[*Australasian Journal of Philosophy* 72(1) (March 1994): 89-102]

Are There More Than Minimal A Priori Constraints on Irrationality?

John Biro and Kirk Ludwig
Department of Philosophy
University of Florida
Gainesville, FL 32611-8545

Our concern in this paper is with the question of how irrational an intentional agent can be, and, in particular, with an argument Stephen Stich has given for the claim that there are only very minimal a priori requirements on the rationality of intentional agents. The argument appears in chapter 2 of *The Fragmentation of Reason.*[1] Stich is concerned there with the prospects for the 'reform-minded epistemologist'. If there are a priori limits on how irrational we can be, there are limits to how much reform we could expect to achieve. With this in mind, Stich sets out to determine what *a priori* limits there are on irrationality by examining `a cluster of influential arguments aimed at showing that there are conceptual constraints on how badly a person can reason' (p. 30). Stich aims to remove the threat of *a priori* limits on the project of reforming our cognitive practices by showing, first, that these influential arguments are bad arguments, and, second, that at best there are only minimal constraints on how irrational we can be.[2] We aim to show three things. The first is that Stich's own arguments against strong *a priori* limits on how badly a person can reason are unsuccessful, because Stich fails to take into account that the concept of rationality is an epistemic, not just a logical concept, and because he fails to take into account the connection between having a concept and being able to recognize conceptually simple inferences involving the concept. The second is that the position Stich argues for, on the basis of Richard Grandy's principle of humanity, turns out not to be distinct from the one he rejects. The third is that, in any case, the position that Stich rejects in order to preserve some scope for the project of improving our reasoning is not only no danger to that project but must be presupposed by it.

I

Stich distinguishes three positions on the question of what limits there are on how irrational we can be; he calls them the *perfect rationality view*, the *fixed bridgehead view,* and the *minimal rationality view*.[3] Stich argues *against* the first two views and *for* the third. He gives only necessary conditions for perfect rationality: we are *perfectly rational* only if our beliefs form a consistent and deductively closed set. The fixed bridgehead view requires neither consistency nor deductive closure, but only that all intentional agents share some special set of true beliefs and rational inferences. The minimal rationality view requires only that every agent share some inferential capacities and beliefs with an ideally rational agent, but not that there be any overlap between different agents who fall short of ideal rationality.

The only points we wish to make about the perfect rationality view are that (1) it is extremely unlikely that anyone has ever held this view, and (2) Daniel Dennett, Stich's primary example of a defender of this position, does not hold it.[4]

Our main concern here is with the status of Stich's arguments against the fixed bridgehead view, which holds that

> there is a special class of inferences and stimuli-induced beliefs which a subject must manifest if his mental states are to admit of any intentional description at all. (p. 40)

What is Stich's argument against the fixed bridgehead view? Stich adapts an argument from Christopher Cherniak, which takes the form of a thought experiment. We are asked to imagine a `hypothetical people' whose `feasibility ordering' for inferences is inverted with respect to ours. The feasibility ordering of a person's inferences is the ordering of the possible inferences he can make in terms of the difficulty he experiences in making them.[5] If we grant the possibility of a people whose feasibility ordering is so inverted, that is, who find the inferences we find hard, easy, and the inferences we find easy, hard, then, Stich argues, we must give up the fixed bridgehead view of rationality.

In what sense are `inference' and `difficulty' used here? As far as the former is concerned, we are presumably concerned with rational inferences, for everyone sharing some special set of irrational inferences would hardly go to show there was a fixed bridgehead of rationality. What about `difficulty'? On one reading, this is a matter of the *psychological* difficulty that an agent experiences in making a certain belief transition. An inference *A* is thus more psychological difficulty than an inference *B* for an agent just in case he experiences less resistance in making *B* than in making *A*. But this is not the natural reading. For on this reading, one might recognize that, first, *Q* followed from if *P* then *Q*, and *P*, and believe both *P* and if *P* then *Q*, but be *psychologically* unable to infer that *Q* (perhaps because of the particular content of *Q*). Would this show that there is no fixed bridgehead of rationality shared by all rational agents? Presumably not. One might also find the inference from *P* and if *P* then *Q* to *Q* psychologically very easy, but be completely unable to see that *Q* followed from if *P* then *Q* and *P*. But this would clearly not constitute a shared bridgehead of *rationality.* Since what is relevant here is rationality, we should take the kind of difficulty here to be difficulty in *seeing that an inference is a reasonable or valid one.*

