STALNAKER ON INTENTIONALITY

Stalnaker offers us a theory of the deliberation and inquiry of intelligent agents that makes heavy use of believing and desiring, construed in a particular way. Let a *Stalnaker-proposition*, or *S-proposition*, be a function from some algebra of possible worlds (not necessarily comprising all the possible worlds) into truth values. (If we ignore the qualification that not all possible worlds need be assigned truth values, we could just as well say that an S-proposition is simply a set of possible worlds.) Stalnaker wants to construe believing and desiring as relations between the cognitive states of the agent and S-propositions. Stalnaker thinks that the S-proposition assigned to an intentional mental state should be viewed as the *content* of that state, or the *object of* that state.

Stalnaker sees two sorts of advantages to his view of the objects of intentional states over rivals (in particular, rivals that ascribe more fine-grained contents): first, that it better fits a pragmatic picture of the purpose of postulating intentional states (in large part because of the identity conditions on contents that it offers); and second, that only by conceiving of the objects of mental states in this way we can solve *the problem of intentionality*, that is, the problem of giving a naturalistic account of what gives mental states the content they have. I will be disagreeing with his views on both points.

I should say at the outset that the whole idea of "the object of" or "the content of" an intentional mental state needs to be treated with care; Stalnaker's book (like virtually all of the literature on this subject, including some of my own previous work) takes this notion too much for granted. I have no doubt that there are useful rough and ready senses in which a mental state can have content and in which different mental states can have the same content. But our relatively ordinary assertions of contentfulness and of sameness of content seem highly context-dependent--especially assertions of sameness of content between the mental states of different agents. (Even more especially, when those agents do not share a language; still more, when they don't even belong to the same species.) Any view according to which we are to assign entities to mental states that are to serve as their contents (and are then to define having-of-content and sameness-of-content in terms of such assigned entities) is clearly ladening our ordinary talk of having-of-content and sameness-of-content with a substantial body of theory; and my first point is that the nature of the theory and the motivation for introducing it deserve serious discussion. My second point, to be elaborated in Sections 2 and 3, is that Stalnaker not only wants to assign *entities* to mental states to serve as their contents, he wants to assign *intrinsically representational* entities, i.e. entities for which it is built in that they represent the real universe as being a certain way; the nature of the theoretical role that such an assignment is to play, and the motivation for introducing such a body of theory, deserve discussion still more. And my third point is that even given that we want to assign intrinsically representational entities to mental states to serve as contents, it is by no means obvious that there is anything like a unique sort of intrinsically representational entity to assign.[1] Stalnaker seems to take it for granted that there is a clear issue between a position according to which we assign S-propositions to mental states as their contents and a

position according to which we assign finer-grained entities as the contents of mental states; but I do not think that this is a clear issue at all.

So, though I will be disagreeing with much that Stalnaker says when he tries to motivate using S-propositions as "the" contents of mental states, still I wouldn't want to put my disagreement by saying that he has chosen "the wrong entities" as contents. In fact, I agree that S-propositions are in a sense the core of content (we can call it *S-content*), and the aspect of content for which the problem of intentionality arises most forcefully. One of my points will be that we need to postulate other "content-like" features of mental states not only in explaining behavior but *even in solving the problem of giving a naturalistic account of S-content*; again, whether we call these features part of content is merely a matter of terminology.

1. The Pragmatic Picture and the Linguistic Picture.

Stalnaker says that there are two very different pictures that can govern people's conceptions of intentional mental states--he calls them *the pragmatic picture* and *the linguistic picture*. He advocates the pragmatic picture, and holds that it leads almost inevitably to

1.   the view of objects of belief as "coarse-grained", in the sense that there can not be distinct but logically equivalent objects of belief; and

2.   the view that there can be no philosophical point to adopting a language of thought hypothesis.

Apparently he holds, then, that advocates of a language of thought, or of finer-grained content, deny the pragmatic picture of intentional states.

It seems to me that his discussion of these matters is quite misleading. The pragmatic picture of belief and desire is described in part as follows:

> Representational mental states should be understood primarily in terms of the role that they play in the characterization of action.... One explains why an agent tends to act in the way he does in terms of ... beliefs and attitudes. ... [O]ur conceptions of belief and of attitude pro and con are conceptions of states which explain why a rational agent does what he does.... Linguistic action, according to this picture, has no special status. Speech is just one kind of action which is to be explained and evaluated according to the same pattern. (p. 4)

The pragmatic picture described here is most compelling. Stalnaker says that philosophers who have advocated thinking of belief and desire in terms of a "language of thought" or an "internal system of representation" have had a different picture, but that does not seem

to me to be true: except for the restriction in the above quotation to *rational* action, the quotation seems to precisely describe the attitude toward belief and desire that Fodor took in [1975] and that Harman took in [1973] and that I took in Chapter 2. Indeed, although Stalnaker begins his characterization of the supposedly contrasting "linguistic picture" by saying that on this picture "rational creatures are essentially speakers" (p. 6), he ends up saying that any view that conceives the intentionality of intentional mental states on analogy with the intentionality of linguistic expressions counts as a version of the linguistic picture (p. 7). But then the pictures do not contrast: the whole point of much of the argumentation in the three works cited was that there was reason to postulate structure resembling linguistic structure in intentional mental states in order to explain non-linguistic as well as linguistic behavior. I'm not here saying that this argumentation is right--in fact, I now think that my claims on this point in Chapter 2 are rather overblown and are stronger than I needed for the points I was primarily concerned to make in that paper--but I think that Stalnaker's long discussion of language of thought views is seriously warped by his suggestion that the point of a language of thought theory is to do something other than explain behavior.

Turning now to the issue of "coarse-grainedness" of content: an almost universal first reaction to a possible worlds view of belief and desire like Stalnaker's is that it can't be adequate, since people can clearly believe that P and disbelieve that Q when P and Q are logically equivalent. Now, Stalnaker has an ingenious reply to this line of objection, which I will consider at the end of the paper, but for now I just want to emphasize that this line of objection against Stalnaker is an initially natural one *from the point of view of the pragmatic picture*. That is, even when one's task is to explain *non-linguistic* behavior, there is often a *prima facie* need to attribute a belief in one proposition and a disbelief in an equivalent proposition (i.e., there is a *prima facie* need for a conception of content more fine-grained than S-content). For instance, if I offer someone who doesn't know much mathematics $1000 for an example of a plane map that requires more than four colors to color (according to the usual coloring conventions), he will behave very differently than he would if I had offered $1000 for a trisection of a Euclidean 60 degree angle by straight edge and compass; to explain this, I need to attribute different beliefs and desires to him in the different cases, and it is *prima facie* difficult to see how I can do this in a relevant way if the desire to do one impossible task is identified with the desire to do any other impossible task.

