

## Chapter 16

### Why Bother with Reflective Equilibrium?

Michael R. DePaul

#### A Short Dialogue

Imagine I am sitting in a bar, in fact, that I am belly up to the bar. It is not the sort of pub one finds around a university—filled with academic types. It is an ordinary joint, filled with ordinary folks. If I strike up a conversation with the person beside me, at some point I'll probably be asked—

Friend: What do you do?

Me: I teach at the university.

(I have always envied colleagues from other disciplines who can dispense with this evasion: "I'm a chemist," "I'm a psychologist," or even "I'm an historian," might pass, but can you imagine coming right out with "I'm a philosopher.")

Friend: Oh! What do you teach?

(My evasion never does much good.)

Me: Philosophy.

(I cringe whenever I say this, since God only knows what ordinary folks think of when they hear "philosophy." As I say it I cannot help thinking of the sorts of books stacked on the "Philosophy" shelf in American shopping mall bookstores. But today I am lucky. I have struck up a conversation with someone who has no preconceptions at all, so a simple request for clarification follows. Historical figures are a safe bet, so I mention some of the usual suspects: Plato, Aristotle, Descartes, Kant, and, to end up with a name that might ring a bell with a mysterious and profound tone, *Wittgenstein*. Such names are enough to scare most people off. But not everyone! Today I am *not* so lucky. I am asked what sort of things these people wrote about.)

Me: Philosophers are interested in very fundamental questions, for example: What is it to know something? What, if anything, can we know? Is ev-

everything that happens caused? What is it for one thing to cause another? What sort of life is best for a human being? What is the nature of virtue? Of moral obligation? And so on.

(Ethics is a lot closer to home than metaphysics or epistemology, and besides, even in these secular times, most people have been exposed to religious approaches to moral questions, so it is a good bet the conversation will strike off in that direction.)

Friend: Now I understand—philosophy is like religion. You study the Bible and perhaps other religions, sacred texts, the *Bhagavadgita*, . . . What do you think about reincarnation?

(I can only take so much, even when pacified by beer!)

Me: *That just isn't right!* Philosophers do *not* approach these questions as religion does. We do *not* rely upon sacred texts or the teachings of some church.

Friend: But then how *do* philosophers go about answering these questions?

(I am afraid the prospects for an honest answer to this question are grim! I believe the method of reflective equilibrium, first described and advocated by Nelson Goodman (1965) and further developed and applied to ethics by John Rawls (1971, 1974), describes the approach the vast majority of philosophers in fact follow. More importantly, it provides an enormously influential answer—in my opinion, close to a correct answer—to the more interesting and philosophical question: How *should* we conduct philosophical inquiry? But can you imagine trying to explain and justify this method to your ordinary Jane or Joe who has been pounding nails or tightening nuts all day?)

Me: Let me try to explain. I'll use moral theory as an example, but remember that philosophical inquiry into other matters, for example, knowledge, causation, reference, or the nature of belief, is similarly conducted. The philosopher must begin her inquiry regarding morality with the moral beliefs she happens to have, such as, beliefs about what is morally good, which acts are right and wrong, or about when guilt is appropriate. Some of her judgments will concern actual things or actions, but others will be about imaginary or hypothetical cases, such as, actions performed by the characters in a movie or novel. And some of her judgments will concern general principles, for example, she will likely judge that two actions cannot differ morally without also differing in some relevant nonmoral feature. The philosopher then attempts to eliminate any beliefs or judgments formed in circumstances that obviously make error likely, such as, when she is ignorant of potentially relevant facts or her personal interest is somehow at stake. Her next task is to construct a "theory" that accounts for the remain-

ing judgments: her "considered" judgments. In attempting to construct this theory the philosopher's considered judgments do not function as a scientist's data is commonly thought to function, for the philosopher seeks to bring her considered judgments into balance with a theory via a process of *mutual adjustment to both her theory and her considered judgments*. Here's what I mean: Suppose that, after she has a good start on an acceptable theory, that is to say, she has constructed a theory that accounts for a wide range of her considered judgments, the philosopher discovers that this promising theory is in conflict with some of her other considered judgments. The philosopher is not bound to revise the theory so that it accords with these judgments. Rather, she must attempt to determine, via further reflection, whether it is the theory or the judgments that, all things considered, she finds more likely to be true, and then revise her beliefs accordingly. If the considered judgments that conflict with her provisional theory are very firmly held and seem to her to be central to her system of moral beliefs, and her reflections do not reveal that these judgments are involved in further conflicts, then it will be the theory that she will have to revise. But if the theory nicely accounts for all her most central and most confidently made considered judgments, and she finds it intuitively attractive on its own, and it seems to her to reveal the deep nature of moral obligation, and perhaps, in addition, the judgments with which this theory conflicts concern unusual cases, then it is the judgments that the philosopher will revise. And so we see that in her effort to construct a coherent system of moral judgments and theoretical principles that account for these judgments neither particular judgment nor theoretical principle is always favored. Whenever conflicts emerge, the philosopher must reflect on the connections among her beliefs and determine what to revise on the basis of what all things considered, seems to her most likely to be correct.

