LEARNING FROM ONE'S MISTAKES: EPISTEMIC MODesty AND THE NATURE OF BELIEF

BY

SIMON J. EVNINE

Abstract: I argue that it is not ideally rational to believe that some of one's current beliefs are false, despite the impressive inductive evidence concerning others and our former selves. One's own current beliefs represent a commitment which would be undermined by taking some of them to be false. The nature of this commitment is examined in the light of Nagel's distinction between subjective and objective points of view. Finally, I suggest how we might acknowledge our fallibility consistently with this special attitude to our own beliefs.

Human beings are fallible creatures and most of us agree that, if we are rational, we should acknowledge this fallibility in some way. That rationality demands of us some acknowledgement of our fallibility I call the Requirement of Epistemic Modesty (REM). But how should we fulfill REM? Most philosophers (indeed, I suspect, most people) would answer that each person should believe (at least) that

(RF) Some of one's own current beliefs are false.¹

(I call this belief RF for Reflexive Falsity.) In this paper, I shall argue that notwithstanding its obvious appeal, RF is not an ideally rational belief to have. There may be circumstances in which it is rational, but these circumstances go beyond those that impose REM on us. I shall further suggest that a powerful argument in favor of the rationality of RF depends on a significant misunderstanding of the nature of belief. Finally, I shall attempt to provide another way of fulfilling REM that does not embody this misunderstanding.
I begin, then, with RF and a simple but powerful argument in its favor. The argument, which I call the Inductive Argument, is this. Let us characterize the notion of a belief set, the entire set of a person’s beliefs at a given time. Then I contend that all belief sets associated with other people than ourselves are such that we take them to include some false beliefs. (This, of course, is an empirical claim and might be questioned, but since I will be arguing against the rationality of RF, I am happy to grant this premise for the sake of argument.) Not only do we take there to be false beliefs in everyone else’s belief sets. If we look at our own past belief sets, they too are all such that we now take them to have included some false beliefs. There is, thus, a vast range of cases in which we take belief sets to include some false beliefs. Given this huge body of evidence concerning the presence of false beliefs in belief sets, it would seem a tiny and uncontroversial step to conclude that our own current belief set likewise contains some false beliefs.

One might say that the Inductive Argument represents the policy of learning from one’s mistakes (and other people’s) in its most abstract form. For it is not the particular mistakes that we and others have made from which we are supposed to learn, but the mere fact of our and others’ having made mistakes at all that is supposed to be instructive.

Despite the considerable suasive power of the Inductive Argument, I believe that RF is not the best way to satisfy REM. In order to support this, I will argue that RF cannot be a member of an entirely rational belief set, i.e. a belief set none of the members of which is irrational. In order more efficiently to conduct the discussion, I shall reformulate RF as a belief about beliefs in a belief set. So RF will now be amended to:

(RF) Some of the beliefs in the belief set of which this belief is a member are false.

The basic idea behind my argument against RF can be best appreciated if we first confine ourselves to finite belief sets. Take some finite belief set, $B_n$, of which RF is a member. Suppose that RF is rational in $B_n$. Presumably, there is no other specific belief, $p$, in $B_n$ in virtue of which RF is rational. If there is, then it must be irrational to believe $p$, and hence $B_n$ contains an irrational belief. If not, then if RF is rational in $B_n$, it must be rational in $B_{n-1}$, where $B_{n-1}$ is formed by dropping any belief, other than
RF, from \( B_n \). By repeating this reasoning, we will be forced to say that RF is rational in a belief set, \( B \), of which it is the only member. But this is absurd. We cannot claim that RF would be rational in such a belief set since it is clear that in that case it would be true if and only if it were false. Of course, this irrationality is not something of which the person whose beliefs are represented by \( B \) could be aware, since by hypothesis, RF is that person's only belief. Nonetheless, we are clearly in a position to see the self-defeating nature of RF in such circumstances. Thus, either RF is irrational in \( B_n \), or there is some other belief in \( B_n \) which is irrational. In either case, \( B_n \) is not an ideally rational belief set.

The same argument can be run in reverse and its scope extended to countably infinite belief sets. First we consider the case in which RF is the only belief in a belief set \( B \). In that case, we cannot judge that RF is rational in \( B \), since we can see that it would be true if and only if false. Now consider a belief set \( B_n \) in which RF is irrational. What about the belief set \( B_{n+1} \) formed by adding some belief \( p \) to \( B_n \)? If \( p \) is rational, then its addition to \( B_n \) can do nothing to make RF rational if it was not rational already. On the other hand, RF could be rational in \( B_{n+1} \) if it was irrational to believe \( p \). But in that case \( B_{n+1} \) would contain a belief, \( p \), that was irrational. In either case, \( B_{n+1} \) is not ideally rational.

It is important to note that this argument does not show that RF is irrational simpliciter (though I shall sometimes speak as if it does in what follows, for ease of exposition). RF may be rational if it is accompanied by other beliefs which one takes to be irrational. However, the fact that it cannot be part of an ideally rational belief set is sufficient to show that it cannot be the best way to fulfil REM. REM is a requirement upon us because of our fallibility. But fallibility concerns the possibility of forming false beliefs, not irrational beliefs. Even if we were perfectly rational we would still be fallible and hence subject to the obligations of REM. Satisfaction of REM should, therefore, not be predicated on our irrationality.

A couple of comments on this argument are in order. First, it must be admitted that the idea of a creature with only one belief is absurd. (Let alone that that one belief should be a second-order semantic belief such as RF.) This absurdity makes itself felt in the difficulty of saying exactly why RF would be irrational if it were someone's only belief. As I indicated, we cannot explain that irrationality in terms of what could be recognized by someone whose only belief RF was. Such a person would not have the beliefs from which to derive that RF, in the circumstances, is true if and only if false. Nor can we appeal to what might be demonstrated by some reasonable expansion of that belief set, since in such an expansion, RF would no longer have the paradoxical quality of being true if and only if false. Nonetheless, given the absurdity of a creature having only one belief, I think we can recognize, from our perspective, a further absurdity, or
irrationality, in that one belief’s being RF. A more thorough treatment would require an examination of the difference between irrationality from the point of view of the believer and irrationality from some point of view external to the believer. But I think the intuitive judgement of RF’s irrationality in the case in which it is someone’s only belief is strong enough to render a lengthier discussion unnecessary here.