In the characterization of the fixed bridgehead view above, all intentional agents were required to share a special class of inferences and stimuli-induced beliefs. Martin Hollis, from whom Stich borrows the expression `fixed bridgehead', discusses for the most part beliefs about perceptually salient features of a shared environment.[6] In contrast, Stich talks mostly, even exclusively, about inferences. What is the special class of inferences which the fixed bridgehead view requires? While Stich is not very explicit about this, it seems reasonable to take them to be inferences which are conceptually simple, that is, inferences which are not to be decomposable into elements which are conceptually simpler, or, at least, which do not have extensive analyses into their simple elements. It will be inferences of this sort to which we will appeal in order justify longer and more complex inferences. Thus, for example, the fixed bridgehead of rationality required of all agents ought to include the ability to understand simple inferences such as those involving *modus ponens.* We shall understand the fixed bridgehead view in this way in the remainder of our discussion.

For Stich's thought experiment to establish that there is no fixed bridgehead of rationality, the feasibility ordering of our hypothetical others must be such that they cannot see that any conceptually simple inferences we recognize, or can recognize, as valid or reasonable are in fact so. It is not at all clear that it would be possible to interpret intentional agents whose feasibility ordering was so different from ours. For in that case, they would presumably not exhibit patterns of reasoning which would allow us to identify logical constants in their language, and it is very difficult to see how interpretation could proceed if we were unable to do this. *Perhaps* this could be done, but it is certainly not obvious that it could. If it could not be done, then *we* would never be able to recognize that there were agents such as the ones Stich asks us to imagine. Stich claims that even if this is true, it is `irrelevant to the issue at hand' because `the fixed bridgehead view is a claim about the conceptual connection between intentional characterizability and the disposition to draw certain inferences' (p. 43). But it is not irrelevant, if the issue is whether or not there are limits on how much reform of our cognitive practices is needed and possible. Even this much would be enough to show that there are some limits to the scope of the reform-minded epistemologist's project, since the reform-minded epistemologist is obviously concerned with intentional agents he can recognize as such. Since, as we will see later (section IV), Stich himself seems to accept the claim that we can recognize others as intentional agents only if we find them sufficiently similar to us in their patterns of reasoning, as far as the concerns of the reform-minded epistemologist go, we could stop here. If we hold further that intentional agents who share with us a large number of concepts must be interpretable by us, it follows that such a radical difference in feasibility orderings as we have described above is not merely not recognizable, but not possible.

II

We do not intend to rest our criticism of Stich's argument on the above assumptions, however. Our criticism is that the thought experiment upon which it is based is internally incoherent in two different, but connected ways.

The first difficulty for Stich's argument arises because of what is required for agents with intentional states to be making rational inferences. The difficulty lies in the requirement that our hypothetical people be unable to recognize as valid those inferences which are most transparent to us; these are the inferences which are conceptually simple, those into which more complex ones are decomposable. This requirement is in conflict with Stich's assumption that they are in fact making *rational* inferences, that is, that the transitions they make between sets of beliefs are justified inferences in addition to being logically valid inferences. Since the hypothesis that their ability to recognize inferences as valid is inverted with respect ours presupposes that they *are* making rational inferences, it follows that the thought experiment is incoherent.

The reason our hypothetical reasoners cannot be making rational inferences is simple. If they are making rational inferences, then it must be the case that (a) their inferences can be broken down into conceptually simpler inferences each of which is valid, and (b) they are able to recognize that each of these simpler inferences is valid. Condition (a) is required because given that *we* are able to make conceptually simple inferences, their inferences must be conceptually complex. We have supposed that they do not share with us a bridgehead of conceptually simple inferences, their feasibility ordering being inverted with respect to ours. Consequently, if the inferences they make are valid, they must be decomposable into conceptually simpler valid inferences. Condition (b) is required because unless it is met, our hypothetical people will be unable to recognize the valid conceptually simple inferences that make their more complex inferences valid. But then they cannot be said to have made a reasoned belief transition, that is, to have justifiably inferred one belief from another, for *they are not in a position to recognize that their inferences are valid*. They would be in the position of someone who, e.g., is inclined to move from the belief that *P is the power set of S* to the belief that *P is larger than S* but who does not and *cannot* understand what makes this inference reasonable. If he does not and cannot understand why this is a reasonable inference, then he did not make a reasonable or justified inference. Stich's argument, then, faces the following dilemma: either our hypothetical people are able to make or at least recognize, even if with psychological difficulty, conceptually simple inferences, or they cannot be making rational inferences at all, contrary to our original hypothesis. However, to maintain that they have intentional attitudes yet are not engaging in rational reasoning is not to deny just the fixed bridgehead view, but to deny that our imagined `reasoners' are rational at all, i.e., to *deny* they are even minimally rational. But that every intentional agent is to some degree rational a view which Stich says he *accepts*. On the other hand, to accept that they can at least recognize the simple inferences we can make is tantamount to *accepting* the fixed bridgehead view, which Stich says he *rejects*. Thus, there is no coherent interpretation of the thought experiment that has the result that Stich wants.