Even if the philosopher manages to bring her considered judgments and moral theory into a state of balance or equilibrium via such a process of mutual adjustment, her work will not be finished. The philosopher must seek an even wider equilibrium. She must also consider the connections between her moral beliefs and principles and the other sorts of beliefs, principles and theories she accepts or rejects. This process was in fact already begun when the philosopher filtered out initial moral beliefs that were formed in circumstances that she is confident entail a high risk of error. When she filtered out these judgments she was merely revising moral beliefs that flagrantly conflict with firmly although perhaps tacitly held epistemic principles. This process can be carried further. As the philosopher works out a more and more complete system of moral beliefs, she will obviously wish to see to it that this system of beliefs attains an appropriately high epistemic status. She must, in effect, see to it that her moral beliefs and the epistemic principles she accepts are coherent. In addition, in order to bring her epistemic principles to bear upon her moral beliefs, she will almost certainly also need to consider various beliefs about moral beliefs, such as, beliefs



about the circumstances in which her moral beliefs were formed, the factors that effect a person's moral judgments, the incompatible moral judgments made by others, and so on. As a result an even wider range of beliefs will be brought into play, some of which will likely lead to other areas of controversy. For example, consideration of the fact that other people make different moral judgments naturally leads to anthropological, sociological, and historical studies of different cultures and the debates about relativism and cultural diversity. Once again, however, the philosopher must seek a coherent system of belief by a process of *mutual adjustment*—neither moral nor epistemic beliefs nor any of the other beliefs that come into play are granted a privileged status.

Thus, for example, if a central, well-established, and intuitively plausible epistemic principle entails that some range of moral judgments are unjustified or irrational, then the philosopher will have to eliminate these moral judgments. But if a tentative epistemic principle yields a similar negative evaluation of central, firmly held moral judgments, then it will be the epistemic principles that will have to be revised. After all, the epistemic principles will themselves have been developed via a similar reflective process, and so will have been based in part upon considered judgments about whether beliefs formed in various circumstances would be rational or justified, and the philosopher may discover upon reflection that she is not as confident of these epistemic judgments as she is of the moral judgments with which, by way of epistemic principles, they conflict. Hence, whenever she encounters a conflict or incoherence within her system of beliefs, the philosopher must consider the conflicting beliefs, the logical and epistemic relations between these beliefs and the other propositions she accepts or rejects, and revise on the basis of what comes to seem likely to be true as a result of her reflections.

It is not too hard to see that there can be conflicts between moral beliefs and other types of beliefs as well. For example, metaphysical theories about the nature of persons or about what sorts of entities are ontologically respectable may well bear upon moral beliefs, as could various psychological theories, for example, regarding the nature of madness, or sociological theories, for example, about the role a moral theory must play in society. The important point is that the philosopher seeks to construct an ever more comprehensive system of beliefs and to bring these beliefs into equilibrium via a process of *mutual adjustment*.

(This description of the philosopher's method is quite a mouthful, but, to put things backwards, supposing my friend managed to digest it all, he would probably refuse to swallow it! The grounds for refusal are familiar.)

Friend: Now I really am confused, or perhaps I should say astonished. I suppose I shouldn't be surprised to discover that philosophers don't do anything, but think about things, that they don't pour over historical documents, *conduct extensive make observations of natural phenomena in the field,*

perform experiments, or worry over complex calculations. But this really is a bit much! If what you say is true, the accounts that philosophers end up advancing, in your example an account of morality—an account that you have grandiosely labeled a "theory"—apparently are entirely determined by nothing more than the philosopher's own intuitive judgments. Isn't it really quite a scandal that people who employ this frivolous method are paid good money to hold positions at the best colleges and universities, and that the results they obtain are published by legitimate journals and presses, and that generations of students have been forced to take courses taught by these people in which they must first buy and then study these books? If seems to me the philosophers have put one over on us. What excuse could there possibly be for investigating such obviously important matters as morality and the other topics addressed by philosophy in such a self-absorbed and self-indulgent manner? It's no wonder nobody cares, or far that matter knows, what philosophers have to say.

## Dead Ends

Before I try to answer the question my level headed interlocutor has posed, I am going to be somewhat high-handed in this section and describe, without much argument, how I think we can expect the method of reflective equilibrium to work. Three important claims about what we cannot expect the method of reflective equilibrium to do for us will emerge. If these claims are correct, they foreclose some of the more obvious routes one might want to take in defending this method of inquiry.