Secondly, I want briefly to consider a variant of RF designed to avoid the problems I have been raising:

(RRF) Some of the beliefs in the belief set of which this belief is a member, other than this belief itself, are false.7

(I call this RRF for Restricted Reflexive Falsity.) The variant fares no better than the original version. I will consider only the additive version of the argument. Since RRF cannot be non-trivially true in a belief set of which it is the only member,8 we start this time with a belief set, B, of two members of which RRF is one and any other belief, p, is the other. In that case, RRF can be rational only if it is rational to believe that p is false. So either p is irrational or RRF is. In either case, B is not ideally rational. Now add another belief, q, to obtain B1. This time there are two cases to consider. If RRF was irrational in B, then the addition of q can only make RRF rational in B1 if q is irrational (or if its addition makes p irrational). If RRF was rational in B (and it was p that was irrational) then the addition of q can either make p rational or not. If it does, then RRF will remain rational in B1 only if q is irrational. If it does not, then p is irrational in B1. In all cases, B1 fails to be ideally rational. Repeat as necessary.

Notice that the ‘base steps’ in the arguments against RF and RRF correspond in a certain way to the Liar Paradox and Moore’s Paradox respectively. In the case of the argument against RF, the claim is that it cannot be rational for one’s only belief to be RF, for in that case, RF could be true if and only if it were false. Although RF does not refer to itself as a false belief (as the Liar sentence does), the effect of such self-reference is produced by the absence of any other belief from the domain of its quantifier. We thus have a kind of extensional equivalent of the situation in the Liar Paradox. The claim of RF’s irrationality in the case of a belief set of one belief is the epistemic side of this extensional equivalent.

With the argument against RRF, the base step concerned the situation in which there are two beliefs, one of which is that the other is false. This resembles the situation of the utterer of Moore’s Paradox, who says “I believe that p, but not-p.” Again, we are dealing with a case in which the restricted domain of quantification produces an extensional equivalent of the self-defeating epistemic situation represented by Moore’s Paradox.
3.

I shall now consider three objections that might be raised against the argument of the previous section. The first challenge I will consider stems from the view that the epistemology of belief is super-imposed on an epistemology of degrees of belief. This is a widely-held, though by no means uncontested, position in epistemology which, following Richard Foley, I shall call the Lockean thesis. The Lockean thesis holds that we countenance both beliefs and degrees of belief, and see the former as derivative from the latter: “belief-talk is a simple way of categorizing degrees of confidence [i.e. degrees of belief] to say that you believe a proposition is just to say that you are sufficiently confident of its truth for your attitude to be one of belief.” Degrees of belief can be measured by real numbers in the unit interval. A degree of belief of 1 represents absolute subjective certainty in a proposition; a degree of 0 represents absolute certainty in the negation of the proposition. Given a threshold value $r$, we can then say that a person believes $p$ if and only if her degree of belief in $p$ is greater than or equal to $r$. (In order to ensure that a set of beliefs does not include contradictory beliefs, $r$ must be greater than 0.5.) The Lockean thesis can be, though it need not be (and is not by Foley), supplemented with the further claim that degrees of belief do, or more likely should, conform to the probability calculus. In that case, they can be interpreted as subjective probabilities. Since this further claim provides the most explicit and comprehensive theory of degrees of belief I will, for simplicity’s sake, assume it in what follows. Nothing that I say depends on taking the Lockean thesis in this way, however.

If one accepts the Lockean thesis, then my argument against RF will no longer go through. To see this, consider a belief set that does not include RF. According to the Lockean thesis, this will consist in a number of propositions to which subjective probabilities equal to or greater than the believer’s threshold value $r$ have been assigned. If the person now has good reason, say on the basis of the Inductive Argument, for adding RF to her belief set, this will naturally affect the subjective probabilities she assigns to the propositions in her belief set. To put it simply, it should lower her confidence in each of those propositions. Now if RF were, for example, accompanied by a belief about which particular beliefs were false, then it would lower that person’s confidence in those particular beliefs to 0, and that would mean that they would drop out of her belief set. But if RF were not directed at any specific beliefs, then its effect on the confidence a person had in each of the propositions she believed would be distributed over all her beliefs. Since the effect would be distributed, the larger the belief set, the less the degree of each individual belief would be diminished if RF were added to the belief set. And if a belief set were large enough, then it is quite possible that few, or even none, of the beliefs
in it would be sufficiently affected to drop them below the threshold level for belief. In that case, one could rationally believe that some of one's beliefs were false while still rationally having all those beliefs. One could sum up the point in this way. My argument against RF depended on there being no correlation between the number of beliefs one has and their rationality. What the Lockean thesis contends is that there is such a correlation. RF may be irrational in a small belief set, but there will be a point in the expansion of a small belief set when it becomes rational simply because of the increase in the number of beliefs that are part of the set.\textsuperscript{12}

In order to rebut this response to my argument against RF it would be necessary to challenge the Lockean thesis. In fact, I believe the Lockean thesis to be mistaken, but arguing so would vastly exceed the scope of this paper. For now, I am content simply to point out that, supposing there are no other convincing objections to my argument, rejecting my conclusion would commit one to this distinctive and not uncontroversial approach to epistemology.\textsuperscript{13}

