Functionalism and Propositions
Author(s): John Martin Fischer
Source: *Philosophical Studies: An International Journal for Philosophy in the Analytic Tradition*, Vol. 48, No. 3 (Nov., 1985), pp. 295-311
Published by: Springer
Stable URL: http://www.jstor.org/stable/4319793
Accessed: 24/02/2011 13:15

JOHN MARTIN FISCHER

FUNCTIONALISM AND PROPOSITIONS

(Received 16 October, 1984)

I. INTRODUCTION

My intention in this paper is to explore the relationship between certain 'functionalist' accounts of belief and the possible worlds account of the objects of belief. The argument has been made that acceptance of a plausible kind of functionalist account of belief requires acceptance of the 'coarse-grained' account of propositions given by the possible worlds theory.[1] This result – that a kind of functionalism entails an account of propositions with the same identity conditions as those of the possible worlds approach – is a dramatic and, at first blush, a surprising one. I shall set forth the argument and present some problems for it. My general strategy will be to show that if the functionalist definition of belief is intuitively plausible, then the argument purporting to establish that functionalism requires coarse-grained propositions will not be uncontroversially valid, and if the functionalist account is construed so as to support the validity of the argument, it will be implausible. Thus, the argument is either not uncontroversially valid, or it is unsound.

I take it that propositions are the objects of such attitudes as hope, fear, doubt, desire, and belief; that they are what can be asserted and communicated in contexts of inquiry and exchange of information; and that they are bearers of truth. One might say that a proposition is, very roughly, a way of determining a truth value, given the facts. For the possible worlds theorist, then, a proposition is a function from possible worlds to truth values. Equivalently, a proposition can be identified with the set of possible worlds in which the value of the function is 'true'. Since necessarily equivalent propositions are true in the same set of possible worlds, they are identical propositions, on the possible worlds account.

But if these coarse-grained propositions are the objects of propositional attitudes, then notorious problems emerge. The possible worlds account of

Philosophical Studies **48** (1985) 295–311. 0031–8116/85.10

propositions requires that if a person believes that P, and P is necessarily equivalent to Q, then he believes that Q. And this has seemed to many to be an unpalatable consequence of the possible worlds approach. Suppose, for example, that a person believes that all bachelors are unmarried men; it follows, on the possible worlds theory, that he also believes that all vixens are female foxes. But it seems obvious that one could have the first belief without having the second; indeed, one could believe that all bachelors are unmarried men without ever having considered female foxes. It seems clear that the information possessed by someone who has the belief about bachelors is different from the information possessed by someone having the belief about vixens – the contents of the two beliefs are different. So it appears that the possible worlds account of propositions is unacceptable.

A similar source of problems for the coarse-grained theory comes from mathematical belief. It seems obvious that a person could hold one true mathematical belief without thereby having all true mathematical beliefs. One could, it might be supposed, believe that $2 + 2 = 4$ without believing that $23 \times 17 = 391$. But since the proposition that $2 + 2 = 4$ and the proposition that $23 \times 17 = 391$ are necessarily equivalent, they are the same proposition, on the possible worlds theory; thus, a person who believes one will thereby believe the other. One's reluctance to accept this sort of conclusion might come from two sources. First, it seems that mathematical error is possible – one might falsely believe that $23 \times 17 = 371$ (while nevertheless believing that $2 + 2 = 4$). It appears to be true of this kind of person that he believes that $23 \times 17 = 371$ *rather* than that $23 \times 17 = 391$.

Recently, Ruth Barcan Marcus has argued that it is incoherent to ascribe to a person a belief in a necessarily false proposition ([7]: 321–338). She argues that, just as we can't ascribe to a person knowledge of a *false* proposition, we also can't genuinely ascribe to a person belief in a *necessarily false* proposition. Since the proposition that $23 \times 17 = 371$ is necessarily false, one might argue, following Marcus, that no agent can genuinely be said to believe this proposition.

