SIMON J. EVNINE

BELIEVING CONJUNCTIONS

1.

Among philosophers who discuss the nature of rational belief, there is considerable controversy over the question: is it rational to believe the conjunction of one's beliefs? In other words, should a theory of rational belief contain a conjunction principle:

(CP) If S is rational, then if S believes *A* and S believes *B*, then S believes *A and B*

I shall argue in this paper that it should.[1] To begin with, I shall defend (CP) against several criticisms that have been launched against it. These criticisms are of two kinds, which I shall call internal and external respectively. Internal objections are that a theory that includes (CP) fails to give an account of what it is rational to believe that is satisfactory by its own standards. In particular, since almost everyone agrees that belief in a contradiction is not rational, (CP) is criticized on the grounds that it would imply that beliefs in some contradictions are rational. External objections maintain that the degree of idealization in a theory including (CP) is so great as to make it irrelevant to various real-world phenomena to which a theory of rational belief should be relevant.[2]

After addressing these objections I shall briefly reject one avenue of support for (CP), based on the claim that rational belief should be closed under entailment. Finally, I proceed to what I see as the right way to argue for (CP). This will involve an examination of the idea that people's beliefs can be divided into fragments that do not rationally affect each other. I shall argue that within such fragments, people do believe the conjunctions of their beliefs. Putting this together with the claim that a rational person should not have a multiply fragmented belief system, we will arrive at an argument for (CP).

Regardless of one's final verdict on (CP) itself, I hope it will emerge uncontroversially from this paper that whether a theory of rational belief

Synthese **118:** 201–227, 1999.

should contain a conjunction principle, so far from being a dry, self-contained and esoteric question, is intimately connected with a number of deep and interesting philosophical themes: How should we represent our own epistemic fallibility, what is the purpose of a theory of rational belief, what kind of a theory of the nature of belief should we adopt? Ultimately, behind the dispute over (CP), we shall run to ground two competing philosophical visions of a very general nature indeed, concerning the nature and unity of the mind.

As I mentioned above, it is almost universally agreed that a theory of rational belief should contain a principle prohibiting belief in a contradiction:

(NC) If S is rational, then S does not believe *A and not-A*.

Unless otherwise stated, I shall assume that any theory of rational belief does contain such an injunction.[3]

2.

Two much discussed paradoxes, the Preface Paradox and the Lottery Paradox, provide the basis for some powerful internal objections to (CP). I start with an examination of the Preface Paradox.

Although it is easy enough to state the Preface Paradox in a casual and intriguing way – an author believes everything she has written in her new book, yet on the basis of past experience, also believes she has made some mistakes in it – it is difficult to state convincingly in a way that casts doubt on (CP).[4] Suppose we have a theory of rational belief, **T**, that does not include (CP) (but according to the assumption made at the end of Section 1, does contain (NC)). The goal for the objector to (CP) is to use the Preface Paradox to show that the theory of rational belief $\mathbf{T}_{(\mathbf{CP})}$, consisting of **T** plus (CP), will count as rational some belief in a contradiction. This will serve as a *reductio* of (CP).

Consider the four following propositions:

(1) everything I believe is true;

(2) something I believe is false;

(3) p_1 and p_2 ... and p_n;

(4) not-p_1 or not-p_2 ... or not-p_n,

where p_1, p_2, ... p_n are in fact all of my beliefs. Let "Bel[ ]" designate the belief in a proposition named in the square brackets. The contradiction belief in which is supposed to be counted as rational by $\mathbf{T}_{(\mathbf{CP})}$, by means of the Preface Paradox, may take one of two forms: either Bel[(1) and (2)] or Bel[(3) and (4)].[5]

In order to derive the rationality of Bel[(1) and (2)] or Bel[(3) and (4)], we require two applications of (CP). One is to get the rationality of Bel[(1) and (2)] from the rationality of Bel[(1)] and Bel[(2)], or Bel[(3) and (4)] from Bel[(3)] and Bel[(4)]. But if that by itself were to be reason to reject (CP), we would have to stipulate that Bel[(1)]–Bel[(4)] were rational. In fact, a second and main role that (CP) is supposed to play in the Preface Paradox is in enabling the derivation of the rationality of Bel[(1)] or Bel[(3)] from *any* rational belief set. Thus, if the Preface Paradox is to be used as a *reductio* to reject (CP), only the rationality of Bel[(2)] or Bel[(4)] need be independently established.

Suppose we try and show that $\mathbf{T}_{(\mathbf{CP})}$ will count as rational Bel[(3) and (4)]. It is easy to see that (CP), applied to *any* set of rational beliefs p_1 ... p_n, will imply the rationality of Bel[(3)]. If we can show that if I rationally believe each of p_1 ... p_n I can also be rational in having Bel[(4)], we can apply (CP) again to obtain the rationality of Bel [(3) and (4)]. Of course, I cannot rationally believe (4) because I rationally believe the negation of any particular one of p_1 ... p_n. Rather, the rationality of Bel[(4)] is urged upon me by the recognition of my own epistemic fallibility. I know from experience that there are many things I have believed before that have turned out to be mistaken. So is it not rational to believe that some of my current beliefs are also mistaken? Well, let us suppose it is rational (I shall dispute this later on). This is to support the rationality of Bel[(2)], not Bel[(4)]. But (2) is not logically equivalent to (4), and hence it does not contradict (3). If all $\mathbf{T}_{(\mathbf{CP})}$ implied was the rationality of believing the conjunction of (2) and (3), it would not have been convicted of implying the rationality of belief in a contradiction.

The objector to (CP) would need to show either that (2) implies (4), or that the recognition of epistemic fallibility directly supports (4). Neither of these options is plausible. (2) is a second-order belief, a belief about my beliefs, while (4) is not. So we cannot derive (4) from (2) unless we make a number of other assumptions. Nor does the recognition of epistemic fallibility directly support (4). Recognition of one's fallibility should not be taken to entail an accurate inventory of one's beliefs. I might well believe that something I believe is false not because I think that one of p_1 ... p_n is false, but because I vaguely think (falsely) that I have some further, unspecified belief that is the culprit.

If (2) is a more plausible form for representing my epistemic fallibility than (4), we might then think to derive the rationality of Bel[(1) and (2)], by showing that an application of (CP) to my belief set leads to the rationality of Bel[(1)]. But here, too, we run into trouble. A simple application to my beliefs $p_1 \ldots p_n$ of (CP) yields the rationality of Bel[(3)], not Bel[(1)]. Nor is (3) logically equivalent to (1). Again, (1) is second-order while (3) is not.

Of course, principles of rational belief might be formulated that would allow us to derive the rationality of Bel[(4)] from that of Bel[(2)], or of Bel[(1)] from Bel[(3)]. But in that case, the Preface Paradox could as well be seen as a *reductio* of any of these other principles as of (CP).