Before I leave this topic, though, I do wish to raise one concern for the Lockean thesis that arises in the current context. How much our confidence in our various beliefs should be lessened, if we have reason to believe RF, should, according to the Lockean, partly be determined by how many of our beliefs we have reason to believe are false. If we have reason to believe that only one is false, then each individual belief in a large belief set will not be much affected. If we have reason to believe they are all false, then each individual belief will have to be given up. So the question arises, for the supporter of RF, of how many beliefs we do have reason to believe are false. I offered the Inductive Argument as support for RF. If we continue to see the Inductive Argument as offering an evidence base of the size I suggested (everyone else's and all our former belief sets), then I assume the reasonable conclusion to draw is the minimal one that \textit{at least one} of one's beliefs is false. And, given the Lockean thesis, it is correct that adding this belief to a large belief set will not have a serious effect on the subjective probabilities of the other beliefs. But now consider a variant of the Inductive Argument in which we confine ourselves to belief sets in which we see not just some, but massive, error. In that case we will have an evidence base which, although much smaller than that of the original version of the argument, will be large enough to warrant some degree of confidence in the conclusion that a massive number of our beliefs are false. And the effect that this would have on the confidence with which we held our other beliefs might be quite substantial. Of course, we can say nothing conclusive without assigning precise values, or ranges of values, to the various parameters, and then doing the math. But underneath the purely quantitative considerations here there is the simple point that, even given the Lockean thesis, the Inductive Argument risks being the first step on a slippery slope that ends with the admission
that it may be rational for us to believe that we are massively in error. This, I take it, is something that even the Lockean ought to be concerned about.\textsuperscript{14}

A second challenge to my objection to RF also seeks to show that the rationality of RF is dependent on the size of the belief set of which it is a member, but in a way that is independent of the Lockean thesis. It might be pointed out that my argument against RF had an extreme artificiality to it. Belief sets are not built up one belief at a time, with each belief having its credentials scrutinized before it is allowed in. We typically inherit large portions of our belief sets, and then proceed, in highly unsystematic ways, to demolish parts of them and annex whole new suburbs. Given this reality, one might feel that a belief like RF is quite rational. The very messiness and disorder of our belief sets is what makes it rational. This fact is precisely what is overlooked in my argument against RF.\textsuperscript{15}

It should be remembered that the conclusion of my argument against RF was not that RF was itself irrational, but that it could not be part of an ideally rational belief system. Either it must be irrational, or some other beliefs in a belief set containing it must be irrational. I went on to suggest that RF was still inadequate as a way of fulfilling REM because our fallibility should not be predicated on our irrationality. Even if our belief systems were more orderly than they are, perhaps built up by considering each belief individually, we would still need to satisfy REM. If the current objection is simply reminding us that typical belief systems do contain irrationalities, and that given this, there need be nothing irrational about holding RF, then I heartily agree with it. But of course, seen in that light, it is no objection to my argument against RF. Alternatively, it might be held that the disorder of a belief set, that which justifies the inclusion of RF, is not to be equated with, or is not even supervenient on, the irrationality of some of its members. Disorder is a purely systemic property that can obtain even if each of the individual members is quite rational. But in that case, I cannot see how the disorder of a belief set would justify the inclusion in it of RF. RF is a second-order belief about beliefs taken individually, not a belief about a systemic property of a belief set as a whole.

A third and final objection to my argument against RF is this.\textsuperscript{16} It might be said that whatever else the Inductive Argument shows, it surely shows that there is a high objective probability that one has some false beliefs. In attacking RF, I am holding that one should not believe that one has some false beliefs. Putting these two claims together, however, seems to imply a violation of David Lewis's Principal Principle, which holds, roughly speaking, that our beliefs should match our assessments of objective probabilities. Specifically, if we believe the probability of p's being true is n, we should accord a subjective probability of degree n to p.\textsuperscript{17} So, in the present case, if the Inductive Argument shows that there is


\textsuperscript{15} This is what I call the "messy" argument against RF.


a high probability that one of our beliefs is false, then we should have a correspondingly high degree of belief that one of our beliefs is false.

The immediate and narrow response to this objection is constrained by the position I have taken on the Lockeian thesis. The Principal Principle is framed in terms of degrees of belief. My position, that it is not ideally rational to believe that some of one's beliefs are false, is framed in terms of belief *simpliciter*. The Lockeian thesis connects the two and implies, given the Principal Principle, that if I take the objective probability that some of my beliefs are false to be high enough, I should believe *simpliciter* that some of my beliefs are false. But I have already stated my position (though I have not argued for it) that the Lockeian thesis should be rejected. Without the Lockeian thesis, it remains open to me to agree with the deliverance of the Principal Principle and accord a high degree of belief to the proposition that some of my beliefs are false, without believing that proposition.18

Nonetheless, despite the availability of this response, one might feel inclined to endorse an amendment of Lewis’s Principal Principle that would directly connect objective probability and belief, by-passing degrees of belief and the Lockeian thesis. So amended, the Principle might go: “believe things that you think are highly objectively probable; don’t believe things you think are highly objectively improbable.” Such an amended version has considerable *prima facie* plausibility. Even if the divorce of belief and degrees of belief makes it logically consistent not to believe something to which one accords a very high degree of belief (and hence, according to the Principal Principle, to which one accords a very high objective probability), there is surely some rational connection between belief *simpliciter* and our judgements of objective probability. The question arises, then, does my position conflict with the amended version of the Principal Principle, and if so, is that a problem? I have a two part response to this question.

