriented implementation point that will be useful shortly. A roboscience may have a representation stored in memory that it will not treat as potentially contributing to an inconsistency (just as humans sometimes daydream). So we will need to give roboscience a way of tagging representations as falling within or outside of the scope of its logical responsibility. Let's make 1 stand for being within that scope, 0 for outside, and reserve values between 0 and 1 for a use I'll introduce momentarily. In honor of the philosopher who initiated the discussion of the feature of personality we're so crudely modeling, we can call these tags on a roboscience's representations *Frankfurt values*.

Now if you think about it, the design approach I have described—and

---

<sup>36</sup>Hoadley *et al.*, 1994, models this technique for focusing resources on a computationally tractable region of a more general coherence problem.

<sup>37</sup>It should go without saying that in asking you to look at persons as a solution to an engineering problem, I am not suggesting that persons have been designed by a divine robotics engineer. I am also not implying that the design solution was selected for over a history of Darwinian evolution. Rather, I am gesturing at a creature construction argument; see Millgram, 2014.

that we implement—is actually more ambitious than I have made out so far. The point of eliminating contradictions and other incoherences in our respective fields of representation is not only to keep the system of local representations in good order, but to initiate and sustain behavior that amounts to exploration and innovation. When you find that your depiction of the world around you includes pairs of representations with the respective forms  $p$ ,  $\sim p$ , you do not simply delete one or the other of them at random: rather, as I have suggested, you come to have the task of figuring out *which one* is correct, which may mean empirical investigation, or looking back to see where you went wrong, or many other things. Contradictions, together with other, sometimes more subtle violations of the etiquette of rationality, are the fuel that drives the most demanding—and also the most necessary—forms of intellectual activity.

If contradictions and incoherences are mental fuel, what do we do when we run out? As it happens, we have already have a label for being out of gas: the roboscientists in our thought experiment are converging on their respective reflective equilibria. From the point of view of the mad metascientist, reflective equilibrium is a *problem*: the roboscientists are no longer producing innovations, are no longer exploring, and are no longer generating surprising or useful results for him to harvest.

You might think that when it's humans, rather than imaginary roboscientists, having arrived at a reflective equilibrium is a good thing. So, time for a reality check.<sup>38</sup> If you're talking to someone who tries to bring you around to his view by saying, "But can't you see how it all *fits*!?"—well, a lot of the time, if it really *does* all fit, you know it *can't* be true. In such cases, it's the reflective equilibrium itself that tells you that you're talking to a crank. In real life, and it behooves philosophers never to forget this, reflective equilibrium is very often a mark of *craziness*. (The rule of thumb is domain dependent: grammar and more generally members the class of special cases discussed in secs. 3–5, above, are reasonable domains to have everything fit; likewise, my view about my own name and the names of family members.)

## 9

No doubt there are various expedients our mad metascientist might consider. For instance, he could restart his army of roboscientists by manually

---

<sup>38</sup>Here I'm riffing on the remarks at Putnam, 1981, p. 129.

resetting their representational fields. But here is an interesting alternative that might prove to be more useful.

The mad metascientist can designate some roboscientists as *neighbors* when there is more than a (fairly high) minimum overlap between their representational fields. (This is meant as a crude proxy for F&B’s “trusted critic” status.) Let’s suppose that there’s a straightforward way to tell when a roboscientist has arrived at a reflective equilibrium. When a roboscientist runs down, a handful of neighbors’ representations that are not in the respective overlaps are copied to the depleted roboscientist’s representational field; they are assigned random Frankfurt values ( $0 \leq f \leq 1$ ). (This is another crude proxy, this time for ambivalence; we can imagine the metascientist settling on surrogate inferential behavior suitable for his model, e.g., making it easier to give up ambivalently held premises of an argument when its conclusion is hard to swallow.) Often enough, the effect will be to provide the roboscientist with new incoherences to resolve, and even straight contradictions; it can resume its exploratory activities; after a suitable period, there will be a new crop of results ready for harvesting.

Let’s introduce a name for the strategy: *intermittent fuzzy scoping*. Intuitively, the idea is to temporarily and selectively make the boundaries of the person vague in order to improve average intellectual or practical performance. Is the strategy a good idea?

In the scenario we are considering, presumably only extensive experimentation will answer that question. And of course the reasonableness of intermittent fuzzy scoping should be expected to vary from domain to domain. (If I’m settled in my views about what my name is, whether I have a cat and so on, that doesn’t give me an occasion to treat those views as needing disruption.) So rather than trying to anticipate its outcome, let’s consider what it would take to apply an answer within the scenario to our own views as to what is rational for us.