There is a real philosophical moral behind all this logic-chopping. Behind the paradox is the intuition that as a rational person, one should possess a sense of one's own epistemic fallibility. And so one should. But we must be careful in stating the form this sense takes. I have already argued that Bel[(4)] is not a convincing form. However, even Bel[(2)] is problematic as a rational representation of epistemic fallibility. Suppose the total set of my beliefs, not including Bel[(2)], contains no false beliefs. Now if we add Bel[(2)] to my belief set, paradox ensues. By hypothesis, none of my other beliefs are false, so if Bel[(2)] is true, it must be false as well. If, on the other hand, Bel[(2)] is false, then I do indeed have a false belief, so Bel[(2)] is true.[6]

This paradox does not immediately show that Bel[(2)] cannot be rational. But further argument can be supplied, though it requires a few assumptions. If a rational creature has some logical acumen and also has concepts of belief and truth (to enable it to have beliefs about the truth of its beliefs) then it would have the resources to recognize that if none of its other beliefs were false, Bel[(2)] would be paradoxical. Furthermore, it is plausible that a rational creature should realistically hope that none of its beliefs are false. I know of no integration of a logic of hope with a logic of belief in the light of which I could assert categorically that it is irrational to have a belief that would be paradoxical if one's realistic hopes were fulfilled. But at the very least, there is reason to be suspicious of the rationality of Bel[(2)].[7]

Epistemic modesty, I submit, is much better represented by the belief that some of our beliefs *might be* false, not that some of our beliefs are false. This belief could be true even if none of our beliefs actually were false. Furthermore, the belief that some of our beliefs might be false seems to get to the heart of the reason for epistemic modesty better than the belief that some of our beliefs are false. Modesty is appropriate because our means of acquiring beliefs are unreliable and prone to failure. But

unreliability, while a defect in its own right, is no guarantee of actual failure.[8]

3.

Richard Foley, who objets to (CP) on just the kind of grounds discussed in the previous section, notes the problem with (2) that we have just discussed and suggests that it can be gotten round with a simple expedient. Instead of attempting to generate the belief in a contradiction using (2), we should generate it with a modified version:

(2′) something I believe, other than this belief, is false.

Let us call the Preface Paradox modified in this way the Modified Preface Paradox. Some of the problems I noted with the Preface Paradox would still obtain, but the Modified Preface Paradox would get round the problem that one of the components of the generated belief in a contradiction is itself inherently paradoxical.[9]

However, the Modified Preface Paradox will now run into a different problem. It is essential to any version of the Preface Paradox that each of the conjuncts of the contradiction be rational. The rationality of (2) was supported by a general appeal to the idea of epistemic fallibility. But what support can be given for the rationality of (2′)? It cannot be a general view about our fallibility, for (2′) makes invidious distinctions among our beliefs and gives a special status to some that it does not give to others, namely, exemption from possible error.[10]

The general form of the Modified Preface Paradox is a specification of a subset of a rational belief set such that there is good reason to believe each member of that subset is true, and a rational belief not in that subset that some member of the subset is false. Can we find another case that conforms to this general pattern, but avoids the pitfall of the Modified Preface Paradox?

We can: the Lottery Paradox.[11] If there is a fair lottery with a large number of tickets, say 1,000, there is an overwhelming probability that for each ticket *n*, *n* will not win. Hence it is rational to believe, for each *n*, that ticket *n* will not win. A first application of (CP) is supposed to yield the rationality of the belief that no ticket will win. Yet since the lottery is fair, it is rational to believe that one ticket will win. A second application of (CP) implies the rationality of the conjunction of these beliefs, but that is a contradiction. As in the Modified Preface Paradox, so here we have a set of beliefs such that it is rational to believe each one of them. Yet

unlike our first attempt to set up a Modified Preface Paradox, the special circumstances of the lottery situation mean that we do not have to rely on unjustified representations of epistemic modesty to have a reason to believe that some of the beliefs in that set are false.

Not only does the Lottery Paradox do better than the Preface Paradox by avoiding the paradoxicality of the unrestricted belief that something one believes is false. It also avoids some of the other problems I raised against the Preface Paradox. In particular, I charged that the conjunction of one's beliefs is not equivalent to the belief that everything one believes is true, nor the belief that something one believes is false equivalent to the disjunction of the negations of all one's beliefs. But the particular nature of the lottery example makes it much more plausible to hold that the conjunction of the beliefs that ticket n will not win, for n = 1–1,000, is equivalent to the belief that no ticket will win. And likewise, the belief that one of the beliefs that ticket n will not win is false is plausibly identified with the belief that either ticket 1 will win, or ticket 2 will win .... This is because when we consider the conjunction of all of one's beliefs, we are dealing with a vague, ill-defined and exceedingly large number of beliefs. Consequently, it is a substantive and controversial assumption that a rational creature will have a belief that $p_1 \ldots p_n$ are all of its beliefs. But in the case of the Lottery Paradox, we are dealing with a relatively small and well-defined set of beliefs. Thus there is no problem about switching between first-order beliefs about tickets and second-order beliefs about whether certain first-order beliefs about tickets are true or false. (This is why the Lottery Paradox can be assimilated to the form of the Modified Preface Paradox.)

Do we, then, have a good argument against including (CP) in a theory of rational belief? The Lottery Paradox has generated a large number of responses, some of them even made with the aim of defending (CP). Without evaluating them all here, let me just state my own preferred response.[12] I believe that the problem comes in the initial move, the claim that it is rational to have the belief that ticket n will not win, for all n. This move is defended by Foley who says, in effect, that by increasing the number of tickets used in the example, we can raise the probability that a given ticket will not win to a point where it is greater than the probability that any given empirical belief is true. Thus, in a lottery with 20 million tickets, the probability that ticket n will not win, for any n, is 20 million to 1, much better odds than we would give to an uncontroversial example of what might be a rational belief under certain circumstances, such as that your car is not being stolen from your garage at this very moment.

This argument presupposes an acceptance rule that makes a belief's rationality contingent only on the probability of its truth.[13] But this ignores

the impact of the existence of other beliefs, except as they factor in to the probability calculations for the given belief. The special circumstances required for an instance of the Lottery Paradox include that we start with a situation where there is a set of propositions that are not all compatible (not all of the tickets will lose) but where there is nothing to say in favor of the truth of any of the set that cannot be said equally for all the others. Let us call such a set an Indifferent Set. In deciding whether it is rational to believe a given proposition that ticket *n* will not win, we must look not only at the probability of its truth, but also at whether it is a member of an Indifferent Set. A theory of rational belief, I suggest, should include a provision that where a proposition is a member of an Indifferent Set, no matter what the probability of its truth, it is not rational to believe it.[14] This solution to the Lottery Paradox clearly requires further defense and exploration, but the issues it raises go beyond the scope of this paper.[15]

4.

There is another issue that we have not yet considered involved in the use of both the Preface and the Lottery Paradoxes to challenge (CP): the distinction between believing a contradiction and having contradictory beliefs. While consistency is generally agreed to be a requirement in a theory of rational belief, the exact form it should take is subject to controversy. Kyburg distinguishes what he calls Weak Consistency from Strong Consistency. Weak consistency requires only that a rational belief set should contain no belief such that any proposition is provable from it. The most obvious case of such a belief would be a belief in a contradiction. Strong Consistency, by contrast, requires that a rational belief set shall contain no beliefs which together entail any proposition. So the belief sets {*p*, *if p then q*, *not-q*} and {*p*, *not-p*} do not violate the Weak Consistency condition but do violate the Strong Consistency condition. (I shall simplify discussion by assuming that the only violations of these consistency conditions are belief in a contradiction and contradictory beliefs.)

As I stated at the outset of this paper, almost everyone accepts that a theory of rational belief should have at least a Weak Consistency requirement. Given that assumption, the question of whether such a theory should contain (CP) is equivalent to the question of whether it should have a Strong Consistency requirement. For if a proposed rational belief set contains beliefs of the form *A* and *not-A*, thereby putting it in violation of a Strong Consistency requirement, then by (CP) it will contain the belief *A and not-A*, putting it in violation of a Weak Consistency requirement.[16] For this

reason, much of the following discussion, though still about (CP), will be framed in terms of the merits of Strong versus merely Weak Consistency.