Stalnaker thinks (pp. ix, 23-5) that the pragmatic picture leads inevitably or almost inevitably to the coarse-grained conception of the objects of intentional states. Why, given the obviousness of examples like the above? The argument appears to be that on a pragmatic picture we ought to think of the deliberation that is involved in determining behavior as consisting of the weighing of possibilities (i.e., weighing which possible states of the world are more likely, what the effects of various courses of actions would

3

be in these different states of the world, and how desirable each of these possible effects might be)--see pp. 4-5 and 85. If deliberation did indeed have to literally consist in the weighing of logical possibilities, Stalnaker's conclusion might well follow. But why not say instead that deliberation consists in the weighing of states of affairs which are *epistemically possible for the agent*--or more accurately, the weighing of state of affair *representations* which are *treated by the agent as* consistent?[2] This would certainly handle the case of the $1000 offer.

Stalnaker considers such a model, but argues against it as follows:

Could we escape the problem of equivalence by individuating [mental states], not by genuine possibilities, but by epistemic possibilities--what the agent takes to be possible? This would avoid imposing implausible identity conditions on [mental states], but unfortunately, it would also introduce intentional notions into the explanation, compromising the strategy for solving the problem of intentionality. If belief and desire are to be explicated in terms of naturalistic relations such as indication and tendency-to-bring-about [i.e. if they are to be explicated directly in information theoretic terms], then the propositions used to individuate [belief- and desire-states] must be the ones that are relevant to these relations, and these clearly must be genuine, and not merely epistemic possibilities. (pp. 24-25)

In short, epistemic probability is an intentional notion that can't be cashed out in the way that he wants intentional relations to be cashed out (namely in terms of "indication relations" and the like, to be discussed shortly).

But as I've said, talk of what is epistemically possible for the agent is really just a loose way of talking about which representations the agent treats as consistent, and one can presumably explain treating a representation as consistent in terms of the employment of that representation in thought (e.g., one treats it as consistent iff one doesn't treat it as implying everything else); no intentional notion of *representing a possibility* enters the picture at all. Indeed, as I will argue at the end of Section 3, there is serious doubt as to whether Stalnaker's strategy for explaining representation in terms of natural relations could work without appeal to antecedent logical relations among states; so it is the "real possibility view" rather than the "epistemic possibility view" that is in danger of using intentional notions that can't be cashed out naturalistically.

2. The Problem of Intentionality (1)

Much of the first two chapters of *Inquiry* is a discussion of *the problem of intentionality*--the problem of giving a naturalistic explanation of intentional relations like believing and desiring that apparently relate people to the objects of belief. Stalnaker does not attempt a detailed solution to this problem, but he does make some sketchy remarks

about how one might develop a solution that was based on the pragmatic picture. And he argues for the superiority of this sort of solution over two solution-sketches which he regards as anti-pragmatic, namely a solution-sketch that I suggested in Chapters 1 and 2, and a solution-sketch that might be suggested by Davidson's writings. As will emerge, I think that Stalnaker misrepresents the differences between his own approach and the approach I urged in the papers he cites. However, my concern here will not be to give a detailed defense of my views against all his criticisms--these are issues on which I have significantly changed my views anyway. Rather, I will just try to clarify the issues a bit.

First I need to develop the distinction, mentioned in my opening remarks, between using "intrinsically representational" entities as objects of belief and using entities that are not "intrinsically representational". Suppose that we think of intentional mental states as follows: most (perhaps all) of them are significant (have-content), and among those that do there is a relation of synonymy or sameness-of-content (applicable between states of different agents as well as between states of the same agent) which is an equivalence relation. Then in one reasonable sense of 'objects of belief', we could use the equivalence classes of states, under the sameness-of-content relation, as objects of belief. I count this as an example of "objects of belief" that are *not* "intrinsically representational": they are simply sets of states, and don't have built into them that they represent the world in any way. Alternatively, if one assumes that the sameness-of-content relation makes clear sense only among states of the same agent, but that among the states of any one agent it is an equivalence relation, then one could instead use equivalence classes *of the states of a single agent* as the objects of belief for that agent; objects of belief would then be "local objects" in that no two agents could literally share the same object of belief. This too would be an example of a conception of objects of belief as not intrinsically representational. (One version of this is equivalent to the view I took of the objects of *non-intentional* psychology in Chapter 2.) By contrast, if we view the object of a state of believing that Caesar crossed the Rubicon as an ordered triple whose components are Caesar, the Rubicon, and the relation of crossing, that is an example of an intrinsically representational object of belief: it is intrinsically representational because its whole point is to represent the world as such that Caesar crossed the Rubicon. Similarly, if we take as the object of the belief-state the set of possible worlds in which Caesar crossed the Rubicon (or, what comes to the same thing, the S-proposition that assigns 'true' to worlds where Caesar crossed the Rubicon and 'false' to the others), that too is a conception of objects of belief as intrinsically representational.

It seems to me that the most central philosophical question about intentionality is whether there are any good reasons to utilize intrinsically representational objects of belief (and if so, what those reasons are). It does not seem to me that one ought to assume without argument that we should utilize intrinsically representational objects. Two points need to be made in clarification of this claim. (1) To refuse to utilize

intrinsically representational objects of belief is not to divorce belief-states and desire-states from the external world: for instance, a theorist of belief who shunned intrinsically representational objects of belief could still theorize about the causal influence of the external world on our belief-states and desire-states, and about the causal influence of our belief-states and desire-states on the external world. (Of course, such a theorist wouldn't have the intrinsically representing objects of states available to aid him in formulating the laws about how belief-states and desire-states influence and are influenced by the external world; but I think there is serious question as to whether intrinsically representational objects of mental states would ultimately be of much use in a serious effort to state such laws. Indeed it seems to me that the issue of whether we ought to use intrinsically representing objects of mental states turns largely on the answer to this question.) Moreover, (2) a theory of belief that shunned intrinsically representational objects of belief could still be used as the basis of 'believes that' and 'desires that' attributions: for instance, one might say that I believe that snow is white if I am in a belief-state that would typically lead to the production of the sentence 'Snow is white', and that someone else (not necessarily a language-user) believed that snow is white if he were in a state whose causal role were sufficiently similar to mine. That's very vague, of course, but working out a theory along these lines (as Stich suggests in [1982] and in Part I of [1983]) would not require the resources required of a solution to the problem of intentionality with intrinsically representing objects: first, because the only relations that would require explaining in the theory of 'believes that' attributions would be state-state relations as opposed to state-world relations;[3] and second, because almost anyone who took this view would hold that the state-state relations were highly vague and context-dependent, thereby inducing an asymmetry between the relative clarity of 'belief that' ascriptions to oneself and the relative unclarity of such ascriptions to others.