Notice, first of all, that the mad metascientist must have available a way of assessing the results that his roboscientists produce. Only if he does can he determine which ones to harvest, and only if he does can he decide whether a particular model of roboscientist is producing good results on average. That sort of standard of assessment is not part of F&B’s understanding of the problem we collectively face, and this is not the place to explain what I think one might use, were one to take on the role of mad metascientist oneself.<sup>39</sup> However, we can say this: if we tie success to problem-solving

---

<sup>39</sup>For a suggestion as to where such standards would have to come from, see Millgram, 2006.

(that is, if we allow results to be of good quality when they solve problems), we have to leave behind the narrowly instrumentalist conception of what it is to be a problem, on which it is set by a goal or objective that you already have.<sup>40</sup>

Second, notice that we have a use for the story of the mad metascientist only if we can position ourselves to think of the investigation of both our logic and our personhood not as discovery problems (i.e., as a matter of finding out what the true theories of logic and of persons *already* are), but as design and invention problems. That is not a trivial shift of perspective. To understand persons and logics in engineering terms, we have to stop thinking of ourselves as facing the so-called logocentric predicament, and instead think of ourselves as being, at any point, equipped with some repertoire of inferential tools, which we can use to design and manufacture more sophisticated inferential tools to add to the repertoire. With unusual exceptions, that is not how we philosophers have been thinking about the logical enterprise.

Let's recap. Reflective equilibrium can function as a form of certification for a constellation of views, but as far as we know, and as far as it is prudent to assume, only locally: in grammar, in settling on political arrangements, and perhaps in other areas also. Elsewhere, however, we can perhaps make reflective equilibrium useful, but in a very different role. If we adopt the procedures in which this technique is embedded, when you find yourself in reflective equilibrium, that will tell you it's time to blur the boundaries between yourself and your neighbors, that is, to take on some of their views, albeit ambivalently, as your own.

If we took the sort of thought experiment we sketched, supplemented with modeling work and experimentation, to be a good reason to accept such procedures, we would do so on the understanding that both logic and the boundaries of persons are in any case to be understood as solutions to engineering problems, and so can be thoughtfully and intelligently reengineered. And in that case, that is, if we adopted these procedures as part of our mode of rationality, F&B's proposal for revising one's views on the basis of trusted criticism would not merely be effective (and would not merely be *temporarily* effective), but *legitimate*.

---

<sup>40</sup>This is F&B's account of problems, and so I am offering this point as a reason to amend their view.

## References

- Chase, J. and Reynolds, J., 2010. *Analytic versus Continental*. Acumen, Durham.
- Cohen, L. J., 1981. Can human irrationality be experimentally demonstrated? *Behavioral and Brain Sciences*, 4, 317–370.
- Cooper, G., 1990. The computational complexity of probabilistic inference using Bayesian belief networks. *Artificial Intelligence*, 42, 393–405.
- Cooper, J., 1980. Aristotle on friendship. In Rorty, A., editor, *Essays on Aristotle's Ethics*, pages 301–340, University of California Press, Berkeley.
- Copp, D., 1984. Considered judgments and moral justification: Conservatism in moral theory. In Copp, D. and Zimmerman, D., editors, *Morality, Reason and Truth*, pages 142–168, Rowman and Allanheld, Totowa.
- Daniels, N., 1996. *Justice and Justification*. Cambridge University Press, Cambridge.
- Daniels, N. 2013. Reflective equilibrium. The Stanford Encyclopedia of Philosophy (Winter 2013 Edition), Edward N. Zalta (ed.), URL = <http://plato.stanford.edu/archives/win2013/entries/reflective-equilibrium/>.
- DePaul, M., 1993. *Balance and Refinement*. Routledge, New York.
- Ebels-Duggan, K., 2008. Against beneficence: A normative account of love. *Ethics*, 119(1), 142–170.
- Elgin, C., 1996. *Considered Judgment*. Princeton University Press, Princeton.
- Fehige, C., 1997. Rawls und Präferenzen. In Hirsch, W. and der Philosophischen Gesellschaft Bad Homburg, editors, *Zur Idee des politischen Liberalismus*, Frankfurt a. M., Suhrkamp.
- Fisch, M. and Benbaji, Y., 2011. *The View from Within*. University of Notre Dame Press, Notre Dame.
- Garey, M. and Johnson, D., 1979. *Computers and Intractability: A Guide to the Theory of NP-Completeness*. W. H. Freeman, New York.