First, I dispute that the Inductive Argument does succeed in showing that there is high probability to the proposition that some of one’s beliefs are false. It would do this only if the evidential base were relevant to the conclusion. But, as I shall explain in the following sections, the Inductive Argument is flawed precisely because it overlooks a certain lack of uniformity between our own current beliefs and the beliefs of others (including our former selves). Secondly, even if the Inductive Argument did show that there was a high probability that RF is true, I think the amended version of the Principal Principle does not adequately capture the highly nuanced relations between probability and belief. In particular, I do not accept that it is rational to believe any proposition to which we accord a high probability. For example, there is a high probability that a given ticket in a large, fair lottery will lose, but I do not think that it is rational to believe that the ticket will lose. This is a controversial
view on a complex issue. Let me just say here that in holding that a proposition's high probability is not always a sufficient condition for believing it, I do not deny that it sometimes may be. In the lottery case, what leads me to say that the high probability that a given ticket will lose is not sufficient to believe that it will lose is that that proposition is one of a large number of similar propositions among which there is good reason to believe that one is false but no reason to prefer any particular one.¹⁹

Furthermore, note that even if we grant that the Inductive Argument does establish that RF has a high probability of being true, I am counseling one not to believe in something which is highly probable. This violates one half of my amendment to Lewis's Principal Principle, which tells us to believe things that we think are highly objectively probable. But nothing I say violates the other half, that one should not believe things that one thinks have an objectively low probability. In other words, while I do say that one should not believe one has false beliefs, I do not claim that one should believe that one has no false beliefs, or that everything one believes is true. These beliefs would clearly not be supported by my case against RF.²⁰

₄.

I have, so far, defended my rejection of the conclusion of the Inductive Argument. If I reject its conclusion, where do I think that the argument has gone wrong? The Inductive Argument attempts to justify a certain attitude to our own current beliefs on the basis of our attitudes to our past beliefs and the beliefs of others. I shall argue that this is just where the argument goes wrong. For there are significant ways in which the fact that some beliefs are our current beliefs precludes us from taking certain attitudes to them that would be quite unproblematic in the case that they were someone else's beliefs. It is, of course, not the content of our beliefs that makes them special to us, but merely the fact that they are ours.

One manifestation of the special status we accord to our own current beliefs occurs in the context of explaining beliefs. We find it quite in order to explain another person's belief by pointing out a cause of it that is not a reason for it. So, we might say knowingly, "Jody only believes the economy is doing poorly because she wants her marriage to fail" (her husband is a booster economist). But this kind of explanation is not something we can do unproblematically for our own beliefs. If Jody were to come to believe that she only believed that the economy was doing poorly because she wanted her marriage to fail, then certainly, if she were rational, she would thereby relinquish her belief about the economy. There are even problems, I would argue, in coming to believe that among several causes of one's belief that p are some that are not reasons for p.
Clearly, if among the causes recognized by the person of her belief that \( p \), there are some that are adequate reasons for \( p \), then there is no problem about her continuing to believe that \( p \) after she acquires the belief that among her causes for believing \( p \) are some that are not reasons for \( p \). But she must either disengage her belief that \( p \) from those non-reasons, or recognize the fact that she would continue to believe that \( p \) even if she did not have adequate reasons for it. The latter is something that approximates the oddity of continuing to believe \( p \) in the face of the belief that the only cause of one’s believing \( p \) is not a reason for it. The former course, of disengaging one’s belief from its non-rational causes, would render false the belief that the belief that \( p \) has non-rational causes. Therefore, in either case, one cannot rationally continue to believe of oneself that one has a given belief with some non-rational causes.

Returning to the main topic of this paper, the way in which our attitude to our current beliefs affects the cogency of the Inductive Argument is that our current beliefs represent a commitment; in believing something, we are committed to its truth. What commits us to the truth of our beliefs is nothing more than the fact that they are our beliefs. There is no such commitment to the truth of other people’s beliefs in the sense that we are committed to them just because they are someone else’s.\(^2\) This commitment to our own beliefs means that, while it makes perfect sense to hold that other people have false beliefs, or that we ourselves did in the past, this is not something that we can rationally hold about our present selves. This requires some explanation.

One can, of course, come to think of a particular belief one holds that it is false. But if one is rational, coming to believe that one’s belief that \( p \) is false means giving up the belief that \( p \). In fact, I would like to say something stronger: that coming to think that a given belief of one’s own is false simply is giving up that belief. There is thus a conceptual or logical connection between coming to think that a belief of one’s own is false and giving up that belief. But even if one does not want to go that far, and wants to allow the possibility that one might continue to have a belief while coming to think that it is false, one would have to resort to various elaborate psychological hypotheses – mental partitioning, self-deception, the unconscious – to give such a case any plausibility. So we could say that there is at the very least a psychological, if not conceptual, impediment to continuing to have a given belief while thinking that that very belief is false.

By contrast, what we have been considering in this paper is the belief that something one believes is false. Here, especially when one has many beliefs, as people do, the psychological impediments to holding such a belief while maintaining one’s other beliefs do not seem great. And it would certainly be wrong to make the stronger claim that there is a conceptual confusion involved in taking someone to have such a belief. There is no
plausibility to the view that coming to think that something one believes is false is, ipso facto, giving up all one’s beliefs. Yet despite the greater psychological ease of holding RF while continuing to hang on to one’s other beliefs, and despite the absence of any conceptual problem in so doing, one still runs afoul of that special commitment to one’s own beliefs.

Consider an analogous case with a different kind of commitment. Suppose one makes a number of promises to someone. The analogue of the Inductive Argument presents itself. One sees that others typically fail to keep all their promises, and indeed that one has failed to do so oneself in the past. Can one now address the promisee and say that one is confident that one will fail to keep some of the promises one is currently making? Surely, to make such an announcement is, in effect, to cancel the very commitment that makes a promise a promise. In both fiction and life, we see over and over again people who continually make promises that they do not keep. We reach a point where we expect them not to keep all their promises, and we long for them to reach a similar degree of self-knowledge. But self-knowledge, in such a case, does not mean that the person would continue to make promises, only now they would be issued with a caveat that some of them would not be kept. Our reaction is “put up, or shut up.” If one feels that one cannot keep one’s promises, one shouldn’t make them at all.