I suspect that a view of the sort just outlined (to which I now have considerable sympathy) underlies much of what Stalnaker had in mind as part of "the linguistic picture" of intentionality; though it certainly does not fit all language of thought views.[4] Even the view outlined, however, does not reject the pragmatic picture of beliefs and desires. For beliefs and desires are still invoked to explain behavior, and the serious explanation of behavior (that envisaged in (1) of the previous paragraph) works without appeal to language. The special appeal to language comes in only in the account of ordinary "belief that" and "desire that" attributions (discussed in (2) of the previous paragraph); and the point of such a theory is that these attributions do not themselves play a serious explanatory role, but serve only as devices for alluding to the more serious sort of explanation envisaged under (1).[5]

In Stalnaker's discussion of the problem of intentionality, it is assumed that "the objects of mental states" are to be intrinsically representational entities. It is illuminating to notice however that nearly all of the evidence that Stalnaker cites for S-propositions (or

what is almost the same thing, sets of possible worlds) as objects of mental states could be obtained from structurally analogous entities that are *not* intrinsically representational. What is mostly relevant to sets of possible worlds as objects of mental states is their structural interrelations: the fact that they form a Boolean algebra. (On Stalnaker's view, a complete atomic Boolean algebra: see pages 54-56 of *Inquiry*.) Consequently, any other (complete atomic) Boolean algebra that was sufficiently large to make the psychological distinctions we need would do just as well for most psychological purposes; for most purposes at least, *there is no need to think of the atomic elements of the algebra as possible worlds; they could be anything at all*. For instance, they could just be numbers (possibly transfinite numbers, let's say, to insure that there are enough of them). Numbers, unlike possible worlds, do not purport to represent how the world is, so if we think of the atomic elements as numbers it will substantially alter how we conceive the problem of intentionality. Of course, numbers like possible worlds have their own special properties, and we would not want the problem of intentionality to be conceived in such a way that these special properties mattered. (For instance, we should take the project of explaining the difference between mental states that are assigned sets that include the number 17 and mental states that are assigned sets that don't include 17 as quite misconceived.) Rather, the point is that unless we can identify psychological laws that use features of sets of possible worlds *that go beyond their structural features*, then (if we buy Stalnaker's views on the identity conditions of the objects of mental states) we should conceive of the problem of intentionality as simply the problem of explaining the attribution of elements of a sufficiently large but otherwise arbitrary complete atomic Boolean algebra to our mental states. The project of explaining the difference between mental states that are assigned sets which include possible worlds according to which Caesar crossed the Rubicon and mental states that are assigned sets which do not include such worlds would be as misconceived as the analogous project about numbers.[6]

As I've said, Stalnaker conceives of the objects of mental states as intrinsically representational, which means that the differences between the set of possible worlds and other complete atomic Boolean algebras are taken to be somehow important (even though not important to much of psychological explanation). Even so, it is useful to divide the problem of intentionality into two parts: first, a part explaining the assignment of non-intrinsically-representing items (e.g., elements of an arbitrary complete atomic Boolean algebra) to our mental states; and second, a further part explaining the "interpretation" of the non-intrinsically representing (e.g. Boolean) elements. [Compare this to the distinction in Fodor [1980] and Stich [1983] and in Chapter 2 above between assigning uninterpreted sentences to mental states and assigning interpretations of those sentences. Here I'm simply making the same distinction, except freed from the presupposition of a language of thought.]

Even the first stage of this project raises some interesting problems, especially if

we assume that, to put it roughly, the important structural relations induced by the objects assigned to mental states must be explainable independently of any talk of assigned objects.[7] In Chapter 2 I interpreted an earlier paper of Stalnaker's (Stalnaker [1976]) as denying this "independence assumption" (and as using what I argued to be a misconstrual of the significance of functionalism to support that denial); but his remarks in the book (e.g. the remarks on the theory of measurement on page 10 and the remarks on functionalism in the first full paragraph on p. 15) suggest that this was a misinterpretation.[8] (And I will give reasons in the next section why he ought to accept the independence assumption.) So let us ask how the Boolean relations among mental states that are induced by assigning S-propositions to those states are to be explained. This is not an issue that Stalnaker discusses, but it is highly relevant to the motivations that have led many to be sympathetic to the idea of a "language of thought".

Let us focus on two particular mental states of a given agent: the first, s_1, a state of idly wishing to trisect a 60 degree angle in a Euclidean plane with straight-edge and compass, and the second, s_2, a state of having the conscious thought that Caesar crossed the Rubicon. (By a Euclidean plane I mean a plane that obeys the Euclidean axioms; I want a *logically* inconsistent wish.) The content assigned to s_1 is to be the empty set of possible worlds, or more generally, the zero element of a Boolean algebra; it is a subset of (or, bears the Boolean *less than* relation) to the content assigned s_2, so we can say that the content of s_1 *entails* that of s_2. Now, in virtue of what facts about s_1 and s_2 does the content of the first entail the content of the second? As part of the first stage of an answer to the problem of intentionality we must answer this question; if it proves unanswerable, the conclusion to draw is presumably that even the assignment of a Boolean structure to mental states is theoretically unjustifiable.

I do not mean to suggest that the question is unanswerable.[9] A simple-minded attempt to answer it would be to suppose that the agent is disposed to logically reason by means of certain rules, and that s_1 entails s_2[10] because there is a possible chain of states beginning with s_1 and ending with s_2 each of which (besides the first) comes from previous members of the chain by one of the rules of logical reasoning. [This is obviously *very* simple-minded. To get a *somewhat* more realistic approach, one would have to deal with two sorts of rules (or *prima facie* rules) that are roughly on par: rules of direct implication as above, and rules for recognizing possibilities (to be used in showing *failure* of implication); and one would have to say something about how they interact, e.g., what happens when they come into conflict with each other.][11] Note however that to postulate anything like "rules of logical reasoning" is to postulate something that goes far beyond the sort of Boolean structure that Stalnaker wants to build into the objects that he assigns mental states. And it is clear that one of the reasons that people have been attracted to language of thought theories over Boolean theories like Stalnaker's is precisely that language of thought theories assign mental states enough structure so that "rules of logical

reasoning" could be represented, thereby making the Boolean structure on the states less mysterious.

I do not contend that any *conceivable* model of the Boolean structure that Stalnaker postulates requires the representation of anything like "rules of logical reasoning": for instance, one might instead imagine that inside an agent's head is a huge array of boxes, one corresponding to each relevant possible world, and that what an agent believes is determined by which of these boxes has an 'X_B' in it. (The agent believes that Caesar crossed the Rubicon if all boxes corresponding to worlds in which Caesar did not cross the Rubicon contain an 'X_B'.) It is clear however that Stalnaker would disavow any such model. (See *Inquiry*, p. 22). And I think that any reasonably *realistic* model has got to give a place to something roughly on the order of "rules of logical reasoning"; and that involves enlarging on the Boolean structure.