Let's begin with a description of what, in essence, the method of reflective equilibrium directs the philosopher to do: (i) to reflect upon her beliefs and the logical and evidential interconnections among her beliefs, (ii) to try to construct "theories" that are intuitively appealing on their own and that account for various categories of beliefs, for example, judgments about right and wrong, epistemic judgments, or judgments regarding what refers to what, and (iii) to resolve such conflicts, as are uncovered in the course of these reflections and efforts, at theory construction on the basis of what comes to seem most likely to be correct as a result of still further reflection. (For a more complete and systematic description of reflective equilibrium see Daniels 1979 or DePaul 1993: ch. 1.)<sup>1</sup> It is no news that since, as this description makes clear, the entire process is guided by nothing more than the inquirer's own beliefs, judgments, and what seems to the inquirer to be correct upon reflection, given enough screwy initial beliefs and unusual judgments about how to resolve conflicts, an inquirer could end up accepting just about anything in reflective equilibrium. Hence, my first and least controversial claim about what reflective equilibrium cannot do: (1) *The method of reflective equilibrium provides no guarantee that it will lead inquirers to true beliefs.*<sup>2</sup>

It is not, however, all that easy to accept any old screwy thing in reflective equilibrium. To do so one must be willing to make suitably screwy adjustments throughout one's entire system of beliefs.



very often find people who have a coherent but wacky system of beliefs. When we discover someone who holds some bizarre belief, what we usually find is that she also holds, or is disposed to accept, other more "normal" beliefs that we can use to persuade her to revise her strange belief. This is to say, what we usually find is that people who hold screwy beliefs have not reached a point of reflective equilibrium, and that they generally can be forced to revise their screwy beliefs in order to bring their beliefs into reflective equilibrium. Hence, one might hope to show that the method of reflective equilibrium is "reliable" even though it cannot guarantee the truth to every inquirer. However, although we admittedly do not often encounter people with coherent but totally bizarre systems of belief, we do not actually need to find such people to have good reason to suspect that the method of reflective equilibrium is not generally reliable. All we need is a sufficient amount of difference of (perfectly nonscrewy) opinion. It is safe to say that philosophers and other thinkers who have addressed similar questions throughout the ages have in fact employed something at least very much like the method of reflective equilibrium. Hence, I think we can safely take the various views about morality, God, how society should be structured, the nature of human persons, beauty, and so forth that have been propounded in one way or another by philosophers, religious and political leaders, novelists, poets, artists, and so on, to provide a fair indication of what sorts of views a person can accept in reflective equilibrium. And we find plenty of sufficiently great difference of opinion here for it to be a very safe bet that (2) *The method of reflective equilibrium will not even reliably lead inquirers to the truth.*<sup>3</sup>

One might, of course, say the same sorts of things about deductive arguments. Given strange enough premises, you can construct a deductively valid argument for any bizarre conclusion you pick. And even ignoring this sort of mere possibility of screwy conclusions being derived by deductive arguments, contrary views have actually been advanced by serious thinkers as the conclusions of deductive arguments often enough for one to argue, in the way I have regarding reflective equilibrium, that deductive arguments will not reliably lead all inquirers to the truth. But of course that is neither here nor there. The interesting thing about deductive arguments is that they are perfectly reliable *conditionally*, that is, given true premises, they yield true conclusions. Similarly, one might claim that the method of reflective equilibrium is to some degree conditionally reliable. Perhaps it is too much to expect that this method is perfectly conditionally reliable like deduction, but perhaps it might have the sort of conditional reliability we think inductive arguments have: given true premises the conclusions of such arguments are very probably true. Thus, while reflective equilibrium may not be able to guarantee that it will lead every inquirer to the truth, nor even that it will lead the majority of inquirers to the truth, it may be asserted that it will reliably lead the rights sorts of inquirers to the truth, that is, inquirers whose intuitive judgments, both in forming beliefs initially and about the resolution of conflicts, are true. I think this is true enough, and I suppose it is worth taking note of that fact. Unfortunately, it will not help answer the question posed above in the absence of some reason for thinking that *our* intuitive judgments are true. But if 'our' here refers to "us philosophers," our all too obvious differences of

would seem we would need some way of picking out those of us whose intuitions are true before we could say for whom the method of reflective equilibrium is reliable, and it obviously is not very likely we will manage to pull that off.

Maybe instead of talking about truth, we should try to answer the question above in terms of justification. But we had better first pin down just what 'justification' is supposed to indicate here. We can initially locate the concept of justification that has received the most attention from epistemologists by factoring it out of knowledge: *Knowledge is justified true belief plus (some feature designed to rule out Gettier problems); hence, justification, in this sense of the term, is whatever a true belief plus must have in order to count as knowledge.*<sup>4</sup> One influential class of theories seeks to account for justification in terms of reliability (see, Goldman 1979, 1986). Moreover, even very many of those philosophers who reject reliability theories still hold that justification must be truth conducive (Alston 1985, Bonjour 1985). If these philosophers are correct, the considerations adduced above seem to indicate that the method of reflective equilibrium will not lead all or even most philosophers to form beliefs that are justified. Of course, it is not universally agreed that justification is truth conducive; perhaps it is just a mistake to think that it is. If we forget about truth conduciveness and attend to the various theories of justification, one type of theory seems to be tailor-made for defending the method of reflective equilibrium, namely, coherence theories. However, the fit here is simply *too* good. No one with the concerns about reflective equilibrium we are out to address will be satisfied upon being told that reflective equilibrium is guaranteed to yield justified beliefs when in the sequel this claim is defended by appeal to a coherence theory of justification. Their concerns about reflective equilibrium will simply reemerge as familiar objections to coherence theories of justification. Finally, without getting involved in the various debates about the nature of epistemic justification, I think it is safe to say that even if justification is not truth conducive, it is in a certain sense "objective." One's beliefs must satisfy certain objective standards to count as justified. There are certain sorts of fallacious patterns of inference, and no matter what sort of rationalization or "justification" the inquirer might be able to construct for using them, these patterns of inference simply cannot yield justified belief. (The gambler's fallacy and hasty generalizations may provide examples of such fallacies.) But unfortunately the coherence constraints imposed by reflective equilibrium are not sufficient to guarantee that any inquirer employing the method will accept only correct epistemic standards. Hence, since reflective equilibrium does require an inquirer to live up to his or her own standards, that is, the standards that he or she accepts in reflective equilibrium, (3) *The method of reflective equilibrium cannot be counted on to yield justified beliefs.*<sup>5</sup>