That treating the acceptability of (CP) as equivalent to the acceptability of a Strong versus Weak Consistency requirement is a legitimate way to proceed in the context is shown by the arguments we have considered against (CP). In the Preface Paradox (and *mutatis mutandis* for the Lottery Paradox), there is no violation of the Weak Consistency condition until (CP) is applied for the second time to yield either Bel[(1) and (2)] or Bel[(3) and (4)]. The conjuncts of each of these conjunctions – (1) and (2), and (3) and (4), believed separately, violate the Strong but not the Weak Consistency condition. Thus, the use of the paradox as an argument against (CP) depends on accepting the Weak, but rejecting the Strong Consistency condition. For someone who thought the Weak condition was still too strong, perhaps because there should be no consistency requirement on a rational belief set, Bel[(1) and (2)] or Bel[(3) and (4)] would not be problematic. Alternatively, if one accepted the Strong Consistency condition, then the scenario envisaged in the paradox would create problems even without (CP) and hence could not be taken as the basis of an argument against that principle. Hence for those who use the Preface and the Lottery Paradoxes as arguments against (CP), the rejection of (CP) is equivalent to the acceptance of Weak Consistency and the rejection of Strong Consistency. I shall argue in the next section, however, that the factors that lead us to want Weak Consistency, i.e., to disallow from rational belief sets single beliefs from which any proposition is provable, should lead us also to want Strong Consistency, i.e., to disallow from rational belief sets any beliefs from which taken together any proposition is provable.

Kyburg supports the distinction between these two phenomena when he writes: “I probably cannot believe a contradiction, or act on one. But I can certainly believe, and even act on, each of a set of statements which, taken conjointly, is inconsistent” (1970, 60). The distinction is here drawn in two different ways. First, believing a contradiction is probably impossible; having contradictory beliefs is certainly possible. Secondly, even if believing a contradiction is possible, one cannot act on one; contradictory beliefs can, by contrast, all be acted on separately. These differences are supposed to support the view that making Weak, but not Strong, Consistency a condition on rational belief captures an important notion of rationality, one on which the rational is tied to what is certainly possible to believe, and to beliefs that can be acted on, but excludes what is probably impossible to believe or what could not be acted on even if believed.

Neither of these considerations provides good grounds for a distinction between Weak and Strong Consistency as conditions on rational belief.

Experience teaches us (at least it has taught me) that all too many people *can* believe even a simple, obvious contradiction. Many apparent beliefs in a contradiction may turn out to depend on some equivocation, but many, I believe, do not. Among those who believe contradictions, some are undoubtedly frivolous and epistemically irresponsible. But some are not. Some are even philosophers! Of course, one who believes it impossible knowingly to believe a contradiction may argue that any appearances to the contrary can be ignored. But in that case, we need some reason for discounting the appearances in the form of an argument that it is impossible to believe a contradiction. A simple assertion that this is impossible, in the face of so much disconfirming evidence, is not enough.

Nor is Kyburg's second point, that it is impossible to act on a contradiction, convincing. First, we might observe that one could be said to be acting on a conjunctive belief any time one acts on one of its conjuncts. Hence, any action could be motivated by, or based on, a belief in a contradiction. But more relevantly, there are many ways in which one could act on both conjuncts of a contradiction. For example, one might defend the truthfulness of it, or its inability to be grasped by the human mind. One might write a book attempting to prove each conjunct, or defending the view that contradictions can, in general, be true.[17]

5.

Perhaps a distinction between the rationality of having contradictory beliefs and the irrationality of believing a contradiction can be made by invoking the concept of awareness. Freudian issues aside, it is plausible to maintain that we can be aware of each belief we have. Anyone who believes a contradiction, therefore, may be aware that he has a belief that cannot be true. But no such conditions guarantee that anyone who has contradictory beliefs can be aware of the problem in his belief set. For awareness of the problem of contradictory beliefs would require the co-awareness of two distinct beliefs. But it is less plausible to hold that we can be co-aware of any two distinct beliefs than to hold the corresponding principle about awareness of single beliefs.[18]

This simple observation has been given a new spin by philosophers working in the tradition of naturalized epistemology.[19] On its basis, they have worked up, sometimes explicitly and sometimes implicitly, an argument against making Strong Consistency a requirement on rational belief, and hence an argument against (CP). This argument takes us from what I called "internal" objections to (CP) to "external" objections to it. It draws on work in Computability Theory to show that the fulfilment of a Strong

Consistency requirement on rational belief would be so far beyond the resources of any finite creature as to make that requirement a pointless idealization.

In Computability Theory, a distinction is made, among tasks that can be performed by the application of an algorithm, between those in which the performance time increases as an exponential function of an input parameter and those in which performance time increases merely as a polynomial function of that parameter. Since for any exponential function and any polynomial function, there will come an argument past which the value of the exponential function will exceed that of the polynomial function, algorithms the execution time of which increases exponentially are considered inherently inefficient.[20] Checking for consistency between different beliefs is a task for which the execution time increases exponentially as a function of the number of beliefs being checked. To give some idea of the scale of the problem, it is worth quoting one impressive statistic. It concerns the question of how large a body of beliefs a computer could check for consistency by the truth-table method.

> Suppose that each line of the truth table for the conjunction of all these beliefs could be checked in the time a light ray takes to traverse the diameter of a proton, an appropriate "supercycle" time, and suppose that the computer was permitted to run for twenty billion years, the estimated time from the "big-bang" dawn of the universe to the present. A belief system containing only 138 logically independent propositions would overwhelm the time resources of this supermachine. (Cherniak 1986, 93)

Since a normal person will certainly have far more than 138 beliefs, we here have a reason for denying Strong Consistency as a requirement on rational belief: a theory of rational belief that included such a requirement would be idealized not only in the sense that it set standards that no person could reasonably be expected to live up to, but that it set standards that were, in some sense, *impossible* for any finite creature to meet. Weak Consistency, by contrast, would only demand that one check each of one's beliefs and this is a task the execution time of which increases polynomially in relation to the number of beliefs. So even if, in fact, we have too many beliefs to check to see if they are contradictions, there is nonetheless an in-principle feasibility to the task, and so it is not pointless to require it for rationality.

This objection to Strong Consistency is external because it depends on a conception of what the goals of a theory of rational belief should be, and an assessment of how well a proposed theory achieves those goals. What, then, are the goals of a theory of rational belief in the light of which we should reject a theory that includes a Strong Consistency requirement? One account of what a theory of rational belief is for is that it tells us "what

a man is obligated, permitted, and forbidden to believe, from the point of view of rationality" (Swain 1970, 27). We must understand the phrase "the point of view of rationality" as referring to some extra-theoretical conception of rationality that we are attempting to do justice to. It is by appeal to such an extra-theoretic conception of rationality, for example, that almost all theorists agree that at least Weak Consistency is an appropriate part of a theory of rational belief. Needless to say, however, such an extra-theoretic conception is likely to be highly vague.