Indeed, this extra structure is important not only in giving an idea of how the Boolean structure might be represented. It is also of direct importance in psychological explanations (indeed, *more* important than the Boolean notion of implying): think of the $1000 offer discussed in Section 1. I do not claim that the enriched structure need be *linguistic* structure in any very obvious sense (and indeed, the bounds of what counts as linguistic structure and what doesn't are quite vague); my point is only that it is important to develop psychological models that contain that extra structure, and that "language of thought" models at least have the virtue of doing this.

Of course, Stalnaker recognizes that there is a distinction between a person whose beliefs we would describe with one sentence and a person whose beliefs we would describe with a logically equivalent (but *unobviously* equivalent) one: he suggests in one place (p. 23) that the *objects* of these two people's beliefs are the same, but that the *forms* by which the common object is represented in the two cases differs. Presumably, then, he could recognize "rules of logical reasoning" (or some such thing): they would have to be stated in terms of the *forms* by which beliefs are represented, not in terms of the *contents* of those beliefs. I don't see that we can call this insistence on excluding form from the objects of belief *wrong*: he's free to use the term 'object of belief' as he likes. But Stalnaker seems to regard it as a substantive difference between his account and "the linguistic picture" that the former recognizes the form/content distinction while the latter doesn't; he says that "the conceptual separation between form and content is ... the central feature which distinguishes the conception of thought implicit in the pragmatic picture from the one implicit in the linguistic picture." (p. 23). But the substance of this "central issue" eludes me. As far as I can see the serious substantive issues are whether one needs to go beyond Boolean structure in giving psychological explanation, and if so what the nature of the extra structure postulated should be.

3. The Problem of Intentionality (2)

Now let us turn to phase two of the problem of intentionality: the special problem that arises when one takes ones objects of belief to be "intrinsically representing" entities.

Half a chapter of *Inquiry* is devoted to a criticism of a strategy for solving this part of the problem of intentionality which I suggested in Chapters 1 and 2 above. I now have serious doubts about the adequacy of that strategy (based largely on doubts about its presupposition that there is anything to be gained by using intrinsically representational objects of belief); but it will be useful to say a few words about Stalnaker's criticisms of it before presenting Stalnaker's own proposals.

The strategy of my Chapter 2 was to assume that mental states of believing, desiring, and so forth could be given a roughly sentential structure, and then to impose on these states a componential semantics which specified truth-conditions for sentences with analogous structures. That is, a person explicitly believes a proposition if he is in a state of believing that has that proposition as its content; where the association of contents with states of believing goes via a compositional semantics that utilizes the sentential structure of the states. (This was to hold for explicit belief–or *core* belief, as I called it there–rather than for belief generally. Belief generally was viewed roughly as a disposition to explicitly believe.)[12] The main ideas of this picture could be easily generalized so as to apply to lots of non-sententially structured states, e.g. states with pictorial structure. But the picture doesn't apply to states with only Boolean structure, and that of course is one of the reasons that Stalnaker is unsympathetic to the picture. Aside from that general reason however, he suggests several specific criticisms of the strategy that I presented, which I will mention after describing the strategy in more detail.

Underlying the strategy in question is a specific view of componential semantics: that the point of a componential semantics is to explain the semantic features of sentences in terms of the semantic features of their component parts. For instance, we might explain the truth-conditions of 'It is not the case that Caesar crossed the Rubicon' in terms of the fact that 'Caesar' stands for Caesar, 'the Rubicon' for the Rubicon, and 'crossed' for the relation of crossing (or if you prefer, the function assigning 'true' to the triples <w, x, y> such that x crossed y in world w), and the fact that 'it is not the case that' is a symbol of negation. Of course, a componential semantics says nothing about *in virtue of what* one of the non-logical components (such as 'Caesar' or 'crossed') stands for what it does; but I expressed hope that something vaguely like a causal theory of reference, applied to the structural components of a belief state (the "morphemes of the language of thought") might serve to fill the lacuna. I skirted the issue of logical connectives (as Stalnaker points out--this is one of his criticisms), but it now seems to me pretty clear what I should have said about them: the facts in virtue of which 'it is not the case that' obeys the truth table for negation are facts about its conceptual role; more specifically, the standard truth-tables for 'not', 'and', etc. are the only truth-tables that makes standard inferences involving these words truth-preserving.[13]

Stalnaker seems to have two main objections to the sort of picture just sketched. The first centers on my mention of causal theories of reference. Stalnaker points out that believable causal theories of reference for words in public languages require intentional notions: for what a word means depends on the attitudes of the users of the language. The only conclusion to draw from this however is that the project of coming up with a non-intentional causal theory of reference has more prospect of success for the "morphemes in the language of thought" than it has for words in public language. The prospects for a non-intentional theory for words in a public language seems irrelevant.

The other main objection is deeper: Stalnaker says that the approach that I advocated is too *atomistic* in that it claims that the most basic kind of representation holds between words as opposed to sentences, or "morphemes in the language of thought" as opposed to whole states. I have some feeling that there may be a legitimate complaint here, but I am unsure exactly what it is. Two points on this:

1.   Stalnaker characterizes my view as holding that "the name-object and predicate-property relations *come first*; the sentence-proposition relation is derivative." (p. 34, my italics.) Is this supposed to mean that first people invented names and predicates, and then some genius thought of putting them together to form sentences? Obviously I never held that view (nor the analogous view at the mental level). Rather, the view was that our goal is to explain the truth-conditions of sentences and belief-states; name-object and predicate-property relations (and their analogues for components of belief states) are theoretical relations needed in characterizing those truth-conditions; and an independent characterization of those theoretical relations is then needed (just as an independent characterization of valence in terms of chemical structure is needed, even though the *goal* of talk of valence is the holistic one of explaining chemical combination).

2.   A theory which is formally atomistic in the sense required above may accomodate a great deal of interaction among the "atoms". Thus a theory of reference for names might well have the form "reference is that relation R which is the first member of a pair <R,X> such that if you assign to each name the R-related object and to each predicate the X-related property and calculate truth-conditions in the usual way, then the resulting truth-conditions are such that blah-blah-blah." (Some sort of restriction, e.g. to causal relations, would of course be needed to eliminate some of the indeterminacy.)

I don't know if these points are enough to make "atomicity" seem unobjectionable. It does seem to me however that Stalnaker's own proposal, to which I now turn, is unworkable largely because it is *not* atomistic in the sense that my proposal was.

Stalnaker's discussion of his own solution to the problem of intentionality divides into three stages. In the first stage he emphasizes that there are unproblematic naturalistic

relations between objects and S-propositions; among them is the relation of *indication*, which Stalnaker takes to be particularly relevant to the problem of intentionality. He says

> Consider an object which has intrinsic states that tend, under normal or optimal conditions, to correlate with its environment in some systematic way, and where the object tends to be in the state it is in *because* the environment is the way it is. For example, the length of a column of mercury in a thermometer tends to vary systematically with the temperature of the surrounding air... (p. 12).