These comments on promising and promise-breaking all stem from the fact that a promise is a kind of commitment. By saying that belief is also a kind of commitment, I intend these remarks to apply, mutatis mutandis, to belief. Promising, of course, unlike belief, is primarily a public, other-directed phenomenon. So for the moment, consider assertion rather than belief. To assert a number of propositions is, in the first instance, to broadcast a commitment. It is to offer the information contained therein to others, for use in action and inference, and to offer it with a kind of guarantee. Should some of the propositions in question turn out to be false, others would have license to criticize the assertor, to hold that person, at least in some measure, responsible for the consequences. To assert those propositions, and at the same time to assert that some of what one is asserting is false, seems to render the point of the original assertions moot. What is one to make of the assertions? Is one, or is one not, supposed to take them as a guide? Can one, or can one not, hold the assertor responsible if one acts on some of the propositions asserted and things turn out badly because what was asserted was false?

Finally, let us turn to belief. Belief is not the same thing as assertion but there are obvious relations between the two phenomena. Some authors, indeed, analyse belief in terms of assertion. If this were correct, then belief would inherit various properties of assertion (exactly which would depend on the details of the analysis). I myself think that this analysis gets things the wrong way round and that belief should explain those features of
assertion at which we have just looked. But I have no analysis of belief to offer here. I shall simply rest content with pointing out the close connections between belief and assertion and the plausibility these connections offer to my claim that belief, too, represents a commitment that is violated by the adoption of the belief that some of one’s beliefs are false.

I mentioned above the compulsive promise-breaker and her attempts to gain self-knowledge. Such self-knowledge, I suggested, must either result in a renewed resolution to keep one’s promises, thereby making one’s current promises unlike one’s previous promises, or a suspension of promising. Both of these represent ways in which the promise-breaker learns from her past experience. But I also claimed that continuing to make promises, while simply adding the warning that some of the promises would not be kept, did not represent an intelligible way of learning from one’s experience. Analogous remarks again apply to the case of belief. If we are impressed by the presence of mistakes in the beliefs of others and in our past selves, we may rationally suspend belief. This is, of course, the skeptical response. But to the extent that we continue to have beliefs, we must remain impervious to our past failures and the failures of others. In the sense of “learning from one’s mistakes” in which this expression means “seeing that we have made some mistakes and concluding from this that we are making some mistakes now,” it must be said that we cannot learn from our mistakes. The Inductive Argument embodies the policy of learning from one’s mistakes in this sense, and it is for this reason that the Inductive Argument fails.

5.

I argued, in the previous section, for a constraint on ideally rational belief, namely, that an ideally rational set of beliefs cannot include the belief that some of the beliefs in that belief set are false. This constraint, I have suggested, flows from the nature of belief as a kind of commitment. It may be objected that even if I am right about the existence of such a constraint on rational belief, it is not a purely epistemic constraint but rather flows from factors external to the epistemic enterprise. The epistemic enterprise is characterized in some way in terms of the acquisition of true beliefs and the avoidance of false beliefs. (I leave unspecified exactly how it is to be so characterized, a notoriously thorny question.) Epistemic constraints on belief are therefore those constraints imposed in virtue of their contributions to goals described in terms of acquisition of true and avoidance of false beliefs. Now we can clearly understand that RF might be true. In that case, it may be objected, no purely epistemic constraint can adjudicate the belief irrational. Furthermore, the very fact that I supported my case against the rationality of RF by comparison
with the oddity of issuing promises along with a caveat that some of them would not be kept emphasizes the pragmatic (as opposed to epistemic) source of the alleged constraint. It is owing to the way in which RF would undermine the uses to which the expressions of one's beliefs might be put that it is to be avoided, not because it is in some way epistemically less than optimal.

There is little point in haggling over how widely or narrowly to construe such terms as "rationality" and "epistemic." I am happy to concede the point that the source of the constraint against RF may be different from that of constraints on, for example, consistency among one's beliefs. But if I am right in the account I have given of why RF is not ideally rational, it is important to recognize that its source is not pragmatic in the following sense: that its adoption simply depends on a choice one is free either to make or not to make. It is not pragmatic, for example, in the way that constraints stemming from prudential sources are pragmatic.25

I think we can better understand the nature of the constraint against believing RF by seeing at work here another instance of Thomas Nagel’s distinction between "subjective" and "objective" points of view.26 The Inductive Argument draws on an evidential base consisting of the belief sets of other people and our past selves. These are all belief sets about which there is no problem in thinking that they contain false beliefs. The Inductive Argument asks us to treat our own current beliefs the same way. If taking an objective point of view towards a set of beliefs means such things as being able rationally to think that that set contains false beliefs, then the Inductive Argument requires that we take an objective view of our own current beliefs. It asks us to treat our own current beliefs just as if they were the beliefs of no-one in particular, with no regard to the fact that our own current beliefs enjoy a special status for us: they represent our commitments as to how things are.

When we consider our own current beliefs in the light of this commitment, we take up an attitude to them in which the fact that they are our beliefs is crucial. Hence it is a subjective view of them. From this perspective, the Inductive Argument has no force since it relies on ignoring exactly what constrains the attitudes we can take to our own beliefs. Hence my contention, at the end of section 3, that the Inductive Argument does not even show that there is an objectively high probability that we currently have some false beliefs. It would only do this if the appropriate reference class by which to judge the likelihood of false beliefs among our current beliefs were people’s belief sets in general. This reference class is disallowed us when we take the subjective point of view of our own beliefs.27

Even supporters of the Inductive Argument, and of RF as the best way of satisfying REM, implicitly allow that we can only imperfectly take up the objective view of our own beliefs. For when we reason that others’
belief sets contain some false beliefs, this must ultimately rest on taking them to have specific beliefs identified as false. But as we have seen, this is not something that anyone takes to have an analogue in our own cases. We cannot unproblematically think of specific beliefs of ours as being false. To think of a given belief that it is false is thereby to give up that belief (or if not, as I indicated, some peculiar psychological mechanisms must be at work). This much objectivity about our own beliefs is, as everyone allows, impossible. I am suggesting that the same kind of impossibility extends even to RF.