But even if one adhered to Marcus's position, there is another reason for holding that a person could believe that $2 + 2 = 4$ without believing that $23 \times 17 = 391$. It appears that a person might hold that $2 + 2 = 4$ without ever considering the proposition that $23 \times 17 = 391$; or one might hold the first proposition while withholding judgment about the second-mathematicians who are quite certain about simple propositions often withhold judgment about more complex ones. Thus, one could deny the possible worlds account of propositions without ascribing to an agent belief in a necessarily false proposition. It is clear, then, that Marcus's thesis about the impossibility of believing the impossible doesn't entail the possible worlds theorist's thesis about the necessity of believing the necessarily equivalent.

One way of denying this sort of problem for the possible worlds account is to claim that a mathematical belief is a belief about the relationship between a string of symbols and the one necessary proposition.[2] Thus, when one believes that $2 + 2 = 4$, one believes the *contingent* proposition that the string of symbols '$2 + 2 = 4$' expresses the necessary proposition. It would follow that a belief that $2 + 2 = 4$ would not require a belief that $23 \times 17 = 391$, since one could believe that '$2 + 2 = 4$' expresses the necessary proposition without believing that '$23 \times 17 = 391$' expresses the necessary proposition. On this approach, mathematical beliefs are beliefs about strings of symbols, and one could differentiate between beliefs which are about different strings of symbols – the belief about '$2 + 2 = 4$' would not be necessarily equivalent to the belief about '$23 \times 17 = 391$'. A parallel response might be given concerning the beiefs about bachelors and vixens. It could be argued that when one believes that all bachelors are married men, one believes that 'All bachelors are unmarried men' expresses the necessary proposition; this is clearly a different belief from the belief that 'All vixens are female foxes' expresses the necessary proposition.

But this way of protecting the possible worlds approach to propositions against the unintuitive consequences comes at a steep price. To say that, when one believes that $2 + 2 = 4$, the object of one's belief is the coarse-grained proposition (the set of worlds in which it is true that $2 + 2 = 4$) is to fail to individuate beliefs sufficiently finely; but to say that one's belief is about a string of symbols is to individuate mathematical beliefs too finely.

In order to see this problem, consider, first, a case of ordinary, non-mathematical, contingent belief. Both a Frenchman and a German see the same black cat in front of them. The Frenchman concludes, '*Le chat est noir*', while the German concludes, '*Die Katze ist schwartz*'. One wants to say that they have the same belief – the belief that the cat is black. But unhappily, if belief is taken to be a relationship between a sentence and a set of possible worlds (or, say, a state of affairs), then the two will be said to have different beliefs, insofar as the beliefs involve different sentences. If symbols or sentences are essential to the individuation of beliefs, then beliefs will be cut too finely.

The same sort of problem afflicts the theory of mathematical belief which is being considered here.[3] Suppose I have a belief which I would express using the string, '2 + 2 = 4'. Now imagine that a Roman emperor had a belief which he would have expressed using the string, 'II + II = IV'. There is a perfectly good sense in which I have the same mathematical belief as the Roman emperor had, and yet, if mathematical belief is contrued as essentially involving strings of symbols, then my belief is a different belief from that of the Roman emperor. But this is unacceptable. (Of course, there is a similar problem with the claim that the belief that all bachelors are unmarried men is a belief about words or, perhaps, sentences.) Since beliefs in necessary truths cannot plausibly be construed as beliefs about strings of symbols, the problem for the possible worlds theorist remains acute: it seems that one can have (say) a true simple mathematical belief without thereby having *all* true mathematical beliefs.

The consequences of the possible worlds theory of propositions discussed above are unattractive, and it would be extremely surprising to discover that accepting an appealing and plausible functionalist account of belief would require acceptance of the coarse-grained theory of propositions given by the possible worlds approach. But Robert Stalnaker has produced a powerful and ingenious kind of argument which purports to establish this surprising result ([12]: 79–92). In investigating Stalnaker's argument, I shall first discuss a rather simple argument (which is clearly not Stalnaker's argument); having diagnosed the failings of this simple argument, I shall elaborate a more sophisticated argument (which is closer to Stalnaker's intentions, and which may be a fair articulation of his argument). I claim that this sophisticated argument is defective in a way which is parallel to the inadequacy of the simple argument. Thus, no reason will have been provided to force a functionalist to accept the unpalatable consequences of the possible worlds theory.