Under these conditions we say that the length of the column of mercury *indicates* the temperature of the surrounding air. The suggestion that will emerge later is that "belief is a version of the propositional relation I called indication" (p. 18).

I think there can be little doubt that if there is to be any hope of solving the problem of intentionality on the supposition of intrinsically representing objects of mental states, then something like indication must play a role in the account. The crucial question is, what exactly is the role it plays? In Section V of both Chapter 1 and Chapter 2, I argued that the motivation for introducing intrinsically representing objects (or equivalently, a correspondence theory of truth) into semantics and philosophy of mind was to enable us to formulate such indication-relations as exist between belief-states (and utterance-states) and the environment. Semantic notions such as 'truth-conditions' were thus to be viewed as theoretical terms in a "reliability theory", i.e. a theory of indication; and semantic notions such as 'refers' were viewed as still more theoretical notions of the theory, needed to formulate truth-conditions which were in turn needed to formulate indication relations. An attractive feature of this way of viewing things is that many of the features which have been argued to be essential to a plausible theory of reference are features that clearly have to enter in to any explanation of why my states or utterances are reliable indicators of the external world. For instance, to the extent that my utterances containing the phrase 'the Lyons silk-weavers strike' are reliable indicators of facts about the Lyons silk-weavers strike, that reliability is obviously totally dependent upon the existence of "experts" whose reliability I inherit; and the reliability of the experts (and hence, indirectly, of me) obviously depends on a causal network of some sort emanating from the silk weavers strike to their (and my) belief states and utterances.[14]

That's one view of the relation between indication and belief. It would however be a bit misleading to sum up *that* view by saying that belief is a species of indication (since on that view there is no reason to suppose that belief states in general reliably indicate the states of affairs that we would regard as their contents); and in any case, since that view is atomistic, it seems clear that Stalnaker's view is different.

But just what is Stalnaker's view? The obvious problem with making belief literally a species of indication is that most people are unreliable about a great many things. Of course, even tree rings and fuel gauges can err: that's why Stalnaker defines

indication in such a way that for the rings of a tree to indicate the tree's age, the correlation between rings and age need hold only under "normal or optimal conditions" (p. 12). Given the restriction to *normal* conditions, it does seem highly plausible that *many* of the most important belief states in people–including most of their perceptual belief states–do indicate the states of the world we take as their truth conditions. (Even in the case of perceptual belief states, though, there appear to be exceptions: that is the point of the thesis of the thesis of the theory ladenness of observation, or at least, of the form of that thesis advocated by Feyerabend and plausibly by the "tower argument"; see sections 6 and 7 of Feyerabend 1975.) When we widen our focus far beyond perceptual beliefs, though, it does not seem as if there is any very clear sense in which reliability is "normal". Are we to say that the unreliability of the average person's beliefs about science or about their own motivations or about politics is due to the fact that the average person is under "abnormal conditions"? It is hard to see what content could be given to the notion of abnormality here that licensed this conclusion. (It should also be noted that belief states that do not reliably indicate the states of affairs that we take to be their truth conditions may well reliably indicate something other than their truth conditions: a member of a religious cult is likely to have lots of beliefs that reliably correlate (in both directions) not with the states of affairs that they represent, but rather with the assertions of the head guru of the cult.)

It is not so odd, to be sure, to say that the average person is under less than *optimal* conditions with respect to such beliefs: the conditions couldn't be optimal, since so many of the person's beliefs are untrue! But if optimal conditions are to be defined as conditions under which a person's beliefs are mostly true, then it presupposes a notion of truth conditions; such a notion of optimality can not be used to explain a notion of indication which is used to explain the notion of truth conditions or representation. And it is very hard to see how to spell out a notion of optimality *in non-intentional terms* so as to make typical errors in non-perceptual matters due to non-optimal conditions. It seems initially more attractive to try to loosen the connection between belief and indication, in such a way as to be compatible with the possibility that even in quite normal and optimal conditions our beliefs be unreliable. The view discussed two paragraphs back was an attempt to do this, but it relied on a formally "atomistic" conception of representation. Can we get the desired result without the "atomism"?

Perhaps the answer to this question will emerge from the second stage of Stalnaker's discussion (pp. 15-18 top). Here he suggests that there are more resources available than just the indication relation. He begins by saying:

> Belief and desire, the [pragmatic] strategy suggests, are correlative dispositional states of a potentially rational agent. To desire that P is to be disposed to act in ways that would tend to bring it about that P in a world in which one's beliefs, whatever they are, were true. To believe that P is to be disposed to act in ways that

would tend to satisfy one's desires, whatever they are, in a world in which P
(together with one's other beliefs) were true. (p. 15)

That these facts hold of belief and desire is of course totally uncontroversial; the only
question at issue is whether they can be used to give a non-circular explanation of the
belief relation and the desire relation. The obvious way to try to use them for that purpose
is as follows:

The belief relation is the first component, and the desire relation the second
component, of the unique pair <B,D> such that B maps an agent's belief-states into
propositions and D maps the agent's desire-states into propositions and the agent is
disposed to act in ways that would tend to bring about the propositions in the
range of D given the truth of the propositions in the range of B.

The problem, of course, is that talk of "the unique pair" is totally inappropriate: indeed,
given any relation B that is a candidate for the belief relation (in the sense of being the
first component of such an ordered pair) and *any function at all that maps the set of
possible worlds 1-1 onto itself*, then (calling that function h) we can define another
candidate for the belief relation as follows: state s stands in the relation B_h to the set of
worlds P if and only if s stands in the relation B to $\{h(w)| w \in P\}$. Similarly for desire.
This leaves the contents of belief and desire *totally undetermined*.[15]

I don't claim that this is news to Stalnaker: he makes essentially the same point
(pp. 17-18), and sums it up by saying that "the *content* of belief and desire cancels out on
the pragmatic analysis. Even if that analysis does give us an account of the structure of
explanations of rational action, it gives us no account at all of how beliefs represent the
world." (p. 18). This suggests that as far as the pragmatic picture goes, there is no need
for intrinsically representing contents; only the Boolean structure of mental states (or
some more complicated computational structure like that discussed in the previous
section) is of interest to psychology. But to accept that conclusion would of course be to
give up on thinking of belief as a kind of indication.[16]

Stalnaker does not accept that conclusion; instead, he says, a pragmatic analysis
gives only part of the picture (p. 18). The part of the picture that it gives is (i) that of
determining which states are belief states and which states are desire states; and (ii) that
of helping to determine the content of desire states *once the contents of belief states have
been independently determined.* What determines the contents of belief-states is the
relation of indication. I quote the crucial passage:

If belief is a dispositional state of the kind postulated by the pragmatic analysis,
and also a kind of indication, then we have a fixed point to break into the circle
that is responsible for the relativity of content [i.e. the fact that the attempted
definition in stage two left the contents of beliefs and desires totally
undetermined]. Beliefs have determinate content because of their presumed causal

connections with the world. Beliefs are *beliefs* rather than some other representational state, because of their connection, through desire, with action. Desires have determinate content because of their dual connection with belief and action. (p. 19)

Here Stalnaker talks of "causal connections with the world" instead of the indication relation, but he soon makes it clear (p. 21, first paragraph) that he does not have in mind incorporating such causal connections via an atomistic theory of representation. Rather, the sort of causal connection he has in mind is precisely that provided by the relation of indication.