Nagel suggests that an inability to integrate a subjective and objective view of the same phenomenon will lead to alternating between points of view, or developing double vision. Does this thought grant anything to the proponent of RF? Could it not be held that while RF may not be rational from the subjective point of view, it is rational from the objective point of view? If, then, we can alternate between subjective and objective views of our own beliefs, RF will be rational for us when we are in the grip of the objective view. Or if we adopt a kind of double vision, RF will be rational in that half of the ‘visual’ field determined by the objective view.

Perhaps, after all, this will turn out to be the best way of putting things. If so, my conclusions would have to be restricted to the subjective point of view. The interesting outcome would be that there is a point of view (an indispensible point of view, though not a complete one) from which RF is not rational and the Inductive Argument not acceptable. But, in fact, I am suspicious of granting even this much to the proponent of RF. It seems to me that the possibility of adopting an objective view, in the case we have been considering, is somewhat illusory. In this respect, there is a major difference between our current example of a subjective/objective distinction and the cases Nagel considers.

Take, by way of comparison, the subjective/objective distinction as it applies to conscious experience. When we look at mental experiences objectively, we see various brain processes and other physiological events. When we take them subjectively, there is a crucial “what it is like” that characterizes them. Here we seem to have a situation that fits the models of double vision or alternating viewpoints. Whatever competition there is between these objective and subjective views arises only when one of them makes a claim to offer a complete picture of reality. As long as each acknowledges its partiality, they can co-exist in one of the ways described by Nagel. Being able to view one’s experiences in terms of their intrinsic phenomenological qualities is in no conflict with seeing them as neurobiological processes in the brain.

In the case of our attitudes to our beliefs, things are more complicated. Taking any kind of view of our own beliefs involves, at least in part, forming beliefs about them. For example, taking the objective view of
our beliefs and the presence of error in them amounts to adopting RF as a higher-order belief. The higher-order activity of taking a view of our beliefs is thus, at least in part, of the same nature as its first-order subject matter. If we adopt certain objective beliefs about our own beliefs, those higher-order beliefs will be as much subject to that special commitment we owe to our own beliefs as are the first-order beliefs of which we are trying to obtain an objective view. While we may be able to fragment our beliefs and use some to obtain an objective view of others, we cannot obtain a comprehensive objective view of our beliefs *in toto.*

It is not my intention here to go deeply into the metaphysics of the self that is implied by these epistemological remarks. But there certainly are such implications. In crude outline, the upshot seems to be that there is a sense in which, as believing subjects, we are not totally visible to ourselves. Like the eye that sees, but cannot directly see itself, the believing subject takes a view on the world that cannot, in some sense, fully include its own believing self. In a remarkable passage that occurs in a discussion of Moore’s Paradox (which clearly raises many of the issues relevant here), Wittgenstein asks, apropos of his observation that one cannot say of oneself that one *seems* to believe something, even though others can say it of one: “Do I myself not see and hear myself, then?” And his answer is: “That can be said.”

6.

If REM cannot be fulfilled by RF, does that mean it cannot be fulfilled at all? If we cannot learn from our mistakes, are we doomed not to acknowledge our fallibility at all? In this final section I shall explore some ways in which one might attempt to recognize one’s epistemic fallibility while respecting the constraint against believing RF.

One might first think to suggest that one can meet REM not by positively believing something, but by avoiding certain beliefs. The idea would be that even if, as I have been arguing, one should not actually believe that some of one’s beliefs are false, one should in some way be open to that belief; it should be a possible belief for one to hold, relative to one’s other beliefs – a kind of permanently unactualized possibility looming on the horizon. Adding it to one’s belief set would indeed induce irrationality, but owing solely to the new belief itself, not to its inconsistency with one’s other beliefs. The weakest way of doing this would be through conforming to the following principle (PRF stands for Possible Reflexive Falsity):

(Weak PRF) One should have no belief which is inconsistent with the belief that some of one’s beliefs are false.
Weak PRF would be violated if one believed that none of one’s beliefs were false, or alternatively, that all of one’s beliefs were true. But recall that, above, I explicitly stated that rejecting the rationality of RF was not tantamount to endorsing the belief that all of one’s beliefs were true. So nothing I have said so far implies that a rational person would violate Weak PRF.

Nonetheless, Weak PRF is very weak and hardly does justice to our sense of epistemic modesty. A stronger version would read:

(Strong PRF) One should have no beliefs which, taken together, are inconsistent with the belief that some of one’s beliefs are false.

I have argued, elsewhere, that one should rationally believe the conjunction of one’s beliefs. Some authors take the conjunction of all of a person’s beliefs to be equivalent to, or at least to imply, the belief that all of one’s beliefs are true.\(^{30}\) If these authors are correct, believing the conjunction of one’s beliefs – as I have argued a rational person does – would put one in violation of Strong PRF. However, I have also argued that the conjunction of one’s beliefs is not equivalent to, and does not imply, the belief that all of one’s beliefs are true.\(^{31}\) So I have said nothing to imply that a rational person violates Strong PRF any more than Weak PRF. What might constitute a violation of Strong PRF is a case in which someone believes the conjunction of all her beliefs, \(p_1, p_2, \ldots, p_n\), and also believes that \(p_1 \ldots p_n\) are all of her beliefs. But, of course, if this does constitute a violation of Strong PRF, then there is a problem even for those who deny the rationality of believing the conjunction of one’s beliefs. There would equally be a violation of Strong PRF in a situation in which one did not believe the conjunction of one’s beliefs, but still had a belief that \(p_1 \ldots p_n\) were all of one’s beliefs. This belief, taken with each of \(p_1 \ldots p_n\), would imply that all of one’s beliefs were true just as surely as this belief taken with the conjunction of \(p_1 \ldots p_n\) would. Therefore, regardless of one’s attitude to the rationality of conjunctive belief, if one wishes to endorse Strong PRF, one had better deny that it is ever rational to believe that a given set of beliefs are all of one’s beliefs.\(^{32}\) (This seems to be a conclusion of some independent interest.)