II. THE FIRST ARGUMENT

Before setting out the argument, a few words about the functionalist approach to mental states such as belief and desire are in order. Various radically different sorts of theories of mental states have been called 'functionalist' theories ([1]: 171–184; [10]: 93–119). What is common to all forms of functionalism is the claim that mental states are to be characterized in terms of their causal roles. Mental states are seen as states which stand in certain sorts of causal relations to sensory stimulations, behavioral outputs, and other mental states. What distinguishes functionalism from 'type physicalism' about mental states is the claim that the same causal role (and thus the same type of mental state) can be 'realized' in physically different sorts of ways – this is the thesis of the 'compositional plasticity' of mental states.[4] And what distinguishes functionalism from behaviorism is the claim that the causal relations in terms of which particular types of mental states are construed may involve other mental states, as well as sensory stimulations and behavioral outputs ([1]: 175–176).

One kind of theory of belief which has been described as a functionalist theory claims that one of the mental states by reference to which belief is to be defined is a 'mental representation'. A mental representation is some sort of internal state with representational properties which is often claimed to have a structure similar to the structure of sentences.[5] Hartry Field presents such a theory in [3]. Field claims that a theory which posits mental representations as part of the definition of belief might be functionalist:

> According to the crude description above, a state of an organism is a state of believing that *p* if the state is causally connected to inputs and outputs and to other psychological states in the right sort of way; but until we know what the other psychological states are that it must be causally connected to, and what is 'the right sort of way' for it to be connected to them, we are in no position to determine whether such a functional theory of belief requires a system of inner representations. (E.g., why couldn't one of the other states to which a state of belief that *p* must be causally connected be a state of having an inner representation of the proposition *p*?) ([3]: 79).

It is unclear to me whether this sort of theory is properly considered functionalist. This is because it might be thought that, in order to be a functionalist theory of belief, the states a theory claims to be causally related to beliefs must *themselves* be functionally definable, and it is certainly unclear whether mental representations can be given functional characterizations. In any case, I will not be focusing here on theories which posit mental representations as essential to belief; on the kind of functionalist theory of belief considered by Stalnaker, mental representations play no essential role. (Of course, Stalnaker needn't deny the existence of mental representations or claim that *no* beliefs involve such states; his position is simply that mental representations are not essential to belief.)

The kind of functionalism about belief and desire which Stalnaker has in mind is a systematization of common sense, intuitive views about the nature of these mental states. The picture is Aristotelian; belief and desire are coordinate states which issue in rational action, and we explain a rational agent's actions by adverting to his desires and beliefs ([12]: 80–81). It will be important, in order properly to assess Stalnaker's argument, to generate explicit functionalist characterizations of belief and desire. This is something which most functionalists decline to do, assuming that *some* precise account could in principle be constructed, and resting content with the more abstract claim that mental states are characterizable (in some way or another) in terms of inputs, outputs, and other mental states. For certain philosophical purposes, the more abstract claim is all that is pertinent; but in order to evaluate the position that a plausible functionalism requires coarse-grained propositions, it is necessary to make explicit the functionalist accounts of belief (on which we shall be focusing) and desire.

Consider a person who sees ominous black clouds gathering in the sky. Assuming that he wants to stay dry, then given that he is disposed to get his umbrella ready for use, it is plausible to say that he believes that it is going to rain.[6] Also, assuming that a person believes that it is going to rain, then if he has a disposition to get his umbrella ready for use, we might say that he desires to stay dry. These simple insights give rise to a very rough first approximation to the functionalist accounts of belief and desire:

(F_1) *S* believes that *P* if and only if *S* is disposed to act in a way which would maximally satisfy his desires, whatever they are or might be, if *P* were true.

S desires that *P* if and only if *S* is disposed to act in a way which would bring it about that *P*, if his beliefs, whatever they are or might be, were true.