## Rationality, Reflective Equilibrium, and Alternative Methods

(I keep hearing what my friend from the bar might say.)



this reflective equilibrium business? The idea was for you to explain why you're in favor of it.

The fundamental reason I am in favor of "this reflective equilibrium business" can be stated easily enough: Any other approach to philosophical inquiry is irrational. But this claim on behalf of reflective equilibrium obviously requires a lot of explanation and not a little justification before it will count for more than name calling. First, although rationality is still esteemed in most (although not all) circles, and it might even be generally (although not universally) agreed that philosophical inquiry *must* be rational, the term 'rationality' is used by different people in so many different senses that it would be nice if I explained how I am using it. Second, given that there are many conceptions of rationality floating about, I ought to explain why rationality, *as I conceive of it*, is an especially good thing. I might as well warn in advance that once you begin to understand how I think of rationality you will probably feel the need for an explanation of why rationality is valuable to be more pressing than you do now, when you are free to think of it in your own favorite way. Third, I obviously cannot simply claim that all alternatives to reflective equilibrium are irrational. I had better do something in the way of convincing you of this claim.

Having raised these three issues, I am going to explicitly focus on only one of them and hope my views regarding the other two become apparent along the way. I will argue that any method of philosophical inquiry that is an alternative to reflective equilibrium is irrational. If all goes well, along the way you will pick up a pretty good idea of the conception of rationality I am working with, even though I will not present an explicit account or analysis. Also, I hope that by portraying clearly exactly what one must do to deviate from reflective equilibrium and how such a deviation is irrational, it will become clear what is wrong with such irrationality, and perhaps begin to become clear why one might value the corresponding sort of rationality.

Friend Sol Are you going to explain why any alternative to reflective equilibrium must be irrational or not?

Well, think very abstractly of what the method of reflective equilibrium recommends and then ask yourself how a method of philosophical inquiry would have to look to be a genuine alternative to reflective equilibrium. From one perspective, reflective equilibrium seems to direct the inquirer simply to take her judgments about something, for example, knowledge, right and wrong, or the nature of belief, and attempt to construct a "theory" that accounts for these judgments. When one views reflective equilibrium from this perspective it is easy to conceive of alternative methods, and alternatives that are obviously preferable at that! But this perspective on reflective equilibrium does not afford us a fair view, for the method does not direct the inquirer simply to construct a theory that accounts for her initial intuitive judgments. Most centrally, the method directs the inquirer to do two things as she attempts to construct a philosophical theory:

(I) Reflect upon the logical and evidential relations that hold between her initial intuitive judgments and the other beliefs and theories she accepts, between these judgments and the emerging theory she is constructing to account for them, between this emerging theory and any relevant background beliefs or theories she accepts, and so on.

(II) Whenever these reflections uncover some sort of conflict or incoherence among beliefs, resolve the conflict by revising beliefs in the way that comes to seem most likely to be correct upon thorough reflection, that is, after taking into account everything she believes that might be relevant.

When one really focuses upon these two directives, it becomes rather more difficult to conceive of an alternative to reflective equilibrium, or more specifically, to conceive of a *rational* alternative to this method.

In order to constitute a real alternative to reflective equilibrium, a method would have to oppose reflective equilibrium with respect to one or the other of its two central directives. In order to do this, such a method must either (A) abandon reflection altogether, or (B) direct the inquirer to reflect, but to do so incompletely, that is, to leave certain beliefs, principles, theories, or what have you out of account, or (C) not allow the results of the inquirer's reflections to determine what the inquirer goes on to believe. I maintain that a method of philosophical inquiry having feature (A), (B), or (C) would be irrational. I will consider these in turn.

### Abandoning Reflection

I am not sure I really need to comment on (A), since it is just about a directive not to think, but here goes. A "method" of inquiry incorporating (A) would surely be a strange beast, indeed, so strange that it is doubtful whether it would constitute a possible method of philosophical inquiry at all. Such a "method" might direct the inquirer simply to believe whatever she happens to believe, without thinking things over at all. I am inclined to think that this already constitutes a good reason for calling the method "irrational." In addition, it is terribly improbable that an inquirer following such a method would end up having any very coherent or systematic view or accepting anything much like the sort of theoretical accounts philosophers seek.<sup>6</sup> It would seem the only "method" that would be likely to lead an inquirer to hold a theory or systematic account without doing any reflective thinking on her own would be one that baldly presents the account or theory and directs the inquirer to believe it without further ado. It is hard for me to imagine the sort of pessimism about one's own cognitive powers that would lead one to adopt such a method. To accept directive (A) one would have to give up entirely upon one's self and either accept whatever one happens to believe without giving it any thought or blindly submit to some sort of wholly external authority and accept what that authority dictates without giving it any thought.