The mistake here is to confuse the concept of a valid inference with the concept of a justified inference, or, in other words, to confuse the concept of logical support with the concept of epistemic support. These are not unconnected, but they are not the same. Stich assumes that because by hypothesis our imaginary `reasoners' are making valid inferences,[7] they must be making rational or justified inferences. But this is not so. The distinction can be seen in the following example. Suppose that S believes Peano's axioms for arithmetic, and infers Goldbach's conjecture from them. (Why he infers the one from the other is immaterial. Perhaps he is `hardwired' this way, or is involved in some terrible confusion, or perhaps a knock on the head has simply given him the feeling that the one is a consequence of the other.) Goldbach's conjecture may very well be a consequence of Peano's axioms. If it is, then S has made a valid inference. But since neither S nor anyone else knows whether Goldbach's conjecture is true, and so *a fortiori* neither S nor anyone else knows whether Goldbach's conjecture is a consequence of Peano's axioms, it follows that, even if it is, S has not made a justified inference. Not every *valid* inference, then, is a *justified* or *rational* inference.

A minimal necessary (though not sufficient) condition on S's inference being justified is that S *be able to recognize* that it is. Short of *magical* epistemic access to the necessity of the one given the other, he could do this only by constructing a proof of Goldbach's conjecture from Peano's axioms, or from other truths he knows, which would require him to be able to recognize the validity of the simpler inferences which make up the proof. Thus no one who is unable to recognize conceptually simple inferences as valid can make justified inferences, even if his inferences are as a matter of fact valid.

The general moral here, of interest independently of its relevance to Stich's argument, is that rationality is an epistemic concept, not a logical one. Rational inference is not just a matter of the logical and causal relations among beliefs, but also of the reasoner's epistemic position with respect to those relations, and, in particular, to the logical relations. A rational inference is one that the reasoner can at least recognize as valid or at least as one more likely than not to lead from true beliefs to true beliefs.[8]

But the difficulty with the thought experiment goes still more deeply than this first problem indicates. We have concentrated on the connection between making a rational inference and being in an epistemic position to see that it is a reasonable inference. This is, so far, to grant that our hypothetical others might yet have intentional states, with contents which employ concepts we share, though they are not even minimally rational. If this were so, then we would still have an argument against the fixed bridgehead view, even if it is an embarrassment for Stich, who that argues there are *some* constraints on irrationality. But there is a serious flaw in this picture. Not being able to recognize conceptually simple inferences involving the concepts they share with us, our hypothetical others cannot be said to be making rational inferences when they make complex inferences. But what so much as gives us the right to talk about their sharing concepts with us, as this presupposes? For having a concept is itself an epistemic state. What constitutes our having a certain concept is our being able to recognize the validity of conceptually simple inferences involving the concept. Mastery of a concept comes to nothing more than our knowledge of how to employ it in its most simple and transparent uses. Conceptually simple inferences are those to which possession of the concepts involved gives us first access. To say that our imagined others make complex inferences involving concepts they share with us, but are unable to recognize as valid conceptually simple inferences involving those concepts, is at once to attribute to them concepts while denying that they meet conditions necessary for their possession.

This does not immediately establish, as one might think, the fixed bridgehead view. That one can have a given concept only if one is able to recognize simple inferences involving it secures immediately that anyone who shares our concepts with us also shares with us the ability to recognize conceptually simple inferences involving those concepts. However, the fixed bridgehead view, as articulated above, is stronger than this, since it requires that all intentional agents share a fixed bridgehead of beliefs and inferences. Nothing in the considerations above show that all intentional agents must share some concepts with each other. Nonetheless, it seems to us that this is an important result, which establishes a relativized version of the fixed bridgehead view: we must share a bridgehead of rational inferences with anyone who shares our concepts with us.

These two criticisms of the thought experiment both focus on failures to recognize that the notions employed in its description have epistemic conditions for their application. A rational inference is not just a valid one, but one which the person making the inference is justified in making. Unable to recognize what makes their inferences justified, our imagined others could not be making rational inferences. But the feature of the thought experiment that underlies this failure engenders a more serious one, for *having a concept* is itself an epistemic notion. One has a concept only if one has a mastery of how to employ it in its simplest and most transparent uses. This is what the thought experiment denies our imagined others can do. If they made inferences, they would not be rational. But for the reason that they cannot make rational inferences, it turns out they cannot be making inferences involving our concepts at all. The thought experiment self-destructs.