Now is it pointless, with this aim in mind, to include a requirement in a theory of rational belief that cannot be met? We need to consider the notion of idealization. A theory is described as idealized if it abstracts from various features of its subject matter to apply to an impossibly or improbably tidied up domain. In scientific theories, this is all that is meant. But in normative cases, idealization carries another implication: that of setting up a goal to be striven for. The fact that the goal may be beyond our reach does not necessarily undermine its value to us as a goal. Consider the analogy often made between theories of rational belief and theories of morality. It is not implausible to argue that we advocate for ourselves moral goals that we believe to be beyond our ability to achieve. Given human nature, for example, it may in some sense be impossible to love one's neighbor as oneself, or to treat others only as ends and never as means. But it would be far from pointless to adopt these as ideals to which one should approximate as far as possible. These norms might form part of a moral theory because, from an extra-theoretic standpoint, we judge that a world that conforms to them has some value to it. Analogously, it may be impossible for us to ensure consistency among all our beliefs, but the goal of an inconsistency-free body of beliefs may still be worth striving for. Indeed, almost every one agrees that it is worth striving for. Even naturalized epistemologists, if it be pointed out to them that they believe p and believe *not-p*, will agree that some effort could reasonably be put into resolving the situation by giving up one of those beliefs. The only question is, how much effort should we invest in hunting out such situations? Naturalized epistemologists may be right that there are better uses we can make with the limited resources we have in time and computational power than obsessively checking our belief system for inconsistencies. But this does nothing to undermine the desirability, the ideal, of a consistent set of beliefs. What is the value of a world that conforms to this norm? I don't know exactly how to describe it, but it seems clear that it cannot be different from the value of a world that conforms to the norm of a Weak Consistency requirement.

There is another goal that the theory of rational belief is sometimes supposed to serve in relation to which the intractability of checking for

consistency might be thought to provide an argument against a Strong Consistency requirement. This goal is to serve as (or as part of) a theory of belief attribution. Among possible theories of belief attribution some make no use of assumptions of rationality. For example, if one were to think that beliefs were written in the brain in some way, neurophysiology might provide a theory of belief attribution that made no assumptions about the rationality of beliefs so attributed. Others do make use of rationality assumptions. Among these latter some, such as Cherniak (1986), use very minimal assumptions while others, for instance those advocated by Davidson (1984) and Dennett (1987), make use of very strong ones. Let us call a theory of belief attribution that *is* informed by some rationality requirement, of whatever degree of strength, a Rational Theory of Belief Attribution (RTBA).

Against this background, we can see the objection to a Strong Consistency requirement as part of the general attack on the use of strong rationality requirements in an RTBA. The general argument against an overly idealized theory of rational belief is this. If an RTBA has an overly idealized rationality constraint, there will be a disparity between the beliefs that the RTBA predicts an agent should have and the beliefs that the agent actually has.[21]

A full discussion of this objection would open some very deep issues in the philosophy of mind. In particular, we would need to consider exactly what the relationship is between the rationality constraints in an RTBA and the kinds of facts in virtue of which an agent has a given set of beliefs. The main question is, are we to suppose that there are ways of telling which beliefs an agent has that do not already play a role in the theory of belief attribution? Only if there are does it make sense to suggest that an RTBA might be unacceptable because it makes an incorrect assignment of beliefs to people. Suppose, then, that there is a way of ascertaining which beliefs someone has independently of the theory of belief attribution, or that the facts in virtue of which someone does have given beliefs are logically independent of the precepts of a theory of belief attribution. In that case, it would be simply dogmatic to insist that a theory of belief attribution should contain any rationality constraint at all. And if it should contain such a constraint, it would be dogmatic to insist that it be of a particular kind. It will be an empirical question whether an agent's beliefs are better predicted by a theory that does employ some rationality constraints of a particular kind or by a theory that does not. The fact that the Strong Consistency requirement is strictly unfulfillable does not show that an RTBA containing it will do worse at predicting which beliefs an agent has than one with only

a Weak Consistency requirement. After all, people clearly do, at least to some extent, seek to avoid inconsistencies among their beliefs.

If, on the other hand, there is no way of determining the beliefs of a rational agent other than through a theory of belief attribution informed by a rationality constraint, then we cannot argue against a Strong Consistency requirement on the grounds that it 'gets things wrong'.

I therefore conclude that the kind of computational intractability of checking for consistency among different beliefs does nothing by itself to show that a theory of rational belief should not contain a Strong Consistency requirement, or that, if it contains a Weak Consistency requirement, it should not also contain (CP). It remains possible that an empirical argument might be made to show that a theory of rational belief intended to serve as part of a theory of belief attribution might be better off without (CP) but as far as I know, such an argument has not yet been attempted. And even if it were successful, it would do nothing to show that (CP) should not form part of a theory of rational belief used for the purposes of an epistemic ideal.

6.

So far, I have been defending the presence of (CP) in a theory of rational belief against various objections made against it. But defense against objections does not add up to a case in favor. So what supports the inclusion of (CP) in such a theory? Before I answer this, let me explain one reason for including (CP) that I do *not* wish to rely on.

In line with his distinction between Weak and Strong Consistency, Kyburg distinguishes between Weak and Strong Deductive Closure (DC):

> (Weak DC) If S is rational, then if S believes A, then for all B such that A entails B, S believes B.
>
> (Strong DC) If S is rational, then if S believes A and S believes B, then for all C such that A and B taken together entail C, S believes C.[22]

As with the consistency requirements, Weak DC plus (CP) is equivalent to Strong DC. So one reason for supporting (CP) would be that one believes in Strong DC.

I do not think, however, that (CP) should be based on Strong DC, since neither Strong nor Weak DC should form part of a theory of rational belief if it is to be used as part of an RTBA. By contrast, the argument I shall give for (CP) will support its inclusion in a theory of rational belief in any of

its uses. The reason why neither form of DC should be part of an RTBA is of the same kind as those raised, but not accepted, against the inclusion of a Strong Consistency requirement at the end of the previous section. Strong DC is clearly a bad idea in an RTBA, for it would imply, for anyone with a handful of modest logical or arithmetical beliefs, that we attribute to her beliefs in things quite remote and arcane to her. This is *a priori* implausible.[23] In a theory which also includes (CP), Weak DC will yield the same unpalatable consequences. But even without (CP), it seems that there will still be cases where Weak DC would have us attribute to someone beliefs that it is clearly unreasonable to take her to have. For example, someone studying number theory may happen to believe not only each of the Peano axioms by itself, but also their conjunction. Yet it would be little less implausible to attribute to such a person beliefs in *all* the consequences of the conjunction of those axioms that it would be to attribute them to one who believed each of the axioms but not their conjunction.

7.

Finally, I come to the reason why I think we should accept (CP). Let me start by considering not (CP) itself, but a non-normative counterpart to it:

> (Conj) If S believes *A* and S believes *B*, then S believes *A and B*.

(Conj) is the claim not that it is rational to believe the conjunction of one's beliefs, but that people do believe the conjunctions of their beliefs. I will not support (Conj) in unqualified form, but an examination of the conditions under which it does and does not hold will be illuminating for a consideration of (CP).

I begin by giving an argument for (Conj). The argument is that, in normal circumstances, its truth is presupposed by our practice of attributing conjunctive beliefs. Take any two propositions that are of no special logical or emotional significance to someone, say that grass is green and that snow is white. I contend (this is intended as an empirical observation about our practice) that if we have good grounds for attributing belief in each of these to a person, then, absent any special circumstances, that is all we need to attribute to that person a belief in their conjunction. Or if, for example, we are summing up a position someone has just explained at some length, we can do so by attributing to that person a large conjunctive belief. This conjunctive belief describes in a single proposition a number of different propositions that were expressed severally over a period of time.