The functionalist account of belief makes use of the notion of desire, and the account of desire makes use of belief. Thus, each particular account does not, by itself, reduce an intentional notion to entirely non-intentional phenomena. However, the two accounts *taken together* could be part of a theory of mental concepts which reduces them – all at once – to non-intentional phenomena ([1]: 176–177).

The functionalist account of belief needs to be elaborated and refined. First, there is an obvious and significant problem. Suppose that, on any intuitive criterion of belief, a person S believes that five plus five are ten and wants ten apples. Further, we stipulate that S believes (on any acceptable criterion of belief) that there are five apples in box x, five apples in box y, and four apples in box z; in fact, there are five apples in box x, four in box y, and five in z. Given S's want to have ten apples and his beliefs about the contents of the boxes, S is disposed to choose boxes x and y. It was understood that S believes that five and five are ten, yet S is *not* disposed to act in a way which would satisfy his desires, if five and five were ten ([5]: 24).

In order to make this point, we needn't choose an example involving mathematical belief. Suppose that (intuitively speaking) a person S believes that there are apples in box x and that he desires an apple. In fact, there are apples in box x, but (unbeknownst to S), the situation is such that if he were to reach for the box, someone else T would be alerted to the availability of apples and would snatch the box away. Now, given S's desire for apples, he is disposed to reach for the box; but he is *not* disposed to act in a way which would *in fact* satisfy his desires, if there were apples in box x, because if he were to reach for the box, he wouldn't succeed in getting it.

It's not true that S is disposed to act in such a way that his desires would in fact be satisfied, if there were apples in box x; rather, S is disposed to act in a way which he *believes* would satisfy his desires, if there were apples in box x. Similarly, in the mathematical case, S is not disposed to act in a way that would satisfy his desires, if five plus five were ten; rather, S is disposed to act in a way which he believes would satisfy his desires, if five plus five were ten. This suggests a refined version of the functionalist account of belief:

(F_2) S believes that P if and only if S is disposed to act in ways which he believes would maximally satisfy his desires, whatever they are or might be, if P were true ([5]: 24).

The revised account of belief avoids the counterexamples to (F_1), but it may seem to have obvious problems of its own. I shall explore the problems with (F_2) below. But, at this point, I wish to set out an argument using (F_2) which is similar in structure to Stalnaker's argument that functionalism requires coarse-grained propositions; this will *not* be Stalnaker's argument, since he is not committed to (F_2). But the discussion of this argument will shed light on a parallel argument which employs a more sophisticated analysis of belief.

Argument I

hypothesis	1. *S* believes that *P*.
(F_2)	2. *S* believes that *P* iff *S* is disposed to act in a way which he believes would maximally satisfy his desires, if *P* were true.
1,2	3. *S* is disposed to act in a way which he believes would maximally satisfy his desires, if *P* were true.
hypothesis	4. *P* and *Q* are necessarily equivalent propositions.
3,4 semantics for conditionals	5. *S* is disposed to act in a way which he believes would satisfy his desires, if *Q* were true.
5, (F_2)	6. *S* believes that *Q*.

This sort of argument might appear to show, from an acceptance of a version of functionalism, that if a person believes that *P*, and *P* is necessarily equivalent to *Q*, then he believes that *Q*. And this is precisely what the possible worlds account of propositions requires. Of course, the argument does nothing to assuage the doubts of a skeptic about possible worlds; that is, the argument is not intended to prove that there exist possible worlds.[7] Rather, the argument claims that, if one is a certain sort of functionalist, one must adopt an account of propositions which has the same identity conditions as those of the possible worlds account. Whereas the skeptical worries about possible worlds might persist, this result would make the possible worlds account of propositions considerably more attractive.

There are, however, some obvious problems with the argument. Immediately, one might wonder about the inference from (3) and (4) to (5). One way of explaining the inference – a way which makes use of the possible worlds apparatus – is as follows. If *P* and *Q* are necessarily equivalent propositions, then *P* and *Q* are true in the same set of possible worlds; thus, in any worlds in which the antecedent of (3) is true, the antecedent of (5) is true. And since the consequents of (3) and (5) are the same, it follows that if (3) and (4) are true, then (5) must be true.[8] But this argument assumes that (3) and (5) are analyzable straightforwardly as conditionals, and it is not at all clear that they are properly construed as conditionals of the sort required by the argument.