III

We turn now to some objections to our argument to clarify it and to ward off misunderstandings.

The first objection is that the notion of simplicity we are appealing to is relative to a logical system, and that it is certainly possible for an inference which is simple in one system to be complex in another. As an example, consider the inference from `--P' to `P' in a natural deduction system. In a system such as that of Myro,[9] this inference requires one step. However, in a natural deduction system such as that of Mates,[10] the same inference requires five steps. Surely, it may be said, the inference in the first of these systems is simpler than in the second, although they are the same inference. And surely it is possible to imagine that someone could use the latter deduction system rather than the former, and so not make any of the inferences we consider to be simple.

This objection confuses a measure of simplicity based on the number of steps in a formal derivation relative to a formal system of rules of inference with what we have called conceptual simplicity. An argument is conceptually simple if it is easy for anyone who understands the concepts involved to see that it is valid. To test whether the inference from `--P' to `P' is conceptually simple, we would first need to ask whether someone who understands the concepts involved in the premise and conclusion would find it an epistemically easy inference. We think the answer is, obviously, `Yes'. The ability to see that such inferences are correct is constitutive of having the concept of negation.[11] This is so even if it is not easy to *derive* the one from the other in a given formal system. Indeed, the validation of formal rules of inferences depends upon our prior ability to recognize that the application of the rules in simple cases, from which more complex inferences are built, yield valid inferences.

We turn now to two objections to our argument raised by Stich in a comment on an earlier version of this paper.[12] Stich has objected to our argument by saying that he is just using `rationality' in a different sense than we are. What he means by one's being rational, he says, is that one displays reliably inference patterns which are in fact valid (or reasonable); he does not require that the inferences that people make be inferences that they can *recognize* to be valid or reasonable. Thus, in Stich's sense, the person who infers Goldbach's conjecture from Peano's axioms, if it is a valid inference, has made a rational inference (provided, perhaps, that the mechanism by which he does so generally produces formally good inferences). Furthermore, Stich charges, if one were to adopt our characterization of what it is to make a rational inference, it would clearly not be a requirement on having intentional state, since animals and young children do not have the conceptual resources to recognize that any of the inferences they make are rational, yet they clearly have intentional states.

The response to the first of these objections is twofold.

First, if one sets out to show that arguments for the claim that a certain degree of rationality is required for having intentional attitudes, or for interpreting people with intentional attitudes, are unsuccessful, then one's argument fails if one uses `rationality' in a different sense from that in the argument one is attacking. We assume that Stich's opponents are using the term in its ordinary sense, the sense in which in everyday life we charge people with being irrational or praise them for being rational. The ordinary notion, as we have demonstrated, is an epistemic notion. If Stich is not concerned with the concept ordinarily expressed by the word 'rationality', he is not a participant in the debate. You can't win an argument by changing the subject.

Second, this response does not survive the second of the objections we have raised to the thought experiment. Even if Stich is just interested in patterns among belief contents in inferences, and not in justified inferences, the thought experiment is still incoherent, for he is still required to attribute to our imagined others the same concepts we deploy, but under conditions which undermine the possibility of their having those concepts.

The second of Stich's objections was that if we insist that making rational inferences requires being in an epistemic position to tell that the inference is reasonable or justified, then since obviously infants and non-linguistic animals are not in such a position, but have intentional states, our position gives him a simple argument for the possibility of having intentional states without being rational. However, the question whether and in what sense non-linguistic animals and infants have intentional states is controversial. It is our belief that our primary grasp on what it is to have an intentional state is derived from our own case, and that our use of these concepts in the case of animals and infants is by analogy with our own case. Thus, our use of intentional concepts in application to infants and animals requires legitimation. We therefore cannot use our practice of attributing intentional states to non-linguistic animals as a standard for evaluating proposed necessary conditions for the possession of intentional states. We must first investigate how these concepts work in their natural home. If we isolate necessary conditions on the application of these concepts that we decide animals and infants cannot meet, this would show that our application of concepts in those cases was mistaken. Furthermore, as we will see below (section IV), for the argument that Stich himself offers to establish even the minimal rationality view, it is necessary for him to deny that non-linguistic animals have intentional states.

Finally, it worth noting that since many of the arguments for constraints on irrationality proceed by appeal to constraints on interpretation, their conclusions in the first place apply only to linguistic beings. Stich's arguments would not have shown that these conclusions were false for linguistic beings, whether or not they held for non-linguistic beings. This point is of special importance given Stich's announced project, because the issue is the possibility of reforming the reasoning of linguistic beings, not of non-linguistic beings.

