Closure and Rational Belief

Simon J. Evnine, Miami

What kinds of principles belong in a theory of rational belief? One contention of this paper is that it is impossible to reach informed conclusions about the answer to this question without a clear sense of what such a theory is intended to do. It may be rejoined that such a theory is always, at least, intended to cash out, in perspicuous terms, the notion of rationality, and hence no further context is needed to try and decide what principles belong in the theory. This will not do: use of the term 'rationality' is so multi-faceted that it is doubtful whether there is a single concept that it picks out. In fact, specifying what a theory of rationality is intended to do may go hand in hand with providing a characterization of one notion (out of many) of rationality. This will prove to be the case in the following.

A second, and in fact the major, contention of this paper is to argue for some particular closure principles for rational belief. I will, following my own advice above, supply a use for a theory of rational belief against which my discussion will take place. I will identify very strong, and very weak views about closure and argue that, against the background of the use of the theory that I have in mind, they both fail. I will then argue for a position somewhere in between.

One - I do not claim the only - task of a theory of rational belief is to enable us to interpret others. By interpreting others, I mean being able to explain and predict their actions by attributing various mental states to them. A number of philosophers have argued that the best way to do this is to assume that the interpreted other is rational, and to ascribe to her mental states that make her actions rational. A theory of rational belief can play an essential role in this process. Of course it cannot be anywhere near the whole story since, by definition, it concerns only belief and no other mental states. But belief is generally admitted to be especially important among mental states and, along with desire, is often taken as the main ingredient of an interpretation of someone. Sometimes, when we are interpreting someone, we have direct evidence for the presence or absence of a given belief. The person, for example, asserts, or denies, that *p*. Often, however, we wish to ascertain whether someone believes, or does not believe, something when we have no direct evidence. It is in such cases that a theory of rational belief will be of vital importance. Since we will have direct evidence for the presence or absence of other beliefs, the assumption that the person being interpreted is rational will allow us to use a theory of rational belief to extend our interpretation into areas for which we do not have direct evidence. This picture provides the background against which I will now discuss various approaches to closure.

Closure principles come in different strengths. Strong closure is expressed thus:

(SC) A set of rational beliefs **S** is such that, for any finite subset **T** of **S**, if **T** $\vdash b$, then *b* belongs to **S**.

By contrast, weak closure requires only:

(WC) A set of rational beliefs **S** is such that, for any member *a* of **S**, if *a* $\vdash b$, then *b* belongs to **S**.

The difference is that, by weak closure, a set of rational beliefs contains what is logically implied by any of its members taken by itself, whereas by strong closure, a rational belief set contains what is implied by any its members taken with each other. Those attempting to provide formal theories of rational belief have typically adhered to one or another form of closure.

Against the background of interpretation, both of these principles seem much too permissive. Even (WC) would license us in attributing to someone many beliefs that would either have no bearing on explaining and predicting her behavior, or would be positively misleading, bringing us to expect actions that were never performed or not expect ones that were. As a special sub-problem of this variety, (WC) might well lead us to attribute beliefs to someone that she positively abjured, thus coming into conflict with our direct evidence.

Many people have argued thus against (WC) or (SC) and several weaker alternatives have been suggested. A modest alternative is to require closure of rational belief under *believed* implication:

(BC) A set of rational beliefs **S** is such that, for any members *a,b,c*... of **S**, if it is believed that *a,b,c*...$\vdash d$, then *d* belongs to **S**.

This does well in dealing with the problems that beset (SC) and (WC), but it faces problems of its own, owing now to its weakness. The underlying problem is that it severs the link between belief and rationality, as this link serves the goal of attributing beliefs to a subject in interpretation. We cannot assume that just because *a,b,c*... do imply *d* that a subject believes that, nor that, given that a subject believes that *a,b,c*... imply *d* that they really do. The interpreter thus cannot use a knowledge of logical relations to extend the beliefs she can attribute to a subject without some well-grounded beliefs about the subject's beliefs about what implies what. Such beliefs about implication are highly theoretical and unlikely to be forthcoming in many cases. We therefore risk ending up with a theory that simply cannot do the job it is designed for.

A more extreme response to the failures of traditional closure is adumbrated by Christopher Cherniak (1986). Cherniak argues that which logical abilities a creature has is determined by its psychology, but that there is no necessary connection between rationality and any particular type of psychology. Hence, equally rational creatures may, owing to their different psychologies, have very different (possibly non-overlapping) sets of logical abilities. As a consequence, Cherniak argues that there can be no theory of rational belief that is *a priori* suitable for all creatures that we might wish to interpret. The only 'closure' rule that applies to a rational creature *per se* is that it believes *some* of the consequences of its beliefs. Nothing more specific can be said. The only way to supply a more detailed theory for interpretation is to rely on a theory that details the specific psychology of the creature to be interpreted.

The real weakness of this approach is masked by the fact that it may well be true that all normally developed, adult human beings do share the relevant psychological features, and that such creatures are, as it happens, the only ones with which we are acquainted that are plausible candidates for interpretation. It could even be argued (though Cherniak does not) that humans have innate access to the type of psychological theory necessary for

interpretation (on his view), or that we can easily - if with some inductive weakness - obtain it by extrapolation from our own cases. But even among normally developed, adult humans, there are strong linguistic and cultural differences and it has been urged that these differences extend to differences in logical abilities. For Cherniak's approach to be generally useful in human interpretation, such relativism must be answered at the empirical level, something that a theory that genuinely dealt with rationality *per se* would not have to do. And of course, should there turn out to be non-human rational beings, there would not even be a *prima facie* reason for thinking, on Cherniak's view, that we could use knowledge of logical relations to extend our belief attributions beyond those for which we had direct evidence. If we had direct evidence for such a creature's believing the conjunction *p and q*, there would be no justification, absent an appropriate psychological theory, for assuming that it believes *p*.