- Goodman, N., 1983. *Fact, Fiction, and Forecast, Fourth Edition*. Harvard University Press, Cambridge, Mass. Foreword by Hilary Putnam.
- Grünwald, P., 2007. *The Minimum Description Length Principle*. MIT Press, Cambridge.
- Hare, R. M., 1981. *Moral Thinking*. Oxford University Press, Oxford.
- Hoadley, C. M., Ranney, M., and Schank, P., 1994. WandereCHO: A connectionist simulation of limited coherence. In Ram, A. and Eiselt, K., editors, *Proceedings of the Sixteenth Annual Conference of the Cognitive Science Society*, pages 421–426, Erlbaum, Hillsdale.
- Keefe, R., 2000. *Theories of Vagueness*. Cambridge University Press, Cambridge.
- Kelly, T. and McGrath, S., 2010. Is reflective equilibrium enough? *Philosophical Perspectives*, 24(1), 325–359.
- Kusch, M., 1995. *Psychologism*. Routledge, New York.
- Lear, J., 1990. *Love and Its Place in Nature*. Farrar, Straus and Giroux, New York.
- MacIntyre, A., 1988. *Whose Justice? Which Rationality?* University of Notre Dame Press, Notre Dame.
- MacIntyre, A., 1990. *Three Rival Versions of Moral Enquiry*. University of Notre Dame Press, Notre Dame.
- Millgram, E., 1987. Aristotle on making other selves. *Canadian Journal of Philosophy*, 17(2), 361–76.
- Millgram, E., 1996. Williams' argument against external reasons. *Nous*, 30(2), 197–220.
- Millgram, E., 1997. *Practical Induction*. Harvard University Press, Cambridge, Mass.
- Millgram, E., 2000. Coherence: The price of the ticket. *Journal of Philosophy*, 97(2), 82–93.
- Millgram, E., 2001. Review of Jonathan Lear, *Love and Its Place in Nature and Open Minded*. *Mind*, 110(440), 1087–1092.

- Millgram, E., 2005. *Ethics Done Right: Practical Reasoning as a Foundation for Moral Theory*. Cambridge University Press, Cambridge.
- Millgram, E., 2006. Review of Bas van Fraassen, *The Empirical Stance*. *Philosophical Review*, 115(3), 404–408.
- Millgram, E., 2008. Specificationism. In Adler, J. and Rips, L., editors, *Reasoning: Studies of Human Inference and its Foundations*, pages 731–747, Cambridge University Press, Cambridge.
- Millgram, E., 2009. *Hard Truths*. Wiley-Blackwell, Oxford.
- Millgram, E., 2014. Private persons and minimal persons. *Journal of Social Philosophy*, 45(3).
- Nehamas, A., 1998. *The Art of Living*. University of California Press, Berkeley.
- Newell, A. and Simon, H., 1972. *Human Problem Solving*. Prentice-Hall, Englewood Cliffs.
- Putnam, H., 1981. *Reason, Truth and History*. Cambridge University Press, Cambridge.
- Rawls, J., 1971. *A Theory of Justice*. Harvard University Press, Cambridge, Mass.
- Rawls, J., 1999. The independence of moral theory. In Freeman, S., editor, *Collected Papers*, pages 286–302, Harvard University Press, Cambridge, Mass.
- Reeves, R., 2007. *John Stuart Mill: Victorian Firebrand*. Atlantic Books, London.
- Ringer, F., 1990. *The Decline of the German Mandarins*. Wesleyan University Press, Hanover.
- Scanlon, T. M., 2003. Rawls on justification. In Freeman, S., editor, *The Cambridge Companion to Rawls*, pages 139–167, Cambridge University Press, Cambridge.
- Schweber, S., 1994. *QED and the Men Who Made It*. Princeton University Press, Princeton.

Sorensen, R. A., 2001. *Vagueness and Contradiction*. Oxford University Press, Oxford.

Stich, S., 1990. *The Fragmentation of Reason*. MIT Press, Cambridge, Mass.

Tugendhat, E., 2007. *Anthropologie statt Metaphysik*. C. H. Beck, Munich.

Williams, B., 2001. Internal and external reasons (with postscript). In Millgram, E., editor, *Varieties of Practical Reasoning*, MIT Press, Cambridge, Mass.

Zabludowski, A., 1974. Concerning a fiction about how facts are forecast. *Journal of Philosophy*, 71(4), 97–112.