IV

We have urged, so far, that Stich's argument against the fixed bridgehead view fails. However, matters are even worse than this for Stich. For, as we will now contend, his own argument for the minimal rationality view in fact establishes the fixed bridgehead view. To see why, we turn to Stich's argument for the minimal rationality view. Initially, Stich characterizes it as follows:

> The view we are left with is that intentional description requires the disposition to draw some reasonable subset of the inferences that would be expected of an ideally rational cognitive agent. (p. 44)

We assume that this means that every agent must draw from his beliefs some of the conclusions that an ideally rational agent with those beliefs would draw.

To argue for the minimal rationality view, Stich invokes Richard Grandy's principle of humanity[13] and argues that it is not just a pragmatic, but a conceptual, constraint on attributing intentional attitudes to others:

> ... adherence to the principle [of humanity] is required if we are to exploit intentional descriptions at all in characterizing a person's mental states. (p. 48)

The principle of humanity holds that

> in choosing a translation we should prefer the one on which `the imputed pattern of relations among beliefs, desires, and the world be as similar to our own as possible' (p. 45).

There is an initial puzzle about why this should be so much as relevant to Stich's aim, which is to establish something about the relation between being an intentional agent and displaying some minimal degree of rationality. For the principle of humanity is a principle about conditions under which it is possible to translate someone else's language into one's own. Even if we discovered that one could do this only if one found others agreeing with one to a certain extent on beliefs and inferences, this would not in itself show anything about all intentional agents. To get a conclusion about all intentional agents, we would have to assume at least

(1) If *A* has intentional states, then *A* speaks a language.

That Stich is committed to this is important, for it undermines his appeal to non-linguistic animals having intentional states in the second objection we considered above in section III. It is still not enough, however, for unless all language speakers can communicate with one another, constraints on shared inferences as a condition on communication would still show nothing about *all* intentional agents. Furthermore, merely assuming that *we* would be able to translate all language speakers would be insufficient, because even if we assume *we* are rational agents, it would be absurd to suppose that we are *ideally* rational agents. But the conclusion that Stich wants is that every intentional agent must share some beliefs and inferences with an ideally rational agent. Thus, we need to make the following two assumptions.

(2) It is possible for there to be an ideally rational intentional agent.

and

(3) Every speaker would be interpretable by every other possible speaker.

Thus, the argument turns out to be a version of Davidson's omniscient interpreter argument[14], and we must wonder what justifies us in making these two assumptions. However, we want to grant Stich these assumptions in what follows, for it turns out that, even granting them, the argument does not establish what Stich wants.

What does the principle of humanity, as stated above, and supplemented with assumptions (1)-(3), actually establish? The principle of humanity, as stated by Stich, does not establish that intentional agents are rational in any degree, for it does not actually entail that there is *any* overlap between the patterns of inferences that one intentional agent makes and the patterns of inferences of intentional agents he interprets. As stated, it is simply a principle for choosing from among otherwise adequate theories of interpretation. It does not rule out any. So if the only theory we had that met all of our constraints did not attribute to some agents we were interpreting any patterns of inferences which we recognized as valid, this principle would not rule it out. The principle as stated here must be strengthened, then, to establish even the minimal rationality view.

As we have noted, Stich borrows the principle of humanity from Richard Grandy, so a natural place to look for an explanation of it is the article in which Grandy introduced it. According to Grandy:

> If a translation tells us that the other person's beliefs and desires are connected in a way that is too bizarre for us to make sense of, then the translation is useless for our purposes. So we have, as a pragmatic constraint on translation, the condition that the imputed pattern of relations among beliefs, desires, and the world be as similar to our own as possible. This principle I call the *principle of humanity*.[15]

Of course, to require that the imputed pattern be as similar to one's own as possible would be to require that the pattern be exactly similar to one's own, but this is clearly not what Grandy has in mind. Rather, one is to make the imputed pattern of relations among beliefs and desires similar enough to one own so that translation may be successful and, within that constraint, as similar to one's own as possible compatibly with one's evidence for attributing propositional attitudes. The trouble with this is that it does not tell us how similar to one's own the pattern of relations one attributes to others has to be for one to be able to make sense of them. But while the principle as Grandy states it does not determine this, it is clear from his discussion that Grandy thought that one had to find quite a lot in common with others to be able to interpret them:

> Quine's insistence on the determinate translatability of observation sentences and on the preservation in some sense of logical truth is subsumable under a general principle of strongly preferring agreement on obvious truths. My constraint on translation also leads to the maxim of preserving agreement on the obvious, but it attempts to derive this directly from the motivation for translation. If a translation plus our observations seem to indicate that a speaker denies the truth of a sentence in circumstances where the truth of

> the translation would be obvious to us, then, unless we have some explanation of this fact on our model, this counts heavily against the translation.[16]

Grandy's intent, then, is that we understand the principle of humanity as enjoining one to find oneself in agreement with those one interprets on obvious truths (by one's own lights) except where one has an explanation of the discrepancy. Thus, we may restate the principle of humanity as follows:

> Do not impute an obvious falsehood to a speaker, or find the speaker failing to believe an obvious truth, unless there is an explanation of how he could have come to believe such a falsehood or to fail to believe such a truth.

If this is how to understand the principle of humanity, it requires that if one finds that others do not recognize the validity of inferences one finds obvious, one must be able to explain their error. As Grandy notes, in cases in which one has gone through a long period of training or poured over a proof of a theorem for a long time, an explanation in terms of a difference in histories will be ready to hand. However, in the case of those inferences which one treats as most basic and most obvious, those inferences which one has not had to have any special training to recognize beyond learning one's language, there will be no possibility of this kind of explanation of error. In this case, Grandy's principle of humanity will require one to find speakers one interprets agreeing with one on the validity of those inferences. Thus, the principle of humanity, as Grandy intended it to be understood, establishes the fixed bridgehead view (given assumptions (1)-(3)). The principle of humanity provides an argument for all interpretable agents agreeing with an ideally rational agent on the most basic and obvious logical truths and valid inferences. Of course, this also thereby establishes the minimal rationality view, but what Stich wants is an argument for the minimal rationality view that is not at the same time an argument for the fixed bridgehead view.

We can see that if we accept the principle of humanity as a constraint on interpretation, and, as Stich assumes, and, thereby, as a constraint on whether others have intentional attitudes, we must deny, given (1)-(3), the coherence of the thought experiment Stich employs to try to undermine the fixed bridgehead view. Note that this argument will require both that others recognize as valid those inferences we find obviously so and that they have the ability to make those inferences, and so are able to display them in their reasoning in appropriate circumstances. Thus, it has the consequence that the thought experiment is doubly flawed, through imagining others could fail both to make and to be able to recognize the validity of conceptually simple inferences we can make and recognize the validity of.

V

Let us now step back from the details of Stich's argument to ask how things stand for the reform-minded epistemologist, in whose interest these attacks are mounted. How important is it to Stich's announced project actually to establish that the only limits on how irrational we can be are those implied by the minimal rationality view? It is hard to see why this is a necessary condition of undertaking the project of reforming our reasoning. A plausible fixed bridgehead view is certainly compatible with the possibility of improving both the ratio of true beliefs to false ones among the beliefs we hold and our practices in evaluating the evidence for our beliefs. Indeed, not only is the assumption that we are rational in this sense compatible with the project of improving our reasoning, it seems required for it. For, (a) unless we are able to recognize when an inferential practice is a good one, it would be hopeless to try to refine those practices we engage in by working out in detail the practical and epistemic consequences of different inferential strategies, and (b) unless we are guaranteed that intentional agents share with us the ability to recognize simple inferences, we could hardly hope to make much headway in reforming their inferential strategies. The reform-minded epistemologist should welcome an *a priori* argument for our being rational, for that we all share a bridgehead of rationality is a methodological presupposition of his project.

We have throughout our discussion been characterizing rationality in terms of an agent's ability to recognize that inferences in a certain class are valid. This is an important difference between our treatment and Stich's. Stich's discussion focuses not on the ability of an agent to recognize an inference as valid but on what patterns of inferences are actually exhibited in his behavior. For example, in the characterization of the fixed bridgehead view above (see section I), Stich says that every intentional agent must `manifest' inferences in some special set. We take Stich literally here.  `manifest' means to make evident by displaying or showing. Thus, Stich holds that the fixed bridgehead view requires that this special class of inferences be displayed so that others have access to them; this can only mean their being displayed in the subject's behavior. It is the failure to mark the distinction between exhibiting a certain pattern of perhaps valid inferences in one's behavior, and being able to recognize that those inferences are valid, which leads Stich to miss that anyone whose feasibility ordering for recognizing inferences as valid was inverted with respect to ours would simply not be making rational inferences, even if the inferences he manifested were valid.   (We put aside for the moment our second complaint with the thought experiment.) It is odd that Stich should fail to make this distinction here because it is closely related to a distinction which he does note and discuss in chapter 4 of *The Fragmentation of Reason*, the familiar competence/performance distinction as applied to knowledge of grammar.[17] Once we have marked the distinction, we can see that what is relevant to the question whether someone is rational is not whether he makes valid inferences, but whether his making them is a reflection of an underlying ability to recognize them as valid. Making valid inferences or exhibiting valid inferences in one's behavior is neither necessary nor sufficient for being rational. This helps us to see that the project of the reform-minded epistemologist is best seen as that of improving our epistemic *practices* rather than our epistemic *abilities.* The project must assume that we are able to recognize, even if with effort, that one strategy is better than another for achieving our epistemic goals. Thus the aim cannot be to provide us with fundamentally new capacities, for unless we already have the ability to recognize reasonable inferences as such, we would have no way to determine which capacities we wanted. What we *can* do, however, is to improve our epistemic practices by attention to their consequences both in the short and the long term.