I come now to my proposals. We should start by thinking about what kind of creatures should fall within the scope of a theory of rational belief if it is to be used for the purposes of interpretation. Clearly, such creatures must be rational, but we need to say more about what this means, since, as noted, the concept has been so diversely used. Here, then, are some conditions, individually necessary but not claimed to be jointly sufficient for one conception of rationality that is tied to the project of interpretation.

I propose that a rational creature must a) be capable of intentional action, and hence of deliberation. b) It must be able to develop and modify, in response to information received piecemeal, some conception of a unified world of which it is a part. c) It must recognize the (at least potential) existence of other creatures with their own representations of the world, and hence have a concept of belief, and therefore concepts of truth and falsity.

It must be said that these conditions add up to both a meager and a stringent conception of rationality. Stringent, in that normally-developed humans are very likely the only example with which we are acquainted of creatures meeting these conditions. Meager, in that the conditions are clearly and obviously met by all such humans, that they do not exhaust all that could be said of the rationality of such humans, and in that any type of being that had anything remotely similar to human culture (that could sustain any forms of art, technology and social institutions) would surely have to meet them.

I have argued at length elsewhere (Evnine 2001) that satisfaction of these conditions entails the possession of certain particular logical abilities. I shall not repeat the arguments here, though of course they form an essential piece of the position I am working out here about closure requirements. Here is a list of the inferential abilities entailed by satisfaction of the conditions that are relevant to our present purposes:

a) possession of a concept of conjunction subject to the usual introduction and elimination rules;

b) possession of at least one of any concept of conditionality that is subject to *modus ponens*.

(Some of the conditions on rationality support logical abilities that are not inferential in nature, and that are connected with consistency conditions on rational belief.) It is on the basis of these 'universal' logical abilities that we can formulate a principle of closure that occupies a middle ground between the overly strong and overly weak versions we have rejected:

(Closure) If **S** is a set of rational beliefs, then i) for any pair of beliefs {*a,b*} in **S**, *a and b* is in **S**; ii) for any belief *a and b* in **S**, *a* is in **S** and *b* is in **S**; iii) for any pair of beliefs {*a, if a then b*} in **S**, *b* is in **S**.

(Closure) requires, thus, not universal closure under entailment, but closure only under conjunction (elimination and introduction) and *modus ponens*. It is weaker than either (SC) or (WC), but unlike (BC), it preserves a genuine link with logic, in allowing the attribution of (some) beliefs to someone merely on the grounds that they follow logically from other beliefs that person has. And unlike Cherniak's view, (Closure) is independent of any psychological theory, being tied only to satisfaction of the highly general conditions contained in the characterization of rationality.

In the remainder, I will mention two objections to (Closure), one specific, the other general. Specifically, clause i) of (Closure) requires that a rational person's beliefs be closed under conjunction introduction. This particular aspect of closure has received a lot of criticism independent of the general problems for closure conditions that we looked at above, criticism centered around the Lottery and Preface paradoxes, which are purported to show that closure of belief under conjunction introduction will lead to irrationality, even to the violation of the weak (WC). I have addressed these objections elsewhere (Evnine 1999). I will say here only this. The requirement expresses the view that all our beliefs should be open to rational interaction with each other; that we should not, in other words, compartmentalize our beliefs. In the context of interpretation, I believe this assumption is the default one - when we interpret, we initially assume that we can bring to bear on our interpretations *any* other beliefs we have reason to attribute to the person whom we are interpreting. There are, of course, special occasions when this default assumption must be modified. The theory of rationality of which the condition of closure under conjunction introduction is a part may thus be said to yield a theory of *prima facie* interpretation rather than interpretation *tout court*.

The general criticism of my proposals is that they have the appearance of being arbitrary. We have closure under certain logical operations but not others (for example, inclusive-or introduction). Why just the conditions that we have, and not others?

To see the answer to this objection, we must remember how we got to those requirements. They were established on the basis of particular logical abilities that themselves were argued to be entailed by the fulfillment of certain general necessary conditions on being rational, in one sense of that term. If the arguments (for which I referred to another work of mine) that link the necessary conditions on rationality with the various logical abilities are good, this means that a theory of rational belief with these closure requirements should be a good method of interpretation for any creature that manifests that form of rationality - any creature, I argued, that is a remotely plausible candidate for interpretation. I have not argued that *only* these inferential abilities can be proven to be 'universal' on the basis of this conception of rationality, but I have not seen a way to establish the universality of others. If anyone can, then further closure requirements might be added to the theory of rational belief for the purposes of interpreting such creatures. Furthermore, if anyone can, by adding further conditions to the conception of rationality, establish the necessity of further inferential abilities, then a theory of rational belief, with additional closure requirements, can be developed for creatures satisfying that narrower conception of rationality. Such a conception of rationality may well

still be wide-enough to include any creature we may want to interpret, given the initial weakness of my own characterization of rationality.

Literature

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

Evnine, S. 1999 "Believing Conjunctions", *Synthese* 118, 201-27.

Evnine, S. 2001 "The Universality of Logic: On the Connection between Rationality and Logical Ability", *Mind* 110, 335-67.