In order to see this, note that there is an ambiguity in (F_2) which plays a crucial role in the argument. Intuitively, there are the following two non-equivalent readings of (F_2):

(F_{2a}) *S* believes that *P* if and only if (*S* is disposed to act in a way which he believes would maximally satisfy his desires, whatever they are or might be), if *P* were true.

(F_{2b}) *S* believes that *P* if and only if *S* is disposed (to act in a way which he believes would maximally satisfy his desires, whatever they are or might be, if *P* were true).

If *S* believes that *P*, (F_{2a}) attributes to *S* an unconditional disposition, on the condition that *P* is true. In contrast, if *S* believes that *P*, (F_{2b}) unconditionally attributes to *S* a *conditional* disposition – the disposition to act in a way which he believes would satisfy his desires on the condition that *P* is true.[9]

We can capture the intuitive difference between (F_{2a}) and (F_{2b}) by regimenting them as follows. Let '$SD_S(r)$' be '*r* maximally satisfies *S*'s desires', '$P_S(x)$' be '*x* is performed by *S*', '$B_S R$' be '*S* believes that *R*', '$D_S(d)$' be '*S* has a total set of desires $d_1, ..., d_n$', '$O_S(r)$' be '*r* is open to *S* (*S* is free to do *r*)', and '>' be the connective for the subjunctive conditional:

$$(F_{2a}) \quad B_S(P) \ \text{iff}\ P > (\forall d)(\forall r)[D_S(d) > B_S(O_S(r) \ \&\ SD_S(r)) > P_S(r)]$$
$$(F_{2b}) \quad B_S(P) \ \text{iff}\ (\forall d)(\forall r)[D_S(d) > B_S(O_S(r) \ \&\ (P > SD_S(r))) > P_S(r)]$$

This regimentation highlights the contrast between the two formulations; in (F_{2a}) the variable '*P*' is not within the scope of the belief-operator, while in (F_{2b}), '*P*' falls within the scope of the belief-operator. If *S* believes that *P*, then (F_{2a}) says that, if *P* were true, then *S* would have a certain disposition – the disposition to maximally satisfy his desires. In contrast, if *S* believes that *P*, then (F_{2b}) says that *S* is disposed (roughly) as follows: for *any* total set of desires he might have, if he were to believe that if *P* were true, doing *r* would maximally satisfy his desires, then he would do *r*.

The dilemma for Argument I is as follows. On (F_{2a}) the argument is valid but not sound (since (F_{2a}) itself is obviously unacceptable), and on (F_{2b}) the argument is not uncontroversially valid. I shall now elaborate on this dilemma.

(F_{2a}) is a manifestly inadequate way of attempting to capture the functionalist insight. Note that since any mathematical truth is true in all possible worlds, the functionalist account of belief in mathematical truth reduces, on (F_{2a}), to:

$$B_S(P) \text{ iff } (\forall d)(\forall r)\,[D_S(d) > B_S(O_S(r)\ \&\ SD_S(r)) > P_S(r)]$$

Here, the truth condition for belief in any mathematical proposition depend only on whether the agent is disposed to behave in a way which he believes would satisfy his desires. That is, it will follow from (F_{2a}) that *all* true mathematical beliefs are held by any agent who is disposed to behave in a way which he believes would satisfy his desires; thus, any rational agent – even one who never considered *any* mathematical proposition – would believe *all* mathematical propositions.

There is even a *worse* problem with (F_{2a}). The problem just discussed applies not only to mathematical truths, but to any true proposition. When the relevant proposition P is a true proposition, (F_{2a}) will ascribe to S a belief in P just in case S is a rational agent (an agent who is disposed to behave in a way which he believes would satisfy his desires); this result is clearly unacceptable.