Neither this distinction nor the project of reforming our cognitive practices is new in epistemology. The most prominent example in modern philosophy of an epistemic reformer who makes this distinction is Descartes, who takes a very optimistic view of our cognitive *capacities,* yet regards our practice as open to considerable improvement. Consider this passage from the

Discourse on Method:

> the power of judging well and distinguishing truth from falsehood, which is what we properly mean by good sense or reason, is naturally equal in all men; and furthermore, ... the diversity of our opinions does not arise because some men are more rational than others, but only because we direct our thoughts along different ways, and do not consider the same things. For it is not enough to have a sound mind; the main thing is to apply it well.[18]

On such a view, reform of our cognitive *practices* is compatible with any degree of perfection in our cognitive *capacities*. And, as we noted above, the very possibility of reforming our cognitive practices depends on our cognitive capacities embodying a certain level of competence in evaluating our patterns of reasoning.[19]

Notes

1. *The Fragmentation of Reason: Preface to a Pragmatic Theory of Cognitive Evaluation,* (MIT Press: Cambridge, 1990). All parenthetical citations of page numbers will be to this book.

2. It is a bit puzzling exactly what the purpose of this chapter is in the overall argument of the book, since Stich in the end argues that rationality is not something we should care about, and, hence, not something we should take as our goal in reforming our cognitive practices. Instead, he suggests that we might adopt different standards for evaluating cognitive systems. Thus, the rhetorical stance of the second chapter sits uneasily with the larger aims of the book. We will treat the second chapter as standing alone, since we think its arguments are of independent interest.

3. Although we do not intend to pursue the point, it is worth observing that all of these views of rationality apply only to relations among beliefs. A more complete account of the limits on irrationality would take into account practical rationality as well as epistemic rationality.

4. Stich attributes the perfect rationality view to Daniel Dennett. But Dennett is not committed to anything stronger than what Stich calls the fixed bridgehead view. In fact, the passage Stich quotes from Dennett,

> ‘there is no coherent intentional description’ of a person who `falls short of perfect rationality and avows beliefs that either are strongly disconfirmed by the available evidence or are self-contradictory or contradict other avowals he has made’ (p. 80),

aside from the expression ‘perfect rationality’, does not suggest anything stronger than the fixed bridgehead view. Stich cites some passages in which Dennett does require deductive closure for an `ideally or perfectly rational system’, but it is clear from the very next clause in one of the sentences Stich cites that n Dennett’s characterization there of perfect rationality it is not supposed to be a constraint on having any intentional states at all, for Dennett continues,

> but any actual intentional system will be imperfect, and so not all logical truths must be ascribed as beliefs to any system. Moreover, not all the inference rules of an actual intentional system may be valid ... (*Brainstorms* (MIT Press: Cambridge, 1978) p. 11)

It is clear enough from this that Dennett is not an advocate of the view that Stich attributes to him.

Stich also suggests that Donald Davidson maintains the perfect rationality view. That this cannot be so should be clear from Davidson’s work on weakness of the will, self-deception, wishful thinking, and other forms of irrational behavior and reasoning, in which Davidson’s aim is to accommodate the possibility of these forms of reasoning while at the same time urging that we can make sense of them only against a background of agreement on truth and standards of inductive and deductive reasoning. Davidson’s actual view also seems much more plausibly classified as a fixed bridgehead view.

5. What this comes to depends in part on how we are individuating inferences. Do we do it by their logical form or by reference to belief content? From Stich’s discussion of examples it seems clear that he has in mind individuating inferences by the content of the beliefs involved, rather than by their logical form. It is not at all clear that this is the appropriate way to think about individuating inferences, and it certainly is not if what we are concerned with is whether others share with us basic logical concepts. For then what becomes important are the formal properties of inference patterns, whether we can find in others’ behavior patterns of inference which allow us to interpret some of their words as expressing logical constants, quantifiers, and modal operators.