It might seem that I am exaggerating here, but notice that to follow the sort of method we are imagining...



here embodied in the account or theory the "method" directs one to accept, *for absolutely no reason*. To have a reason for doing so one would have had to reflect at least enough to uncover the reason, and at least to that small extent have trusted the results of one's own reflection. This was what I was trying to indicate by calling the authority "wholly external." The inquirer is not here accepting the dictates of the authority because she has come to believe that, at least regarding certain matters, the authority is either reliable or more reliable than she is. There is obviously nothing unusual or irrational about accepting an authority in such a way. We all do it countless times, for example, when we accept a pathologist's identification of a tissue sample, a theoretical physicist's nontechnical explanation of a complex mathematical theory, or the result of a mathematical operation indicated by a pocket calculator. Not only is there nothing unusual or irrational about such acceptance of authority, it is perfectly compatible with the method of reflective equilibrium. To take a simple case, if upon complete reflection a person comes to be certain that some authority is perfectly reliable about a certain range of propositions, then unless the authority should endorse some proposition within that range that the person is certain is false, the way for the person to maintain coherence within her system of beliefs is to accept the dictates of the authority and revise her other beliefs accordingly.

Another problem with abandoning reflection is that a person who fails to reflect is liable, and I think virtually certain, to end up accepting what he himself does not really find acceptable. Here is what I mean. It very commonly happens that when we reflect upon something we believe, we uncover among the other things we believe, or come to believe as a result of our reflections, reasons for doubting or rejecting our initial beliefs, reasons which, all things considered, we are much more strongly committed to than the original belief, and which would therefore lead us to abandon the original belief. Thus, I say, a person who fails to reflect and goes along believing what he has always believed is liable to believe what he himself does not find most acceptable. The danger of believing what one does not find acceptable must be even greater for one who adopts the teachings of some outside authority without reflection. But no matter how one comes to do so, I think it is irrational to believe what one finds unacceptable, and a method of philosophical inquiry that directs inquirers to put themselves into such a position cannot be rational.

### Reflecting Incompletely

It might seem that methods of inquiry that incorporate (B), and thus direct merely *incomplete* reflection, do not entail the sort of irrationality—the pessimism, submission to external authority or danger of self-contradiction—I have claimed is involved when one abandons reflection. Indeed, methods of inquiry that leave out certain sorts of judgments are quite familiar and seem unobjectionable. We expect, for example, that evolutionary biologists or physicists studying cosmology will ignore any religious beliefs they might have during their scientific inquiries, and we surely would be somewhat disturbed to find that this expectation was not fulfilled. Another example might be provided by legal proceedings, where we expect decisions to be made only on the basis of what qualifies as

evidence according to the relevant legal standards, not on the basis of any beliefs a judge or juror might happen to have that bear upon the case.<sup>7</sup> Moreover, many of the criticisms of reflective equilibrium have focused on the use the method makes of intuitive judgments, so critics pretty clearly do not want us to abandon reflection entirely, but to reflect in a more limited way, a way that gives our intuitive judgments no weight. That critics do not wish us to abandon reflection altogether of course comes as no surprise—for it would be a very odd thing to spell out a set of reasons for not forming beliefs in the way that someone is inclined to form them, and present these reasons to that person for consideration, when what you want the person to stop doing is reflecting about her beliefs altogether!

Although there are many examples of acceptable methods of limited reflection, they are not really to the point. In order to be following a genuine alternative to reflective equilibrium the inquirer must not merely set aside or eliminate certain beliefs, or even whole classes of beliefs. To be employing an alternative method the inquirer must eliminate beliefs *without any ground for doing so*. If she had a reason for leaving some of her judgments out of account, for example, those judgments do not meet her own epistemic standards, then she would *not* simply be leaving them out of account. They *would* have been taken into account, have been reflected upon, and this reflection would have revealed these judgments to be in conflict with other more strongly held beliefs or principles, and this realization would in turn have led to the revision, and perhaps the rejection, of these judgments.<sup>8</sup> When we focus our attention on the relevant kind of incomplete reflection, the kind that really is incompatible with reflective equilibrium, it becomes apparent that it is irrational after all, that it involves the same sort of pessimism, submission to external authority, and high risk of self-contradiction we saw when considering the complete abandonment of reflection.