Adopting (F_{2b}) avoids these problems, but the validity of the argument is now in question. On (F_{2b}) the transition from (3) and (4) to (5) is illegitimate, since it involves an illicit substitution into a belief context. Premises (3) and (5) are not conditionals of the form required by the argument; substitution of necessary equivalents within belief contexts is supposed to be licensed by the *conclusion* of the argument, but clearly cannot be used to generate the conclusion without begging the question.

Not only does (F_{2b}) fail to support the inference from (3) and (4) to (5), but (F_{2b}) appears to be viciously circular ([5]: 24). (F_{2b}) purports to give an account of believing that P, but the *definiens* specifies a conditional belief of which P is a component. This clearly involves an unacceptable circularity.

Let me summarize the discussion of the argument. Argument I uses a functionalist account of belief which is ambiguous. Reading (F_{2a}) makes the argument valid but obviously unsound. And reading (F_{2b}), while not subject to the same sort of implausible consequences as (F_{2a}), does not make the argument valid; further, (F_{2b}) appears to be viciously circular. What is needed is a non-circular, plausible functionalist account of belief employing which renders the argument valid.

III. THE SECOND ARGUMENT

Consider again the person who believes that five plus five are ten, wants ten apples, believes that there are five apples in box x, five in y, and four in z.

(Remember that, in fact, there are five apples in x, four in y, and five in z.) This person is *not* disposed to act in a way which would satisfy his desires, if five and five were ten, but he *is* disposed to act in a way which would satisfy his desires, if five and five were ten and his other beliefs were true. Also, consider again the person who wants apples and believes there are apples in box x; while he is not disposed to act in a way which would satisfy his desires, if there were apples in box x (since he would cause T to intervene), he is disposed to act in a way which would satisfy his desires, if there were apples in box x and his other beliefs (about the efficacy of reaching for the box, etc.) were true. The following sort of account, then, might seem to avoid some of the problems considered above:

(F_3) Let $\{\Gamma - P\}$ be the set of S's beliefs other than P. S believes that P if and only if P is a member of a set of propositions Γ such that S is disposed to do what would maximally satisfy his desires, whatever they are or might be, if all members of Γ were true.[10]

(F_3) might appear to be circular insofar as it uses the notion of belief in the *analysans*. But it is not circular, since it defines belief that P in terms of beliefs *other than P*; thus, as above, it could be a part of a total psychological theory which reduces mental concepts – all at once– to non-intentional phenomena. This sort of account of belief can be used in an argument structurally parallel to the first argument. I begin by setting out the argument, after which I evaluate (F_3).

Argument II

hypothesis	1′.	S believes that P.
(F_3)	2′.	S believes that P iff P is a member of a set of propositions Γ (where $\{\Gamma - P\}$ is the set of S's beliefs other than P) such that S is disposed to do what would maximally satisfy his desires, whatever they are or might be, if all members of Γ were true.
1′, 2′	3′.	P is a member of a set of propositions Γ such that S is disposed to do what would maximally satisfy his desires, whatever they are or might be, if all members of Γ were true.
hypothesis	4′.	P and Q are necessarily equivalent propositions.
3′, 4′	5′.	Let Γ' be the same as Γ, except with Q replacing P. Q is a

semantics for conditionals		member of a set of propositions Γ' such that S is disposed to do what would maximally satisfy his desires, whatever they are or might be, if all members of Γ' were true.
$5'$, (F_3)	$6'$.	S believes that Q.

The second argument yields Stalnaker's conclusion without the obvious defects of the first argument. But there are significant problems with (F_3). The first problem is that it seems to imply that any agent with inconsistent beliefs will believe *all* propositions. If S's beliefs are inconsistent, then all his beliefs are true together in *no* possible world; this would mean that the conditional in (F_3) would be vacuously true, insofar as the antecedent would be true in no possible world. Since almost all persons hold inconsistent beliefs, (F_3) faces a severe problem.

I shall now briefly sketch a possible response on behalf of (F_3).[11] Just as an agent may have various internally consistent but uncoordinated sets of desires, he may also have various internally consistent sets of beliefs which he fails to integrate properly. Thus, (F_3) would only apply to suitable internally consistent belief systems of an agent. That is, $\{\Gamma - P\}$ wouldn't include *all* of S's beliefs other than P, but it would be an internally consistent subset of S's beliefs other than P.