But if that is the solution, it is hard to see how there is much of an advance beyond stage one. Our task in stage one was to provide an account of content. It was not part of the task to provide a solution to the question of which states were belief states and which desire states; everybody knew that we needed to use the causal role of the states to determine that.[17] It is still the indication relation that Stalnaker uses to account for the *content* of belief states; and the problem I raised in my discussion of stage one is totally unresolved: how are we to bring in indication in our account of belief-content without flying in the face of the obvious fact that most people's beliefs about a great many things are extremely unreliable? Above I sketched one (admittedly quite programmatic) strategy for solving this problem, the strategy that I advocated in Chapters 1 and 2; but that is precisely the strategy to which Stalnaker is trying to find an alternative. It is quite unclear to me what the alternative strategy is to be.

Indeed, there is strong reason to think that *any* remotely promising strategy for solving the problem of dealing with the widespread unreliability in human beliefs must do something that Stalnaker doesn't want: it must invoke a finer structure in belief states than Boolean structure. For when we are interpreting the belief states of a person who is deceived about his own motivations or about what various politicians will do, *why don't we* maximize the agent's reliability by assigning nonstandard truth conditions that will make these belief states express truths? The reason is clear: those belief states are linked to others, via inferential links (how the agent is disposed to reason) and/or by composition from common constituents, and we want any assignment of truth conditions to the agent's states to be *systematic* in that it should respect such inferential and/or compositional links. So any reasonable approach to naturalizing the assignment of S-content (possible worlds content) to belief states must put a premium on assignments of content that are systematic in this way. But that means that *even in giving a naturalistic account of the S-content of belief states*, we must postulate the inferential and/or compositional links among belief states that the assignment of S-content is to represent; moreover, these links must be explainable independently of the assignment of S-content to the states, for otherwise the systematicity condition would be vacuous. (This point is argued in more detail in Field 1990.) The finer structure of belief states is needed not simply in explaining behavior,

but in providing a naturalistic account of S-content: quite the reverse of what Stalnaker claimed in his critique of the use of epistemic possibility, in the passage quoted near the end of Section 1).

Stalnaker does make one suggestion about the naturalization of content that I have not discussed so far. He distinguishes *backward-looking relations* like indication, in which "a state of an object is defined in terms of what tends to cause it", and *forward looking relations*, in which "a state of an object is defined in terms of what it tends to cause" (p. 14). In the main text, the only role that forward-looking considerations are given in solving the problem of intentionality is in determining what states are belief-states and what states are desire-states, and in determining the content of desire states by their role in action once the content of belief-states have already been determined. So the main burden of determining content still rests upon the backward looking (causal and indicational) considerations. Still, a natural suggestion, and one that Stalnaker makes in a footnote, would be to give the forward looking considerations a more prominent role in determining content. This is an attractive idea. It is, of course, neutral to the issue of whether a theory of representation be "atomistic" in Stalnaker's sense, in that an atomistic theory of representation could just as well invoke forward looking considerations in its theory of reference as could a non-atomistic theory. But it suggests a kind of "explicitly holist" approach to determining content, which Stalnaker might find attractive: a crude version of the holist approach would be that the S-content of a state is what is assigned to the state by that systematic function from states to sets of worlds which maximizes "backward looking" and "forward looking" constraints together. Nonetheless, the point of the last paragraph remains: (i) the systematicity condition is important, in that without it the account would never lead to the ascription of familiar kinds of false beliefs, and (ii) the only way to explain the idea of systematicity is to postulate an internal structure in belief states of the kind that Stalnaker is adamantly opposed.

4. Failure of Logical Omniscience

I now want to return to the difficulties that are apparently raised for Stalnaker's view by the fact that people's beliefs appear to be neither consistent nor closed under logical consequence. That there is a *prima facie* problem here is clear: Stalnaker wants to describe the beliefs of not only the "ideally rational believer", but of ordinary believers as well, as relations to sets of worlds that are (at least)[18] *logically* possible. How then can inconsistency and failure of closure under consequence be represented? I will confine my discussion here to the problem raised by inconsistency.

Stalnaker has a discussion in Chapter 5 which appears to me to plausibly handle many cases of inconsistency. Suppose that, in the course of theorizing about the

philosophy of mind, I come to accept the existence of some sort of entity (perhaps propositions or properties or functions or sets) which I reject while doing ontology. If I notice the conflict, then presumably I will take an attitude of less than belief toward at least one of the conflicting existence claims. (There is a good discussion in this chapter of types of acceptance that do not involve full belief.) But if I *don't* notice the conflict, it is quite conceivable that I will believe in the entities in question in one context while disbelieving in them in another. Is this sort of inconsistency among beliefs a problem for one who wants to describe beliefs as relations to S-propositions or sets of possible worlds? Stalnaker points out that it is not. For the natural way to describe what is going on is that I have *two* belief states involving the existence of these entities--one belief-state according to which the entities exist, the other according to which they don't. In one sort of context, one of these belief-states enters into the explanation of my thought and action, in other contexts the other does. If I noticed the conflict, I would presumably try to integrate these states; but prior to my doing so, the description of me in terms of these two conflicting belief states seems intuitively correct. And there is no problem with describing the content of either of the two belief states in terms of S-propositions (or sets of logically possible worlds).

I have no doubt that the above provides an attractive solution to the problem raised by this sort of example of inconsistency of belief. However, there is another sort of example, which Stalnaker does not discuss. Consider the belief-states of Cantor while he was developing set theory. Cantor assumed a "naive comprehension schema" for set theory; one instance of this proved logically inconsistent (Russell's paradox). Surely there is no plausible way to explain this along the lines above: rather, it would appear that one must attribute logical inconsistency to the content of a single belief state.

Stalnaker can not accept the idea that a single belief-state has inconsistent content. The reason is that on Stalnaker's view of content, two things hold. First, any belief-state with inconsistent content has the same content as a state of believing that snow is both white and not white would have, and second, belief-states with the same content have to enter into the explanation of behavior in essentially the same way. If these assumptions are both correct, then we have to find some way of describing Cantor's belief so that it has a consistent content. If we can not do this, that is good reason to reject at least one of the two assumptions. (Which one to give up is largely a matter of convention as to the use of the word 'content'.)