It is easy to see that a method of inquiry with feature (B) puts the inquirer at risk of accepting what she does not find acceptable, and thereby contradicting herself in a sense, in much the same way a method of inquiry incorporating feature (A) does: Some of the beliefs or theories that are left out of account might very well conflict with the system of belief the inquirer is led to accept by following his method of limited reflection. If this happens, and the inquirer is in fact more strongly committed to the beliefs that were left out of account and would remain so after duly considering the conflict and how best to resolve it, then it can hardly be rational for the inquirer to accept the system of beliefs the method of inquiry led him to, since this system is contradicted by other things he believes more firmly and would continue to find preferable if he were to consider the matter. I do not believe a philosophical method that puts an inquirer into such a position is rational.<sup>9</sup>

It may not be apparent why I think methods of limited reflection involve pessimism and submission to external authority. But what other explanation could there be for a person ignoring certain of her judgments, even when she is very strongly committed to them and has no reason of her own for doubting them? Such a person must either be alienated from the part of herself responsible for the judgments being excluded and have given up on this part of herself without having any reason for distrusting this part, or she is submitting to some method,



approach, or authority that directs her to ignore certain judgments of which she is confident, even though she cannot really believe this method, approach, or authority will lead her to the truth.

I think the point I am trying to make is important enough that I am going to risk laboring it. I want to focus on the aspect of submission to external authority by considering what must be happening when a person sets out to criticize the method of reflective equilibrium. The inquirer employing reflective equilibrium has started out with a set of initial judgments about some area, and through a process of mutual adjustment constructed a theory that accounts for these judgments. The inquirer will also have reflected on the connections that hold between this system of theory and corresponding judgments and any other beliefs that might be relevant, and once again brought her beliefs into a stable equilibrium via a process of mutual adjustment. A critic must advance some sort of argument against some element of the resultant system of belief. To pick a pertinent example, the critic might cite studies by cognitive psychologists showing that our intuitive judgments about the area in question are unreliable, and press the inquirer not to allow these intuitive judgments any weight at all in determining the theory she ends up accepting. In all likelihood, what the critic is thinking is that the inquirer is ignorant of the relevant studies and that she accepts many of the same forms of argument and background epistemological views that he does, so that when she is made aware of the studies and the implications of these studies for her considered judgments and the theories in part supported by these judgments, she will accept the studies and implications just as the critic has.

If this is what the critic is doing, then he really is not doing anything that conflicts with the method of reflective equilibrium. He is merely providing the inquirer with information acquired in ways the inquirer accepts which conflicts with some of what the inquirer believes and expecting the inquirer to make appropriate adjustments so that her system of beliefs is again coherent. This is nothing more nor less than what reflective equilibrium dictates.

But suppose that the inquirer is not ignorant of the psychological studies and that she has already incorporated her belief in the results of these studies into her system of beliefs in a way that does not require her to give her intuitive judgments no weight in her deliberations. To do so consistently, she obviously must differ with the critic somewhere else, for example, with respect to some epistemic principle, rule of inference, or judgments about what interpretation of data is most plausible. But if her beliefs are indeed in a state of reflective equilibrium, she will have considered the opinions about which she does not agree with her critic, and there will be a coherent story to tell in support of her own views that the inquirer finds most acceptable upon thorough reflection. This is the case we must consider to get an alternative to reflective equilibrium, and to get the alternative we must imagine the critic still demands that the inquirer agree with him in this case. What would bring the inquirer to do such a thing? She would have to abandon the results of her own reflection, give up, at least in part, on thinking for herself, and simply knuckle even though she firmly believes, after careful reflection on all the relevant considerations, that doing so will lead her away

such a thing. Indeed, is this not one paradigm of what we consider to be irrational? Is it not exactly what we think happens when inquisitions and pogroms succeed most thoroughly, rather than merely in provoking cynical head nodding?

Perhaps it is unfair to imagine the critic *demanding* agreement—we have no pogroms or inquisitions. Critics simply present their arguments, aggressively perhaps, but with the sort of tame linguistic aggression one finds in graduate seminars, discussions at conferences, and in the pages of journals. What, in such polite society, would bring the inquirer to abandon the results of her inquiry? It is naive to think the answer is not pretty much the same. She would still have to give up and knuckle under even though thorough reflection had led her to believe this means accepting what is false. She may not be doing this because she has been shown “the instruments” or to avoid the ghetto. But she must be motivated by considerations she does not regard as having any more relation to attaining the truth, such as, by a desire to fit in, have others to talk with, earn degrees, publish, achieve a professional reputation, and so on.

### Not Believing What Seems Most Likely to be True

A philosophical method that does not allow the results of a person's reflection to determine what she accepts, as in (C), is perhaps most obviously irrational. Such a method would have to direct the inquirer to reflect, but after she had completed her reflections, to believe something other than what these reflections had led her to consider most likely to be correct. And it surely could not be rational, if it is so much as possible, for a person who had fully reflected upon some conflict among her beliefs to believe what these reflections have convinced her is mistaken and to leave off believing what her reflections have convinced her is correct. An inquirer who did such a thing would have given up on herself as an intellectual being and not merely risked believing what she herself does not find acceptable, but in fact have accepted what she herself does not find acceptable. In addition, it seems that, when it comes down to the final resolution of conflicts among beliefs, such an inquirer must be submitting to some sort of external or alien authority, since she goes with beliefs that she herself does not consider most likely to be correct. And so we see that a method that does not allow the theory the inquirer accepts to be determined by the results of her reflection is irrational as well. (See Onora O'Neill 1992 for a useful discussion of how submission to alien authority figures in Kant's attempt to vindicate reason.)