This approach relativizes belief to a belief system; one may believe a proposition relative to one belief system but not another. There are certainly problems with the elaboration of such an approach; certain systems of belief will be more 'relevant' or important by reference to which to fix an agent's beliefs. For example, there will clearly be 'trivial' internally consistent subsets of an agent's beliefs – one could simply choose very small subsets; on the relativized approach to belief, some way of distinguishing the trivial from the important belief subsystems is required, and it is by no means easy to see how this could be done. I believe that, even if such an approach to belief could be worked out, there are other problems with (F_3) – problems which can be seen to be parallel to the problems with (F_2).[12]

Note again that (F_3) ascribes to an agent who believes that P a certain sort of disposition:

(F_3) S believes that P (relative to Γ) if and only if P is a member of a set of propositions Γ (where $\{\Gamma - P\}$ is a suitable set of S's beliefs other than P) such that he is disposed to do what would maximally satisfy his desires, whatever they are or might be, if all members of Γ were true.

How is such an ascription to be understood, consistently with the non-question-begging validity of the inference from (3′) and (4′) to (5′)? I claim that (F_3) is ambiguous in a way which is parallel to the ambiguity of (F_2). Consider, first, the following way of capturing (F_3), a way which is similar to (F_{2a}). Let '$B_S(P)_\Gamma$' be 'S believes that P, relative to Γ', and '$T\Gamma$' be 'all members of Γ are true' (where $\{\Gamma - P\}$ is a suitable set of S's beliefs other than P):

$$(F_{3a}) \quad B_S(P)_\Gamma \text{ iff } (\exists\Gamma)\,[(P \in \Gamma)\,\&\,T\Gamma] > (\forall d)(\forall r)\,[D_S(d) > SD(r) > P_S(r)]$$

It is clear that in (F_{3a}), 'Γ' occurs in a 'transparent' context; if (3′) and (4′) are true, (5′) must be true – this way of understanding (F_3) supports the validity of the argument. But, whereas (F_{3a}) allows the inference from (3′) and (4′) to (5′), it is an implausible account of belief.

To see this, let us consider a case of a person S who has never entertained the *true* proposition, 'It was minus ten degrees Fahrenheit at the North Pole on July 4, 1980.' Call this proposition, '*NP*'. Imagine that, not only has S never explicitly entertained *NP*, but his behavior cannot be said to 'presuppose' its truth – *NP* is remote and irrelevant to S's concerns. Surely, we shouldn't ascribe to S this belief. But *NP* is apparently a member of a set Γ of propositions (where $\{\Gamma - NP\}$ is an internally consistent set of S's beliefs other than *NP*) such that S is disposed to satisfy his desires, if all members of Γ were true. This is because it is plausible to suppose that there is some suitable set $\{\Gamma - NP\}$ of S's beliefs such that S is disposed to satisfy his desires, if $\{\Gamma - NP\}$ were true. And since *NP* is actually true and not ruled out by any of S's beliefs, *NP* would also be true, if $\{\Gamma - NP\}$ were true; thus, it follows that S is disposed to satisfy his desires, if Γ were true. Here, (F_{3a}) seems to ascribe to S belief in *NP*. If (F_{3a}) in fact had this sort of result, then it would expand the scope of belief in a way parallel to the way in which (F_{2a}) unduly expands the scope of belief – it would ascribe to S a belief in a proposition never considered and not presupposed by his behavior.

There might be a way of blocking this unwelcome result.[13] Note that (F_{3a}) asks us to consider not just actual desires of the agent, but possible ones as well. And (F_{3a}) requires, for belief that P, that S have the appropriate disposition, for *any* total set of desires. It is surely possible that S have a desire of the following sort: the desire to eat an apple if and only if *NP*. This kind of 'biconditional' desire is different from an ordinary unconditional desire to eat an apple. Now, if S were to have such a desire, he *wouldn't* eat an apple (since he would have no inclination to think *NP* true); but since *NP* is true (and would still be true if S had the biconditional desire), eating an apple would satisfy his desire to eat an apple iff *NP*. This claim rests on the plausible assumption that if one has a (biconditional) desire to X iff P, then if P is true, one satisfies the desire insofar as one X's. But we can now see that S is *not* disposed to satisfy his desires, whatever they are or *might be*, if all members of Γ were true; if he had a certain kind of biconditional desire, he wouldn't satisfy it. Thus, he needn't be said to believe *NP*, on (F_{3a}) – there exists a total set of desires relative to which he does not have the appropriate disposition.