Stalnaker discusses one other method which might be employed in an attempt to describe Cantor's belief state as having a consistent content: the method of going metalinguistic. That is, instead of ascribing to Cantor a belief-state whose content is the infinite conjunction of the comprehension axioms of naive set theory (or some such thing), we ascribe to him a belief-state whose content is that these sentences *express truths*. This maneuver strikes me as having little to intuitively recommend it, but I will

not press that point (except to note that it is an accident of the example that the *prima facie* subject matter of the belief is mathematical; the same sort of inconsistent characterization could arise in describing a subject matter with an uncontroversially non-linguistic content). A more basic doubt is whether the metalinguistic move will work. The move does succeed in giving Cantor's belief-states *consistent* content, since it is consistent to hold that the sentences that in fact characterize sets characterize the decline of the Roman empire instead. The trouble is that it appears to be of little use to Stalnaker to ascribe to Cantor's belief state a content that is *merely* logically consistent; the content needs to be consistent with other facts that Cantor perfectly well knew. And among the facts that he knew, we may assume, were the fact that '$\in$' means set membership, that '&' means and, and so forth, and hence the fact that the axioms of set theory mean just what they do in fact mean. If we now consider the metalinguistic belief *that the naive comprehension axioms express truths and that they mean so and so* (or, *that the naive comprehension axioms express truths and that their meaning is determined by such and such rules*), then that metalinguistic belief is itself inconsistent. So the ascent from object-level propositions to metalinguistic propositions appears to have gained nothing.

A number of people made essentially this objection to an earlier presentation of Stalnaker's views (in his [1976]). In *Inquiry*, Stalnaker considers the objection (in a more abstract setting, not tied to the Cantor example)--see p. 76. His reply is to grant that even the metalinguistic beliefs ascribed to, say, Cantor form an inconsistent set; but then to introduce the discussion (which I've summarized in the second paragraph of this section) of how having an inconsistent set of beliefs is not ruled out by the possible worlds analysis, since the belief-states that compose this inconsistent set may themselves be consistent. It seems to me however that this is plainly inadequate: Cantor's metalinguistic belief that the comprehension axioms express a truth and his other metalinguistic belief about the meaning of these axioms are not at all like a belief in properties while doing philosophy of mind and a disbelief in them while doing ontology; it isn't as if Cantor had two distinct belief-states that could not both be operative at once. Rather, the two belief-states were involved together in explaining his actions, and the discussion in Chapter 5 of "compartmentalized" belief just seems irrelevant to the case at hand.

If, as I have been suggesting, we can not find a reasonable way to describe belief states such as Cantor's as having consistent content, then we must give up Stalnaker's fundamental idea (pp. 4-5, 85) that the explanation of behavior is to be done in terms of mental states described in terms (*solely* in terms) of the real possibilities that they represent. And if we give that up, there is little point in insisting on using the term 'content' in such a way that only the real possibilities that a state represents can be part of the state's content.

It will help clarify the nature of the argument I have given in this section if I contrast it with another argument, the argument described in Section 1 as expressing "an

18

almost universal first reaction" to a possible worlds view of belief and desire. The argument is that such a view of belief and desire can not be adequate, since people can clearly believe that P and disbelieve that Q even when P and Q are logically equivalent. I think that Stalnaker does a good deal in Chapters 4 and 5 to undercut the force of that argument taken by itself (i.e., divorced from an argument against the model of explaining behavior that underlies it). For if we have good theoretical reasons for thinking that the explanation of behavior is to be given in terms of mental states described solely in terms of the real possibilities that they represent, then in many situations where a person sincerely assents to a sentence 'p' and sincerely dissents from a logically equivalent sentence 'q' it seems quite natural to conclude either that one of the sentences 'p' and 'not q' doesn't correspond to his beliefs (as would be determinable from his behavior) or that he is in a state of "compartmentalized belief" not wholly unlike that described early in this section. Stalnaker also points out (pp. 72-74) that even the first alternative does not force one to deny that the person believes that p and disbelieves that q. For phrases of the form 'believes that p' needn't be analyzed straightforwardly in terms of content (e.g. as 'is in a state whose content is the set of possible worlds in which 'p' is true'). Stalnaker does not provide an alternative analysis of phrases of the form 'believes that p', but it does not seem to me that invoking the prospects of such an alternative analysis is at all *ad hoc*.

Still, two things must be noted. First, it is not at all clear that these points will help with *all* instances of "the problem of equivalence": for instance, it is difficult to see how they help with the problem of the $1000 offer in Section 1. That problem arose not because of any assumptions about 'believes that' or 'desires that' locutions, but because the differences between a person trying to trisect a 60 degree angle and a person trying to find a map that needs more than five colors to color seem essential to explaining their non-linguistic behavior. It does not seem to me that one can plausibly explain this difference in behavior in terms of different attitudes to *sentences* like 'I will trisect a 60 degree angle'; and invoking "compartmentalized belief" does not seem substantially more promising.

Second, even if all instances of the problem of equivalence could be treated by going metalinguistic or by compartmentalized belief, that does not help in the least with the problem that has primarily occupied us in this section, namely the problem posed by the apparent existence of belief-states with an inconsistent content. Stalnaker's denial that there can be such belief-states does not stem simply from his decision to treat logically equivalent contents as identical: it would be possible to recognize inconsistent belief-states compatibly with that decision, for that decision merely requires that inconsistent belief-states all have the same content (the empty set of worlds). Rather, what forces Stalnaker to deny that there can be inconsistent belief-states (or at least, to deny that such belief-states can be used in the explanation of intelligent behavior) is his picture of action: the picture according to which the explanation of behavior is to be given in

terms of mental states that are described solely in terms of the possibilities that they represent. It is that basic picture which I have argued in this section to be flawed. And my argument nowhere turned on the use of locutions such as 'believes that'; rather, the argument was that there is no plausible way to describe apparent cases of inconsistent belief-states (such as in the Cantor example) so that they accord with Stalnaker's picture.

Again, the basic difficulty with Stalnaker's views is not his use of the term 'content', but his picture of rational action. But once that picture of rational action is undercut (as I have tried to do in this section), then there is no longer any motivation for insisting that the word 'content' be used in such a way that only the real possibilities that a state represents can be part of its content.

Notes

1. Also, there is no obvious reason to assume that there is anything like a uniquely best assignment of one entity of a given kind to a given mental state--cf. Quinean indeterminacy.

2. Stalnaker himself agrees that beliefs and desires must have representations (p. 22). He wants to avoid taking a stand on the nature of those representations (e.g. on how similar they are to linguistic representation), but the suggestion outlined in the text does not require a stand on that.

3. Admittedly, the state-state relation in question would be based on the overall similarity of the states in their state-world relations. [This overall similarity, incidentally, could take into account both similarity with respect to "pragmatic" relations and similarity with respect to "indication" relations; for a discussion of this distinction (Stalnaker's), see Section 3.]