### Why Give a Damn About Being “Rational”?

Friend: OK. I see why you say that any alternative to reflective equilibrium is irrational. I might even cut you some slack and grant that reflective equilibrium is the only rational method of philosophical inquiry, since I think I see how the additional argument might go. But so what? You say it would be irrational for a person to leave some of her beliefs out of account, or to believe something that conflicts with what she is



abstract. Particularly a person who believes one thing when her own reflections have led her to be more strongly committed to something else seems to be contradicting herself somehow. But what if some of this person's beliefs are nothing but crazy superstitions? What if she is paranoid or obsessive and they are the result of some sort of chemical imbalance in her brain? Couldn't we fairly say that it would be irrational to take such beliefs into account? And wouldn't it be even worse if the person were very strongly committed to these superstitions or delusions, and revised other beliefs to conform with them? You say that to knuckle under and go along without reasons for doing so drawn from one's own system of belief is to behave in the same way as the most pathetic victims of inquiries. But I could just as well point out that to stick with one's own superstitions or delusions against the testimony of the rest of the world and continue making whatever adjustments are necessary to maintain a coherent system of belief is to follow the path of the dogmatist or lunatic. I would say that this is irrational, but you say just the opposite. There is not much sense haggling over who gets to keep a word, so you win. Let's say 'rationality' refers to the sort of self-consistency you seem to have in mind. What I want to ask about is the value of this subjective type of rationality. Philosophers once thought that if they proceeded rationally they would be led to knowledge and truth, or at the very least to justified beliefs. But you've already admitted this doesn't hold for your sense of rationality. So why should we care about proceeding rationally in this sense? What's so bad about being irrational?

Me: I've tried to portray very clearly what a person must do in order to follow a method other than reflective equilibrium. I wanted to try to get you to form a clear picture of what one is involved in believing irrationally in the way I have described. If you have this clearly in mind, if you really see what it is that one would be doing, . . . Wasn't it Louise Armstrong who said "If you have to ask, you ain't never gonna get it." The answer should be intuitively obvious.

(This ending obviously is not entirely fair. Insofar as the question concerns merely the disvalue of the sort of irrationality I have described, I do not think it terribly unfair. There is of course more to be said. One might ask, for example, whether the disvalue of irrational belief is best understood in terms of the violation of an obligation or whether it is more a matter of failing to attain something desirable or valuable. And if it is best understood in terms of obligations, one must ask after the ground of the obligation and whether it is a moral obligation or an obligation of some other sort. Or one might wonder whether irrational belief is something to feel guilty about, ashamed of, or regret for. But these are further questions. It might be lazy of me to begin a discussion that raises them and then let them drop, but I do not think it unfair. If the question really just is whether irrational belief of the sort I have described is a bad thing, then I do not

actly what such a belief involves and still cannot just see that it is a bad thing. What makes my ending unfair is that although my friend asks what is so bad about irrational belief, the way he raises this question, by pointing to the possibility that rational belief might come into conflict with other values such as believing the truth or avoiding dogmatism or even not being a little bit insane, suggests that what he really wants to raise is a comparative question. He wants to be convinced not that irrational belief is a bad thing, but that it is more important than other relevant values, for example, the value of believing the truth or, for that matter, the value of being a part of an intellectual community. He wants a reason for thinking that, when faced with the possibility that in attaining the one value we will lose the others, we should act to guarantee that we believe rationally and hope that the cost of doing this is not that we miss out on other good things. For my claim has been not merely that irrationality is a bad thing, and rationality a good thing, but that this is the value around which philosophical inquiry should be structured. Particularly when by making this claim I go against the dominant intellectual tradition that sees truth as the value that structures inquiry in general, philosophy in particular, and as the fundamental concept in terms of which all epistemic concepts (such as rationality) are to be defined, it does seem unfair to sidestep this issue. My excuse is that it is too large an issue to address within the confines of this chapter, and the truth is that I am not at all sure what to say.)

## Notes

I began working on an earlier incarnation of this chapter while in Germany with a grant from the Alexander von Humboldt-Stiftung. I am grateful to the AvH for their support and to Prof. Andreas Kemmerling, my host at the Seminar für Philosophie, Logik, und Wissenschaftstheorie, Ludwig Maximilian Universität-München. I had the opportunity to present the earlier incarnation of this chapter at the Ludwig Maximilian Universität, Universität Konstanz, University of Reading, University of Kent, and the University of Notre Dame, and I received many helpful comments from members of the audience on each occasion. I also had an opportunity to present this chapter in more or less its current form to a Summer Seminar organized by the Dutch Research School for Advanced Studies in Ethics, where the comments from the audience were once again very helpful. Finally, I have benefited from detailed comments by Marian David on both of the earlier versions of this chapter.

1. Because it does not play an important role in the argument I plan to offer here, I have neglected one very significant element of reflective equilibrium. This is the consideration of alternative theories available to one that Rawls stresses when distinguishing wide from narrow reflective equilibrium. I give this element thorough consideration in DePaul 1993.