Although, as we have seen, Strong PRF does impose some interesting restrictions on one’s beliefs, it is still too weak to capture the full import of REM. So far, our explications of PRF have been of a negative kind, the imposition of consistency requirements on our beliefs. What we need is to provide some positive content, some substantive conception of ourselves and our beliefs that will constitute an acknowledgement of our fallibility. A first step would be to require a belief that the consistency requirements just discussed are in fact met. That is, we might require, for REM, that one believe:
Believing B-PRF, along with actual satisfaction of Strong PRF, would provide a reasonably robust way of meeting REM. But I think we can require even more.

The problem with RF, I argued, was that it was supported by an argument, the Inductive Argument, that overlooked an important distinction between attitudes one can take to one’s own beliefs and attitudes one can take to the beliefs of others. Given the way in which one is committed to the truth of one’s beliefs simply because they are one’s own, I held that one cannot rationally see oneself as a member of a community of error. But this does not rule out a different kind of community. We can, and should, recognize that the methods we have of acquiring beliefs, methods that we share with others, are not guaranteed to yield exclusively true beliefs. The methods we have are unreliable. (By “unreliable” I mean only “not perfectly reliable.”) We have good reason to believe that methods of belief acquisition such as perception, memory and inference are such as to allow information to become lost, garbled, misinterpreted or otherwise corrupted. But community of unreliability in the process does not entail community of error in the product. So recognition of the limits to our belief-acquiring methods does not bring with it a belief that some of our beliefs are actually false.

These observations give a distinctive content to the modality in the belief that, I submit, is the right way for a rational creature to satisfy REM. That belief (I call it RU for Reflexive Unreliability) is:

(RU) Some of one’s beliefs may be false.

What the modality here indicates is that it is consistent with their methods of acquisition that some of our beliefs are false. This gets closer to the heart of epistemic modesty than does the purely formal requirement of B-PRF. Unlike RF, RU can include itself within its own scope without potential paradox. There is nothing specially problematic about a belief set of which the only member is the belief that some of the beliefs in the set may be false. To put the point more generally, should it turn out that we are in fact in an environment in which none of our beliefs could be false, if we were subject to the machinations of a benevolent rather than an evil demon, it would of course be false that some of our beliefs may be false, but there need be nothing irrational about holding such a belief. Whether it is true or false, we may rationally hold RU in a doxastically perfect world.

It would, of course, be a mistake to see in RU the beginnings of an argument for RF. Such an argument might go something like this.\(^3\) We
recognize that our beliefs are the products of unreliable processes. Unreliable processes will yield defective products (i.e. false beliefs) in a certain percentage of cases. So probably a certain percentage of our beliefs are false. If the probability that a certain percentage of our beliefs are false is high enough, we should simply believe that that percentage of them are false. This argument makes just those mistakes discussed above in connection with the Inductive Argument, and in particular with the applicability of Lewis’s Principal Argument. It asks us to treat our beliefs as if they were the beliefs of no-one in particular in order to derive the claim that the fallibility of the means of acquisition, a fallibility which leads to a certain percentage of errors in general, has led to that percentage of errors in our case. The argument also assumes that high objective probability is sufficient for belief, a claim I challenged above.

Some may feel that RU is still too weak to do justice to the full measure of our fallibility. I have some sympathy with this misgiving, but I can find nothing stronger that will not fall foul of the strictures against RF and rely on the blindness embodied in the Inductive Argument to the differences in the ways we must view our own beliefs and the beliefs of others. If something stronger than RU can be found that still falls far enough short of RF to be sensitive to these crucial differences, then I will welcome it as a better means of satisfying REM. Failing that, I maintain that any degree of epistemic modesty greater than that expressed in RU is false modesty.\textsuperscript{34}

Department of Philosophy  
California Polytechnic State University  
San Luis Obispo  
CA 93407

NOTES

\textsuperscript{1} Strictly, RF is a belief-schema, but for ease of exposition, I shall treat RF as if it were a belief. No confusion should arise from this. A similar warning applies to a number of other belief schemata that I treat in the following as if they were beliefs.

\textsuperscript{2} This requires some qualification. In fact, the claim splits into two cases, corresponding to two basic ways in which a given belief set can change. One is by having added to it a new belief which we come to think is true; the other is by having dropped from it a belief we now think is false. In the first case, it will not be true that we think the former belief set contained some false beliefs unless we think that the new one does. Since I am arguing against RF, I am committed to saying that, in this case, we have a former belief set of ours which we do not think contains some false beliefs. In the second case, by contrast, it must be true that we think the former belief set contained some false beliefs. In making my unqualified claim in the text, I am simply ignoring past belief sets of ours that are subsets of our current belief set.

\textsuperscript{3} An explicit statement of the Inductive Argument is given by Jonathan Roorda, “Fallibilism, Ambivalence, and Belief,” Journal of Philosophy 94 (1997), p. 135. The argument...

4 Since nothing will ever hinge on distinguishing the two versions of RF formulated in the text, I shall use RF, equivocally, to refer to each of them. Whichever version fits best into the context should be understood.

5 The argument’s conclusion will not be shown to extend to uncountably infinite belief sets. I don’t know whether it is plausible to think of people as having uncountably many beliefs, but even if it is, I assume no-one will be attracted by the thought that RF’s rationality is dependent on being a member of an uncountable belief set.

6 If you feel that since humans are not perfectly rational, it is of little moment if we satisfy REM in a way that relies on our being irrational, then treat this paper as if it were an investigation into how some hypothetical fallible but perfectly rational beings ought to satisfy REM.