We do not require evidence that, for example, some concrete psychological process has occurred, a process in which the two individual beliefs are put together or conjoined.[24] Actual belief in the conjunction of one's beliefs is the default mode in belief attribution. With the restriction to normal circumstances, therefore, (Conj) acts as a presupposition of our practice of attribution of conjunctive beliefs.

What is the status of (Conj) itself? There are a number of different ways in which it could be taken. It could be, simply, an empirical generalization: this is what in fact happens, or usually happens. Hence we assume its truth in belief attribution. Or it could be seen as a kind of *a priori* presumption of the applicability of a normative principle such as (CP). In that case, of course, we could not use (Conj) to establish (CP). Finally, and this is the way in which I like to think of it (though nothing in what follows will depend on this), it could be seen as a conceptual claim, resting on the thesis that, in normal circumstances, being in a state of believing a conjunction simply is being in a state of believing its conjuncts. There is no state of believing *A and B*, distinct from the state of believing *A* and believing *B*. This is not a popular view, but it has received some support in the literature.[25] Clearly, a full discussion of it would require an extensive examination of the nature of belief, something quite beyond the scope of this paper.[26]

I said above that I shall not defend (Conj) in unqualified form. What qualifications are necessary? The first and most important qualification concerns the 'normal circumstances' that I held must obtain for the truth of (Conj) to be presupposed by our practice of attributing conjunctive beliefs. What are these normal circumstances?

(Conj), of course, does not apply across different people. If S_1 believes *p*, and S_2 believes *q*, there is no reason to expect either of them to believe *p and q*. If, then, there are cases where a single person includes within himself different person-like entities, (Conj) might fail to apply across these entities as well. Relevant cases need to meet two sets of constraints. On the one hand, we must be dealing with a single person. In the eponymous operetta, Cox and Box share a single room, one working by day, the other by night, neither knowing of the other's existence, each thinking of himself as the sole occupant of the room. If a single body were inhabited in this way by two minds, then it might be true that one believed *p*, the other *q*, but neither *p and q*. But apart from sameness of body, there is little to justify the claim that we are here dealing with a single person, so although this situation seems definitely possible, it would not be a counter-example to (Conj). On the other hand, if we are dealing with a single person, we must be sure that the distinction between the person-like entities is sufficient to

override what I described as the *prima facie*, default status of (Conj). S-in-a-happy-mood and S-in-a-sad-mood may, in some ways, be like different people, but not enough to allow one to disavow the beliefs of the other. Hence (Conj) would apply by default.[27]

Let us suppose that there are intermediate cases between these two extremes. (If there are not, then (Conj) can remain unqualified in this respect and the thread of my argument for (CP) can be picked up below in Section 8.) We may suppose that a person's belief set can be, in David Lewis' terminology, "fragmented". The fragments will be sub-systems of belief. A sub-system of S's beliefs is a subset of S's beliefs such that various principles governing belief apply to beliefs within the subset but do not necessarily relate the beliefs in the subset to beliefs outside of it.[28] (Note, for future reference, that the set of a person's beliefs can form a sub-system of itself according to this definition.) Instead of (Conj), I propose its restriction to sub-systems:

> (Conj*) If S has the beliefs *A* and *B* within a single sub-system of beliefs, then S believes *A and B*.

By qualifying (Conj) in this way, we insulate it from counter-examples of cases in which a person has two beliefs but somehow keeps them isolated from each other.

Sub-systems of belief have often been resorted to with the aim of explaining the existence of contradictory beliefs in a person's belief set. The argument goes: people have contradictory beliefs; one cannot believe a contradiction; unless beliefs are in some way isolated from each other, (Conj) implies that their conjunction is also believed; therefore, there must be sub-systems within a person's beliefs across which (Conj) does not apply. This will mean that if two beliefs are contradictory, that is sufficient to assign them to different sub-systems. I have already argued that people can, and do, believe contradictions. This suggests that we cannot assume that the fact that two beliefs are contradictory is sufficient to show that they belong to different sub-systems.

There is clearly something attractive to the idea that contradictory beliefs do belong in different sub-systems. The hypothesis of sub-systems gains in plausibility to the extent that it can join forces with everyday observations about the ways in which people's beliefs seem to be fragmented. And such everyday judgements seem most pertinent to cases of conflict within a person's beliefs. Take, for example, the famous case of the Japanese astronomer, recounted by Geach.

> [He] seemed to succeed very well in treating the sun alternately as an inanimate natural body whose properties can be investigated by the techniques of mathematical physics, and

as a divinity, the ancestress of the Japanese imperial dynasty; when challenged about the matter by a European colleague, he said 'Here in Europe I know it's all nonsense, but in Japan I believe it'. (Geach 1976, 9)[29]

Clearly, in this case, the assignment of the contradictory beliefs that the sun is a natural object and that it is not a natural object (but rather a divine person) to different sub-systems matches in a satisfying way a more obvious and less theoretical distinction between the astronomer's professional, scientific activity and his religious beliefs. This is amplified by the astronomer's reference to the cultural and ultimately geographical division between Europe and Japan. The sub-systems accommodating the contradictory beliefs match the fault lines independently visible in the person's life.

Nonetheless, I think that the fact that two beliefs are contradictory is neither necessary nor sufficient for assigning them to different sub-systems. Let us deal with the issue of necessity first. The hypothesis of sub-systems of belief seems most plausible when it matches intuitively different areas of a person's mental life. But such different areas might exist without containing any explicitly contradictory beliefs. One might simply compartmentalize one's thoughts in an extreme way. Also, where two contradictory beliefs are assigned to different sub-systems, they may each attract around them all sorts of other supporting beliefs that do not contradict each other.

That contradictoriness is not sufficient for two beliefs to be assigned to different sub-systems is shown by the existence of paraconsistent logicians. Reasoning about Russell's Paradox, Graham Priest came to believe that the set of all sets that are not members of themselves is a member of itself, and that it is not a member of itself. But it would be absurd to assign those beliefs to different sub-systems when it was the very same piece of reasoning that led him to each of them, and when he explicitly believes their conjunction![30]

It might be thought that believing that contradictions cannot be true is itself necessary and sufficient to make the contradictoriness of beliefs sufficient for assigning them to different sub-systems. In other words, contradictoriness is a sufficient condition for different sub-systems for all and only those who believe that contradictions cannot be true. It is precisely because Priest does not believe that contradictions cannot be true that he is led by the same piece of reasoning to contradictory conclusions. The Japanese astronomer, by contrast, (let us stipulate) does believe contradictions cannot be true, and hence his contradictory beliefs should be assigned to different sub-systems. But in fact, believing that contradictions cannot

be true seems neither necessary nor sufficient for automatic assignment of contradictory beliefs to different sub-systems.

That it is not necessary is shown by the case of the paraconsistent pagan. Imagine a pagan who believes the supreme god is a great lion. One of the things that makes this lion mad is if people go around saying that it is both material and not material, real and not real, completely and essentially leonine and not completely and essentially leonine, etc. In fact, the pagan thinks he is reliably informed that the great lion has become so offended by sloppy thinking that he has issued a pronouncement on the subject: I am that I am. This is interpreted by the priests to mean that for no property F is the lion both F and not F. Years go by, and our pagan takes up the study of logic. For various purely logical reasons, he concludes that contradictions can be true. He is very impressed by Russell's Paradox, and comes to accept that there is a property G – being such that the set of all sets that are not members of themselves is a member of itself – such that for all objects O, O is G and O is not G. (He reasons on the basis of a complex argument that may or may not be fallacious about the relations of falsity and negation.) This, of course, contradicts the revealed truth of religion. Logic, he comes to believe, does not apply to god, and henceforth he is careful to keep his religious and logical beliefs quite separate.