2. (1) should be interpreted to mean: it is not the case that every inquirer who employs the method of reflective equilibrium will be led to form true beliefs. It is consistent with (1), so interpreted, that some inquirers are such that if they were to employ the method of reflective equilibrium they would be led to form true beliefs. The method might be said to guarantee truth to such special inquirers.

Obviously I here assume that truth is not properly defined as what a person would accept in some ideal state which ends the process.



equilibrium. Such a definition gets us a guarantee of truth, but on the cheap.

3. There are a number of plausible ways of interpreting (2). Perhaps it is most natural to interpret it to mean that each inquirer is such that if he or she were to employ the method of reflective equilibrium, he or she would not be led to form beliefs that are mostly true. However, this is not what I mean by (2). Rather, I mean that it is not the case that the vast majority of inquirers are such that if they employ the method of reflective equilibrium, then they will be led to form true (or mostly true) beliefs. It is compatible with this interpretation, and with the considerations on the basis of which I made claim (2), that some inquirers are such that if they employ the method of reflective equilibrium, they will be led to form true (or mostly true) beliefs. Some may take the considerations I cited to support the stronger, and perhaps more natural reading of (2), but I here make no such claim. If half of those who employ a method of inquiry are led to form one belief, and the rest of those who employ the method are led to form some incompatible belief, then we can be just about certain that the method is at best leading 50 percent of those who use it to the truth. Suppose it is the case that 50 percent of those employing the method end up accepting the truth. What is going on? It might be that there are not any relevant differences among the inquirers employing the method, so that some feature of the method is responsible for the outcome. In this case we might be able to conclude that any inquirer who employs the method will have only a 50 percent chance of coming to form a true belief as a result. But it might also be the case that there is some relevant difference among inquirers. Some feature that only 50 percent of inquirers have, and that the method will invariably lead inquirers with this feature to the truth. I have said nothing that might decide between these alternatives, so I wish (2) to be interpreted in a way that is compatible with both.

I should mention another possible, but perhaps less likely, confusion. I intend the set of "beliefs a person following the method of reflective equilibrium is led to form" to contain only those beliefs formed as a result of employing this method of inquiry. I assume the vast majority of the beliefs of the vast majority of people are ordinary perceptual, memory, introspective, and testimonial beliefs. I also assume that the vast majority of these beliefs are true. Most people who employ the method of reflective equilibrium will retain these beliefs throughout the process, but I do not think this is relevant to the reliability of this method of inquiry.

4. Plantinga (1993) identifies the epistemic concept he calls 'warrant' in this way, reserving the term 'justification' for a more deontological concept of epistemic evaluation. I have chosen to stick with 'justification' simply because that is the more commonly used term.

5. After my remarks about (1) and (2) it should be clear that I intend (3) to be interpreted as the claim that it is not the case that all or even the vast majority of inquirers who employ the method of reflective equilibrium will be led to form beliefs that are justified. For a more thorough presentation of my reasons for thinking that reflective equilibrium can guarantee neither truth, reliability or justification see DePaul 1993: chs. 1 & 2.

6. I have in mind here explicit theories and self-consciously systematic accounts. It is of course a common practice to view ordinary cognizers as having implicit theories or systematic views about various matters in order to explain certain aspects of the cognizer's behavior. Surely the most familiar example of this approach is provided by the Chomskian explanation of the ability of native speakers to recognize grammatical sentences of their own language.

7. The legal example is not entirely happy, since I think we do not really believe that the judge or juror should base her *belief* only upon the admissible evidence. Her

the evidence she has.

8. In another kind of case, perhaps this is what happens with science and religion, when either the person of an exclusively scientific temperament rejects religion or the devout religious believer rejects science: The inquirer is systematically more confident of judgments in one area than of judgments in a potentially relevant area. So the inquirer decides to work out her views in the area where she is most confident first, and only then consider how her judgments in the other area relate, consistently revising the judgments in the second area so they conform to the judgments in the first. Once again, although at a certain stage of inquiry it might seem that some judgments are simply being left out of account, and hence that some alternative to reflective equilibrium is being employed, if we consider the person's practice more broadly, we can see that this practice does indeed qualify as a sort of reflective equilibrium.

9. I should say that what worries me here is perhaps not best captured in terms of "risk." I expect any method of inquiry that is not Cartesian—moving from indubitable premises by indubitable steps—will put the inquirer at risk of error, perhaps even the error of self-contradiction. Of course, I do believe that the risk involved in methods having feature (A) or (B) represents much more than an abstract possibility. As I noted above, I think we can pretty much count on methods having feature (A) leading inquirers to accept things they do not find acceptable, and in practice I think the probability of a method having feature (B) leading to such a result is nearly as great. After all, if it were not very likely that following a method of limited reflection would lead the inquirer to different conclusions, there would not be much point in following such a method rather than a method of complete reflection, and one would not expect those who argue for the exclusion of one or another class of judgments to be so vociferous. But although I cannot quite put my finger on it, I think there is something more that is worrisome here than the fact that there is a very real possibility or even a probability that methods having feature (A) or (B) will lead inquirers astray, something about the way such methods, particularly those having feature (B), march inquirer's into trouble without acknowledging that there is any trouble.