4. For instance, I explicitly argued for the use of intrinsically representing objects to serve to describe (at least *part of*) the content of mental states in Section V of Chapter 2. (I no longer find the argument offered there conclusive. See note 16 below.)

5. The idea of separating the question of belief-that and desire-that attributions from the question of the contents of belief (and of putting the main theoretical burden on the latter rather than on the former) is not unique to this approach. Indeed, Stalnaker himself argues for such a separation (and for such a view about where the burden of theory should go): see pp. 73-4.

6. For simplicity I have ignored here the fact that Stalnaker's S-propositions needn't be defined at all worlds, and hence that they don't quite correspond to sets of worlds. The discussion would not need to be altered very much were this feature of Stalnaker's views taken into account. (In effect we would simply have to allow mention of the possibility of

extending a given algebra of states to a larger algebra of which the original was a subalgebra.)

7. A less rough statement requires some ideas from the theory of measurement: see Section 2 of the postscript to Chapter 2, or the discussion of an earlier version of this material on p. 10 of Stalnaker's book.

8. My misinterpretation of Stalnaker has apparently led him to a serious misinterpretation of my argument in Chapter 2. Stalnaker points out that there is no difficulty in finding rather unproblematic naturalistic relations that connect objects to S-propositions, and he takes this to undermine my conclusion (Chapter 2, p. ???) that "functionalism does not either solve or dissolve the problem of intentionality". (See *Inquiry*, pp. 14-15). In so taking it, he misconstrues the import of the quoted remark on functionalism. [Indeed, his misconstrual is rather strange, since the goal of the first two sections of my paper (where I was presupposing for the sake of argument that we were to use sets of possible worlds as objects of mental states) was precisely to offer a proposal for how to naturalistically connect people to propositions in ways relevant to belief and desire.] The import of the quoted remark on functionalism was that on a naturalistic view, a *functional* relation between people and propositions is only possible if there are *non-functional* relations between people and propositions which ground each instantiation of the functional relation; so that if we are to solve the problem of intentionality, we need to find such non-functional relations. This is part of what I had been taking Stalnaker to be denying, but it is what he accepts in his first full paragraph on p. 15.

9. Kripke has forcefully stated ([1982], Ch. 2) some serious *doubts* as to whether it is answerable; the simple-minded suggestions that follow in this paragraph ignore the important challenge that Kripke poses.

10. Or more accurately, that the assignment of contents to s_1 and s_2 is such that the content of the first entails the content of the second.

11. The closest thing I've seen to a discussion of this is in Pollock [1974], Ch. 10; the task Pollock is concerned with there is not one of providing psychological models, but it seems closely related.

12. Somewhat more accurately: to believe that p is to be disposed to realize that p follows from your explicit beliefs, and to believe it for that reason, when you think of p or when p is suggested to you. For a good discussion of explicit belief and its relation to non-explicit belief, see Lycan [forthcoming]. (I have lifted the term 'explicit belief' from this paper, and the "somewhat more accurate" formulation above is influenced by Lycan's discussion.)

13. A rather similar suggestion, but concentrating more on inductive inference than logical inference, is obtainable from my [1977]; the idea is developed explicitly in Wagner [1981].

14. My claim that we need a correspondence notion of truth if we are to formulate such indication relations as exist between belief-states (and utterance-states) and the environment was disputed by Grover, Camp and Belnap ([1975], pp. 114-115). It is not entirely easy to evaluate their discussion--for one thing, they appear to smuggle what amounts to a correspondence notion of truth into their alternative account of indication relations, by taking the notions of *believing a proposition* and *asserting a proposition* for granted--but I think that their line of criticism may ultimately be correct. Indeed, I am now tempted by the position not only that a correspondence notion won't be needed to formulate specific indication relations, but also that such a notion won't be needed in theories that make generalizations about indication relations, and won't be automatically generated by theories involving indication either; so that *a fortiori*, notions like reference won't be automatically generated by theories involving indication relations. Consequently, I am tempted to view such notions as truth and reference in their correspondence senses (as opposed to their disquotational senses) as having no important value. But it still seems to me that we need some sort of explanation of whatever indication relations hold between mental states (or utterances) and the external world; and the parenthetical remark at the end of the paragraph indicates that much of the flavor of recent theories of reference would survive in this new guise.

15. Except for the cardinalities of the sets of worlds assigned to each state; and even that could be shown largely undetermined in a more extended discussion.

16. This is *not* to say that in order to give up on the idea of intrinsically representing objects of mental states, one would have to follow Stich [1983] in giving up on the prospects of a theory of indication or in excluding such a theory from psychology. It is merely to say that the results of a theory of indication should not be forced into the format of a theory involving intrinsically representing contents. (I think that some sort of theory of indication, not necessarily forced into the mold of a theory of content, is of considerable importance even within psychology, e.g. in giving an account of how psychological explanations are possible in absence of detailed information about the agent's psychological state and the state of his environment. See Loar [1981], pp. 194-195, and Schiffer [1981], pp. 218-219.)

17. Admittedly, Stalnaker's discussion does make explicit something that I slurred over in the discussion of stage one: that the content of desires needn't be explained in terms of indication. Rather, we can use the indication relation solely to determine the contents of belief-states, and to then use the pragmatic analysis in determining the contents of

desire-states in terms of the contents of belief-states.

18. In fact, he assumes that the worlds are possible in some stronger sense, a sense in which not only denials of logical truths are impossible, but so are denials of mathematical truths (pp. 73-7), denials of true *a posteriori* identity statements between names (pp. 85-6), and denials of certain sorts of true "essentialist" claims (p. 75). The assumption that the worlds are possible in such a stronger sense worsens the difficulties that I will be discussing in Stalnaker's position. However, I have chosen not to stress that fact since I am unclear why Stalnaker wants to use such a strong notion of possibility anyway: none of the advantages he sees for the possible worlds conception seem to require it, and the fact that the boundaries of his strong notion of possibility are left unclarified serves only to make his notion of possible world more obscure than it would be had he stuck to logical possibility.

Incidentally, on p. 58 Stalnaker attaches considerable weight to a rejection of the idea "that there is one domain of all metaphysically possible worlds from which the restricted domains relevant to interpreting different kinds of possibility and necessity are drawn." He does not seem to be merely saying that there are too many possible worlds to form a set; rather, it sounds from the context like a denial of any strongest sort of impossibility, even logical inconsistency. But it is clearly essential to much of Stalnaker's discussion that logical falsehoods (and perhaps mathematical falsehoods, false identity statements between names, etc.) are absolutely impossible: that is an essential premise to Stalnaker's conclusion that they can not be believed. So what he intends by the remark on p. 58 (which is part of an intriguing but obscure discussion of ontological commitment) is unclear.