The issue of whether believing that contradictions cannot be true is sufficient for contradictory beliefs to be *ipso facto* assigned to different sub-systems is more complicated, and I defer discussion of it until Section 9 below, when I re-address the topic of believing contradictions.

I have argued that, because of the possible existence of multiple sub-systems of belief within a single person's belief set, (Conj) should be replaced by (Conj*). (Conj*) will be shown, in the next section, to be a premise in an argument for (CP). There are, however, further qualifications that could be made to (Conj*), for those who feel it is still too strong in various ways. (I shall discuss the qualifications with respect to (Conj) for the sake of simplicity. Nothing here will hinge on the difference between (Conj) and (Conj*).)

One such qualification of (Conj) is for those who see it as an empirical generalization and hold that it should be taken probabilistically, as

> (Conj-Prob) If S believes *A* and S believes *B*, then probably S believes *A and B*.

This would lead ultimately to a probabilistic counterpart of (CP), (CP-Prob), according to which rational beings probably believe the conjunctions of their beliefs.

Other ways of qualifying (Conj) might be adopted in response to an objection to (Conj) made by Richard Foley. He writes:

> If persons have a very large number of beliefs, [(Conj)] can be true only if persons have incredibly complex beliefs. But in order to believe something *p* . . . a person must be psychologically able to consider *p*. Yet, it seems very unlikely that any person is ever able to consider the conjunction of the contents of all his beliefs. Thus, [(Conj)] seems to be false. (1979, 249, fn. 7)

This objection can be easily disposed of if one takes (Conj) as resting on the claim that the state of believing a conjunction just is the state of believing its conjuncts. In that case, all that is required for someone to be psychologically able to consider a conjunction is that she be psychologically able to consider its conjuncts. But for anyone who does not like this way of taking (Conj), Foley's objection presents some serious obstacles and would require weakening of (Conj). This could probably happen in a variety of ways, perhaps depending on a fuller statement of Foley's argument. Here is a sketch of one such way.

First we characterize recursively a notion of a belief's difficulty, the strain it puts on one's psychological resources to consider that belief. The difficulty of logically simple beliefs is a function of a range of factors, perhaps including length, the difficulty of the concepts involved, and other things. The difficulty of logically complex beliefs is some function of the degree of logical complexity of the belief and the difficulty of its components. Foley's objection to (Conj) could be put by saying that for each creature, there is a limit to how difficult a belief can be if it is to be psychologically considerable by that creature. (Conj) would imply that creatures with realistic sets of logically simple beliefs would have complex beliefs the degree of difficulty of which would exceed their limit. In response we can amend (Conj) to:

> (Conj-Diff) If S believes *A* and S believes *B*, and the degree of difficulty of *A and B* does not exceed S's limit, then S believes *A and B*.

This, in turn, would support a version of (CP), (CP-Diff), according to which it would be rational to believe the conjunction of one's beliefs, so long as those conjunctions did not exceed one's difficulty-limit.

(CP-Prob) and (CP-Diff) are certainly much weaker than (CP). They do not represent capitulation, however, to the original opponents of (CP). Neither the Lottery nor Preface Paradox, for example, in any way depended on notions of psychological difficulty, as discussed, or on the implausibility of exceptionless versus probabilistic principles. In what follows, I shall continue to work with (Conj*), to derive (CP). But readers should feel free to substitute weaker versions of these, if they feel so inclined.

8.

Suppose, then, that we accept that a person's beliefs can be fragmented into sub-systems, and that these sub-systems are correctly characterized by by (Conj*). How does this connect with (CP)? (CP), after all, is expressed as a principle that applies to all of a person's beliefs. Should we simply abandon it in that form and substitute a version of it relativized to sub-systems of belief? Not at all. The connection between (Conj*), a principle about beliefs within a sub-system, and (CP), a principle about the rationality of a person's beliefs *in toto*, is made through the principle:

(R) If S is rational, then S's belief set is a sub-system of itself.

(R) is another way of saying that if someone is rational, then her beliefs are not fragmented into two or more sub-systems across which principles like (Conj*) do not apply. If a belief set is a sub-system of itself, then principles like (Conj*) that apply within a sub-system of belief will apply across the person's entire belief set.

If (R) is accepted, we have a simple argument for (CP). Suppose S is rational. By (R), S's beliefs form a single integrated system. If S's beliefs form a single integrated system, then whenever S believes *A* and believes *B*, it follows, by (Conj*), that S believes *A and B*. Hence, if S is rational, then if S believes *A* and believes *B*, S believes *A and B*. And that is what (CP) says.

It turns out, then, that (CP) depends on (R). It is this that makes the discussion of what may seem a fairly dry and limited question – should a theory of rational belief contain (CP)? – into a deep and interesting philosophical problem. For the decision of whether or not to accept (R), I contend, will be determined by fundamental aspects of one's philosophical outlook, aspects that involve what one might call philosophical temperament as much as philosophical argument.

The rallying cry of those who reject (R) is likely to be a striking passage from Walt Whitman's "Song of Myself" (1855, ll. 1314–6):

> Do I contradict myself?
> Very well then . . . I contradict myself;
> I am large . . . I contain multitudes.

Here we have the idea that contradictory beliefs co-exist within a person as parts of different sub-systems. A single person – "I" – is a large container of many person-like entities. But more important is the unapologetic, even defiant, attitude towards psychic fracture. The lines come from an incantatory celebration of a self that is described over and over again in

contradictory and conflicting ways. It is as if life demands this internal multitudinousness.

Descending from these elevated generalities, opponents of (R) might draw on various different philosophical currents: post-modern theories of the death of the subject, Dennett's multiple draft model of the mind, modularity in cognitive psychology,[31] and various others. I will briefly describe one such line of thought, since its connection with the rejection of (R) may not be immediately obvious. William James made a famous response to a paper by Clifford in which the latter had argued the more or less traditional view that the strength of one's beliefs ought to be proportionate to the strength of one's evidence. Clifford put a new spin on this by arguing that failure so to proportion one's beliefs was a moral as well as an epistemic failing. James, echoing Kierkegaard, responded that life often calls us to fix on a belief even with insufficient evidence. Sometimes it is important to take a chance at being right, even if we thereby give up the chance of being rational (in the sense defended by Clifford). The kinds of cases which James thought called for this bold policy would include just the kinds of areas where we are likely to be tempted by hypothesizing sub-systems of belief – religious faith, being in love, and so on. Since Clifford's principle requires the rational influence on each belief of all other relevant beliefs, James' denial of it can be seen as a claim that people should, in some cases, isolate beliefs from the rational influence of other beliefs that may be relevant as forms of evidence. In other words, James could be seen as denying (R).[32]

On the other side, there are those who support (R). As one such supporter, Stalnaker, puts it, "one cannot agree to disagree with oneself" (1984, 84). Although humans, as limited and imperfect creatures, may well find within themselves pockets of belief isolated from the rational effects of the rest of their beliefs, this cannot be a healthy or acceptable situation. The goal is surely to arrive at a single, rationally integrated representation of the world (or as much of one as we can). Even Freud, the theorist *par excellence* of divisions within the mind, described the goal of the therapeutic process of analysis with the words "Where Id was, there Ego shall be", thus implying the undesirability of psychic division.