7 The suggestion of restricting the scope of RF in this way is due to Richard Foley, “Justified Inconsistent Beliefs,” American Philosophical Quarterly 16 (1979), p. 252, n.12.

8 It will be false if taken to express an existential commitment to another belief; otherwise it will be trivially true.


10 Many authors argue, quite reasonably, that degrees of belief are likely to be vague. Various technical expedients have been introduced into theories of degrees of belief to accommodate this insight. I shall ignore them in what follows for ease of exposition.

11 Or perhaps all of them except some subset of Cartesian-style indubitable beliefs. I will ignore this caveat.

12 See Hawthorne and Bovens, “The Preface, the Lottery and the Logic of Belief,” for some precisely worked out examples of how the number of beliefs is relevant to the rationality of the belief that some of one’s beliefs are false. The number of beliefs required for this to be rational, of course, will be affected by the value of one’s belief threshold.

13 The Lockean thesis is pressed from two directions. On the one hand, it is challenged by Bayesianism, which seeks to replace the notion of belief with that of degrees of belief, and not to accommodate the two notions. See Richard Jeffrey, “Dracula Meets Wolfman: Acceptance vs. Partial Belief,” in Marshall Swain (ed.) Induction, Acceptance, and Rational Belief (Dordrecht: Reidel, 1970), pp. 157–85. I ask about the rationality of RF. But a) RF is itself a candidate for belief; b) RF concerns other beliefs, not degrees of belief; and c) RF assesses beliefs in the light of truth and falsity, which do not unproblematically apply to degrees of belief. Hence, the issues dealt with in this paper could not even be raised in the Bayesian context. However, analogous issues do arise in the discussion of van Fraassen’s Principle of Reflection (especially its synchronic part). See Bas van Fraassen, “Belief and the Will,” Journal of Philosophy 81 (1984), pp. 235–56. I hope to take this up in another paper.

The other direction from which the Lockean thesis is challenged, which represents my, unargued for, position, is the view either that the notion of degrees of belief is illegitimate or that the two notions of belief and degrees of belief are logically unrelated to each other. See, for example, Isaac Levi, Gambling with Truth: An Essay on Induction and the Aims of Science (New York: Alfred A. Knopf, 1967); Mark Kaplan, Decision Theory as Philosophy (Cambridge: Cambridge University Press, 1996).
Of course, it might be argued that there are other reasons for believing RF than the Inductive Argument, and that these other reasons will not support a similar conclusion that many of our beliefs are false.

I am grateful to Tyler Burge for helping me see that there is an objection to my argument against RF here that is independent of the Lockean objection.

I am grateful to a referee for this journal for raising this objection.


This is Mark Kaplan’s position. See Kaplan, Decision Theory as Philosophy, pp. 102–54.

I discuss these issues at slightly greater length in “Believing Conjunctions,” Synthese 118 (1999), pp. 201–27.

I have argued elsewhere that we should believe the conjunction of all our beliefs, and some take this to be equivalent to believing that everything we believe is true. But the conjunction of all one’s beliefs is not equivalent to the belief that everything one believes is true; one may have an incorrect view, or more likely no view at all, about whether a given conjunction exhausts all of one’s beliefs. See my “Believing Conjunctions” and section 6 below for more extensive discussion of this point.

There are, perhaps, a few exceptions to this. Believers in Papal Infallibility might be described as being committed to accepting (a few of) the beliefs of the Pope just because they are his beliefs. Such situations are clearly anomalous from an epistemological point of view.

The analogy between promising and believing was suggested to me by van Fraassen’s interesting paper, “Belief and the Will.” One important disanalogy between promising and believing that I don’t consider concerns the huge difference in the number of beliefs a person typically has and the number of promises he typically makes. This was pointed out to me by Bernie Kobes.

See Kaplan’s definition of belief in terms of assertion in Decision Theory as Philosophy, pp. 107–11. See also De Sousa, “How to Give a Piece of Your Mind”; van Fraassen, “Belief and the Will”; Roorda, “Fallibilism, Ambivalence, and Belief.”

We may limit the restriction so that it does not affect some subset of our beliefs such that among beliefs of that kind, we do not take ourselves or others to have been in error.

Pascal’s Wager may be seen as a constraint on rational belief (that it should include a belief in the existence of God) stemming from prudential sources. Whether or not one subjects oneself to that constraint depends on a choice influenced by how much we care about the possible gains and losses of adopting such a belief.


Strictly speaking, it is not the reference class – belief sets in general – that is disallowed us by the subjective view of our own beliefs, but the combination of that reference class with the particular property, that of containing some false beliefs, the instantiation of which in our own case is at issue.

There is an analogy between the idea of an infinite hierarchy of levels of belief, each higher-order level containing an objective assessment of the beliefs in the level below, and the Tarskian response to the Liar Paradox, namely an infinitely extending hierarchy of
languages, each higher-order one able to attribute truth and falsity to the sentences of the language in the level below. This similarity is hardly a surprise, given the connections we have seen above between the Liar Paradox and the problems of RF.


31 See Evnine, “Believing Conjunctions,” for my argument that one should believe the conjunction of one’s beliefs and for my claim that such a conjunction does not imply that all of one’s beliefs are true.

32 Alternatively, one could try and argue that having a belief that $p$ does not imply having a belief that one’s belief that $p$ is true. A fuller discussion of which of these strategies is preferable as a way of avoiding a violation of Strong PRF must be deferred for a further paper.

33 Ivan Fox pressed this objection on me.

34 Many thanks to Giovanna Pompele, Otávio Bueno, Ivan Fox, Bernie Kobes, an anonymous referee and especially Tyler Burge for help in improving this paper. I am also grateful to Cal Poly for a State Faculty Support Grant in Winter quarter of 1999, during my enjoyment of which much of this paper was written.