6. Martin Hollis, ‘Reason and Ritual’, in *Rationality* ed. by B. Wilson (Oxford: Blackwell, 1970), and `The Social Destruction of Reality’, in *Rationality and Relativism,* ed. by Hollis and Lukes, (Oxford: Oxford University Press, 1982).

7. We have talked sometimes about belief transitions and sometimes about inferences. These are not the same. Though every inference is a belief transition, not every belief transition is an inference; hence, not every valid belief transition is a valid inference. Call any causal process which generates new beliefs from old beliefs a `belief transition’. Call the old beliefs `antecedent beliefs’ and the new beliefs `consequent beliefs’. A belief transition is a valid belief transition if and only if the propositions which are the objects of the consequent beliefs are consequences of the propositions which are the objects of the antecedent beliefs. We can see that not every belief transition is an inference from the following example. Suppose that A’s belief that there is a cat on the mat suddenly (or even regularly) causes him to believe that there is a dog on the porch. In this case, although clearly there is a belief transition, A’s moving from the first belief to the second is not an inference. This would be so even if he simultaneously acquired the belief that he had inferred from his belief that there is a cat on the mat to his belief that there is a dog on the porch, for otherwise the belief that it was an inference would be self-verifying. It will not be necessary for our purposes to go into what distinguishes belief transitions in general from inferences. We will be granting (for now) that in the case Stich imagines the others are in fact making inferences. It will be sufficient for our purposes if we can show that even if they are making valid inferences, their inferences are not justified.

8. This point is connected with larger issues in epistemology, namely, the dispute between internalists and externalists about justification. It is not easy to characterize the dispute between these two views because each wishes to lay claim to our epistemic vocabulary for its own purposes. But, roughly, the externalist believes that the conditions conceptually necessary for being justified in believing that p do not have to be accessible to the believer in order for him to be justified. The internalist holds that it is at least a necessary condition that the believer has epistemic access to those conditions. Our proposal here is an internalist one. Whether or not externalism is a plausible story about the justification of our beliefs about the world around us, it is hardly, as the examples in the text show, a plausible account of what it is to make a rational inference.

[9] George Myro, Mark Bedau, and Tim Monroe, *Rudiments of Logic*, (Englewood Cliffs: Prentice-Hall, 1987).

[10] Benson Mates, *Elementary Logic,* (Oxford: Oxford University Press, 1972).

[11] In this example, our understanding of the negation sign enables us to recognize that the expressions '--P' and 'P' express the same proposition, which guarantees the validity of the inference.

[12] Stich's comments were delivered at the April 1992 meeting of the *Southern Society for Philosophy and Psychology*.

[13] 'Reference, Meaning, and Belief', *Journal of Philosophy,* 70 (1973): pp. 439-452.

[14] See 'The Method of Truth in Metaphysics', reprinted in *Inquires into Truth and Interpretation* (Oxford: Clarendon Press, 1984), and 'A Coherence Theory of Truth and Knowledge', reprinted in *Truth and Interpretation: Perspectives on the Philosophy of Donald Davidson,* ed. Ernest LePore (New York: Basil Blackwell, 1986). For a criticism of Davidson's argument, see Kirk Ludwig, 'Skepticism and Interpretation', *Philosophy and Phenomenological Research*, 52 (June 1992): 317-339.

[15] *op. cit.* p. 443.

[16] *op. cit.,* p. 443.

[17] Stich discusses the competence/performance distinction in chapter 4 in response to Jonathan Cohen's argument in "Can Human Irrationality Be Experimentally Demonstrated?" (*Behavioral and Brain Sciences*, 4 (1981): 321). Stich argues that there are important differences between the case of grammatical competence and that of reasoning competence. But whether he is right or not, this will not affect the present point unless there is *no* competence/performance distinction at all to be drawn in the case of reasoning. This is very unlikely.

[18] G. E. M. Anscombe and P. T. Geach, eds., *Descartes Philosophical Writings* (Bobbs-Merrill: Indianapolis, 1971), p. 7.

[19] We would like to thank Piers Rawling, Stephen Stich, Corliss Swain, Chris Swoyer and anonymous referees for this journal for helpful comments and criticisms, as well as audiences at the University of Manchester, the 1992 Southern Society for Philosophy and Psychology meetings, the 1993 IUC conference on Connectionism and the Philosophy of Mind, and members of the 1991-92 philosophy reading group at the University of Florida.