Since we are debating about what is rational, and rationality is such an elastic and vague concept, it is difficult to imagine arguments for or against (R) that will not be thought simply to beg the question by adherents of the opposing view. For example, one might argue for (R) by pointing out that to the extent that there are multiple sub-systems within a single person's beliefs, there will be ways in which that person fails to put together pieces of information that when put together can support the drawing of

various conclusions. So multiple sub-systems seem to imply a failure of optimization of information available to a person – and surely that is to be counted as a failure of rationality. But naturalized epistemologists, such as Cherniak and Kornblith, will respond that full use of information available would overload any finite creature, and that a divide and conquer strategy is actually the best option for any creature less than God – so it cannot be irrational to pursue that strategy. There are good points on both sides here, but I have suggested that in the last analysis, one's reaction to (R) will come down to a bedrock level of philosophical temperament or vision.

9.

I want, finally, to return to the issue of believing contradictions. As I mentioned, many of those who think that people's beliefs can fall into different sub-systems think this because, or partly because, they are sympathetic to (Conj), accept that people can have contradictory beliefs, but do not think that people can actually believe contradictions.[33] I do not agree. It seems to me plain that people do believe contradictions. These cases fall into two sorts: those who believe them thinking that they might be true (paraconsistent logicians, religious mystics, and others); and those who believe them and think that there is something wrong with believing them. The following remarks apply only to the second sort of cases. As an example of such a case we may return to the Japanese astronomer. He believes that the sun is a natural object and that it is not. This by itself does not entail that he believes the contradictory belief that the sun both is and is not a natural object, since each of the components may exist within a different sub-system; but his own diagnosis of his situation ("Here in Europe I know it's all nonsense, but in Japan I believe it") implies that he does, by the very act of giving this description, come to believe their conjunction as well. He is, in other words, entirely aware of the two conjuncts and could only repudiate belief in their conjunction by repudiating belief in one of the conjuncts.

This case presents a problem. All beliefs, including the Japanese astronomer's belief in the contradiction, exist within a sub-system. (This is true since a belief set can be a sub-system of itself.) It seems intuitively right to say that where a belief in a conjunction exists within a given sub-system, beliefs in each of the conjuncts also exist within that same sub-system. In that case, we would have to say that the belief that the sun is a natural object and the belief that it is not exist within the same sub-system. But this conflicts with our initial analysis of the situation, which seemed to find it natural to assign them to different sub-systems. (Indeed,

this case was one which was used to make the hypothesis of different sub-systems most plausible.) It seems that in cases where someone does believe a contradiction, the attempt to see the contradictory beliefs as belonging to different sub-systems must be undermined. (It is for this reason, and because it is possible to believe a contradiction even though one thinks that contradictions cannot be true, that I did not want to say, above, that thinking that contradictions cannot be true is sufficient for contradictory beliefs to be *ipso facto* assigned to different sub-systems of belief.)

I am not sure how to deal with this issue. At some point, we will want to introduce into a theory of rational belief considerations of the dynamics of belief change. Any hypothesis about sub-systems of belief will need to be complicated accordingly. In dynamic terms, there is surely something unstable about believing a contradiction and thinking that contradictions cannot be true. To put it dramatically, we might say that it is a state that contains within it the seeds of its own destruction. (This is doubtless why many philosophers have asserted that it is impossible to believe a contradiction.) Something of this dynamic instability is suggested by the telling phrase "being in two minds about something". Take the Japanese astronomer. Perhaps the division of beliefs into sub-systems has hitherto allowed him to evade the fact that he believes that the sun is a natural object and that it is not. Prior to his being confronted about it, it would not be correct to say that he was in two minds about whether the sun is a natural object, though this is as close as he will get to making that phrase literally true. When he is confronted by the situation, he comes to realize that he has contradictory beliefs, and in so doing, comes to believe their conjunction. He does not just give up one of the beliefs, or become agnostic on the whole issue. But what he realizes is not just that he has had two contradictory beliefs all this time. Given that he is unwilling to abandon either belief, this is painful enough. But what is worse is the realization that these contradictory beliefs were maintained simultaneously by a process of psychic division. He comes to recognize the existence of a division within himself, a falling away from the ideal of (R). At this point, the Japanese astronomer is like the cartoon character who runs off the edge of a cliff and remains impossibly suspended to contemplate his imminent fall. The two sub-systems are seen to be there, and yet this very acknowledgment of their existence means that they can no longer be maintained as distinct sub-systems. He is now inevitably faced with the task of integrating them. This is the state of being in two minds about something – a state that obtains only when enough psychic unity has been achieved to make it literally false.

ACKNOWLEDGEMENTS

Thanks to Tyler Burge, David Craig Harlan, Dana Nelkin, Giovanna Pompele and an anonymous referee for this journal for help in improving this paper.

NOTES

[1] As do, for example, Hintikka 1962, Stalnaker 1984 and Swain 1970.

[2] The internal criticisms have been made strongly by Foley 1979, 1993; Kyburg 1970. The external criticisms are found implicitly in Cherniak 1986; Kornblith 1989 and explicitly in MacPherson 1993.

[3] There is a minority challenge to this view, stemming from those logicians who believe that contradictions can be true. See Priest 1985/6.

[4] Foley 1993, pp. 165ff presents the Preface Paradox as part of his sustained attack on (CP). The discussion in the text is largely based on remarks in Kyburg 1970, p. 59, though he is not there explicitly discussing the Preface Paradox.

[5] Neither *(1) and (2)* nor *(3) and (4)* has the form of a contradiction as stated. To show that (CP) led someone to a belief of the form *A and not-A*, we would need to invoke another principle, to the effect that if S is rational, then if S believes *A* and S believes that *A* is logically equivalent to *B*, then S believes *B*, along with a stipulation that a rational person does believe that (2) is logically equivalent to "not everything I believe is true" (and *mutatis mutandis* for (4)). Let all this be granted.

[6] This is related to the Liar Paradox. The Liar Paradox, of course, depends on self-reference. Here, the possibility of self-reference is raised by the second-order nature of Bel [(2)]; its actual occurrence is produced by the stipulation that none of my other beliefs are false.

[7] Clearly, there are various options and strategies that could be appealed to here to attempt to avoid this paradox.

[8] I deal at greater length with the form that recognition of one's fallibility should take, and its relation to the Preface Paradox, in Evnine (unpublished).

[9] See Foley 1979, p. 252, n. 12.

[10] We may want to make some exceptions to a general belief in fallibility to accommodate Cartesian-style, infallible beliefs. But (2′) is not one of these. It is a straightforward empirical belief and exempts itself from its own scope only to avoid paradox. It is a case of special pleading.

[11] The Lottery Paradox was first introduced in Kyburg 1961 (p. 197) as an argument against (CP). It is presented and defended at greater length in Foley 1979 and 1993 and Kyburg 1970. See Ackermann 1972, pp. 39–50, for a non-partisan discussion of the paradox.

[12] The response is described in Ackermann 1972, pp. 47–50.

[13] To be more precise, the acceptance rule implies that, in the case of the proposition that a given ticket will not win, its high probability is sufficient for it to be rational to believe it. This does not mean that the acceptance rule must hold that high probability is *always* enough for rational belief.

[14] Of course, it may be rational to act *as if* one did believe it. This is why, in a lottery with a large number of tickets, it is rational not to buy a ticket at all, i.e., it is rational to act as if you believed that for any ticket you can buy, it will not win.

[15] Another common response to the paradox is that we should adopt a theory of degrees of belief and accord a degree of belief less than 1 to the various propositions that ticket *n* will not win (see Ackermann 1972, pp. 42–7 and Stalnaker 1984, p. 91). While my response is compatible with this, it does not entail it. My response is thus not affected by the probabilistic version of the Lottery Paradox that Kyburg has given against the degrees-of-belief response (1970, pp. 57–8).

[16] For a fuller exposition of this, see Kyburg 1970, pp. 55–61.

[17] Some of these examples are taken from Priest 1985/6.

[18] Let me just note here that this invocation of facts of awareness cannot stand behind the distinction between Strong and Weak Consistency as it features in the Preface or Lottery Paradoxes. For the use of those paradoxes as arguments against (CP) is not taken to depend on any lack of awareness of the inconsistency of the those beliefs which, if conjoined by (CP), would yield belief in a contradiction.

[19] See Cherniak 1986 and Kornblith 1989.

[20] See Lewis and Papadimitriou 1978 for a good introduction to these issues.

[21] Cherniak 1986, p. 7 expresses this objection to what he calls theories of ideal rationality. See also MacPherson 1993.

[22] Kyburg 1970, pp. 55, 58–9. Kyburg calls the principles the "Weak Deduction Theorem" and the "Principle of Deductive Closure" respectively, and expresses them in terms of bodies of reasonably accepted statements, rather than rational believers.

[23] This reliance on *a priori* considerations thus marks a difference between the cases of closure conditions and consistency conditions on rational belief. I argued above that, assuming that there was independent evidence of people's beliefs, it would be an *empirical* question what kind of consistency constraint, if any, would work best in an RTBA. This difference is, I believe, a consequence of the fact that the requirements of a closure condition are open-ended in a way that the requirements of a consistency condition are not. Hence it is possible to say that a closure condition in an RTBA is *a priori* implausible.

[24] That a conjunctive belief requires some mental act of bringing together the two conjuncts is argued for by Armstrong 1973, p. 107.

[25] See Routley and Routley 1975, pp. 210–15 and Stalnaker 1984, pp. 82–4.

[26] The approaches to belief evinced in Davidson 1984, Dennett 1987, and Stalnaker 1984 would all be conducive to seeing (Conj) in this light. Routley and Routley 1975 explicitly discuss the impact of accepting (Conj) on a theory of the nature of belief – cf. pp. 211–2.

[27] Here would be the place, in a longer treatment of the subject, for discussion of such phenomena as the unconscious, self-deception, multiple personality disorder and so on.

[28] A number of philosophers have invoked this idea of partitioning, or fragmentation within a corpus of beliefs to account for various phenomena relating to irrationality in belief and action. This is often, though not always, directly related to a failure of (Conj) to apply across divisions. See, for example, Davidson 1982, Lewis 1982, Pears 1984, ch. V, and Stalnaker 1984, ch. 5.

[29] The words attributed to the Japanese astronomer might suggest that his is a case of someone whose beliefs simply change according to where he is. In that case, it would not be an example of someone with contradictory beliefs at a given time. Since it was originally given as an example of just that phenomenon, I do not take it as a case of change of belief according to location.

[30] See Priest 1985/6, p. 103.

[31] See, for example, respectively, Dunning 1993; Dennett 1991; Kornblith 1989, pp. 211–4.

[32] One might respond by saying that James is not denying (R), a condition on what is rational, but denying that we should be rational. However, I have suggested that at least one of the functions of the concept of rationality in both (CP) and (R) is to represent the epistemic ideal. This, I believe, would give a common meaning to what James rejects and what (CP) and (R) express. My understanding of James' position and the issues of this paragraph has been much influenced by Adams 1984.

[33] Lewis 1982 and Stalnaker 1984 both seem to fit this description.

REFERENCES

Ackermann, R. J.: 1972, *Belief and Knowledge*, Doubleday, Garden City, NY.

Adams, R.: 1984, 'The Virtue of Faith', in Robert Adams, *The Virtue of Faith*, Oxford University Press, Oxford (1987).

Armstrong, D. M.: 1973, *Belief, Truth and Knowledge*, Cambridge University Press, Cambridge, MA.

Cherniak, C.: 1986, *Minimal Rationality*, MIT Press, Cambridge, MA.

Clifford, W. K.: 1879, 'The Ethics of Belief', in W. K. Clifford, *Lectures and Essay*, Macmillan, London.

Davidson, D.: 1982, 'Paradoxes of Irrationality', in R. Wollheim and J. Hopkins (eds.), *Philosophical Essays on Freud*, Cambridge University Press, Cambridge.

Davidson, D.: 1984, *Inquiries into Truth and Interpretation*, Clarendon Press, Oxford, UK.

Dennett, D.: 1987, *The Intentional Stance*, MIT Press, Cambridge, MA.

Dennett, D.: 1991, *Consciousness Explained*, Little, Brown and Company, Boston.

Dunning, W.: 1993, 'Post-Modernism and the Construct of the Divisible Self', *British Journal of Aesthetics* **33**, 132–41.

Evnine, S.: unpublished, 'Learning from One's Mistakes: Epistemic Modesty and the Rationality of Conjunctive Belief'.

Foley, R.: 1979, 'Justified Inconsistent Beliefs', *American Philosophical Quarterly* **16**, 247–57.

Foley, R.: 1993, *Working Without a Net: A Study of Egocentric Epistemology*, Oxford University Press, New York.

Geach, P.: 1976, *Reason and Argument*, Blackwell, Oxford, UK.

Hintikka, J.: 1962, *Knowledge and Belief. An Introduction to the Logic of the Two Notions*, Cornell University Press, Ithaca.

James, W.: 1897, 'The Will to Believe', in William James, *The Will to Believe*, Longmans, New York.

Kornblith, H.: 1989, 'The Unattainability of Coherence', in John Bender (ed.), *The Current State of the Coherence Theory*, Kluwer, Dordrecht, The Netherlands.

Kyburg, H.: 1961, *Probability and the Logic of Rational Belief*, Wesleyan University Press, Middletown.

Kyburg, H.: 1970, 'Conjunctivitis', in Marshall Swain (ed.), *Induction, Acceptance, and Rational Belief*, Reidel, Dordrecht, The Netherlands.

Lewis, D.: 1982, 'Logic for Equivocators', in David Lewis, *Papers in Philosophical Logic*, Cambridge University Press, Cambridge (1998).

Lewis, H. and C. Papadimitriou: 1978, 'The Efficiency of Algorithms', *Scientific American* **238**, 96–109.
MacPherson, B.: 1993, 'Is it Possible that Belief Isn't Necessary?', *Notre Dame Journal of Formal Logic* **34**, 12–28.
Pears, D.: 1984, *Motivated Irrationality*, Clarendon Press, Oxford, UK.
Priest, G.: 1985–6, 'Contradiction, Belief and Rationality', *Proceedings of the Aristotelian Society* **86**, 99–116.
Routley, R. and V. Routley: 1975, 'The Role of Inconsistent and Incomplete Theories in the Logic of Belief', *Communication and Cognition* **8**, 185–235.
Stalnaker, R.: 1984, *Inquiry*, MIT Press, Cambridge, MA.
Swain, M.: 1970, 'The Consistency of Rational Belief', in Marshall Swain (ed.), *Induction, Acceptance, and Rational Belief*, Reidel, Dordrecht, The Netherlands.
Whitman, W.: 1855, 'Song of Myself', in Walt Whitman, *Leaves of Grass*, Penguin Books, Harmondsworth (1986).

Department of Philosophy
California Polytechnic State University
San Luis Obispo
CA 93407
U.S.A.