Are there biconditional desires such as the desire to eat an apple iff *NP*? It's hard to see why not, but I claim that consideration of such desires avoids the broadening of the scope of belief (on (F_{3a})) only at the expense of excessively restricting it. Imagine a very ordinary case in which a person S is thought (by any intuitively acceptable criterion of belief) to believe a proposition. Suppose that S clearly sees Muffin the cat sitting on the mat and that S's behavior is obviously appropriate to a belief that the cat is on the mat – S avoids treading on the mat on his way to the refrigerator, etc. We want to say that S believes that the cat is on the mat, but I claim that (F_{3a}) will not allow us to say this. Call the proposition that the cat is on the mat '*CM*'. But suppose that, unbeknownst to S, *NP* is actually true. (Assume that S has no beliefs which bear on *NP*.) I shall argue that, no matter how one chooses the non-trivial set $\{\Gamma - CM\}$ of S's beliefs other than *CM*, S will not be said (on (F_{3a})) to believe *CM*.

For any appropriate set of S's beliefs, $\{\Gamma - CM\}$, if $\{\Gamma - CM\}$ were true, then *NP* would also be true (since it is actually true and not ruled out by S's beliefs). Thus, if Γ were true, *NP* would be true. Now, exactly as above, we can see that *CM* is *not* a member of a suitable set of propositions Γ such that S is disposed to maximally satisfy his desires, whatever they are or might be, if all members of Γ were true. For suppose that he had a desire to eat an apple iff *NP*; he wouldn't be disposed to eat an apple, and thus he wouldn't be disposed to do what would satisfy his desires, if all members of Γ were true. Thus on (F_{3a}), we wouldn't be able to say that S believes that the cat is on the mat; the range of belief is unacceptably diminished. The point is that precisely the way in which we blocked the expansion of belief – consideration of certain biconditional desires – generates an equally unacceptable contraction of belief.

The problem may seem to be that, in asking whether S believes that *CM*, we've considered a desire which is 'irrelevant' to *CM* – the desire to eat an apple iff *NP*. This suggests a restriction of desires considered by (F_{3a}) to desires relevant to the proposition belief in which is being assessed. But I don't see any non-question-begging way to separate the 'relevant' from the 'irrelevant' desires so that a plausible account of belief is maintained. Suppose, for instance, that (in the case discussed above) Muffin, is, unbeknownst to S, Max's favorite cat. Now, I don't see why the desire to eat an apple iff Max's favorite cat is on the mat is not 'relevant' to *CM* – the proposition that the cat is on the mat. But if this is so, then even the restricted version of (F_{3a}) will rule out S's believing that the cat is on the mat.

To avoid the implausible results of (F_{3a}), we could modify it in a way which is parallel to (F_{2b}):

$$(F_{3b}) \quad B_S(P)_\Gamma \text{ iff } (\exists\Gamma)\ [(P \in \Gamma)\ \&\ (\forall d)(\forall r)(D_S(d) > B_S(T\Gamma > SD_S(r)) > P_S(r))]$$

If S believes that P, then (F_{3b}) says (roughly) that S is so disposed that for any total set of desires, if he were to believe that if all members of Γ were true, then doing r would satisfy his desires, then he would do r. (F_{3b}) allows us to say that S believes *CM* in virtue of the following: if S were to believe that if *CM* were true, doing r would satisfy his desires (even including a desire to eat an apple iff *NP*), then he would perform r.

But the problems with (F_{3b}) are similar to those of (F_{2b}). In (F_{3b}), 'Γ' occurs within a belief context, so the inference from (3′) and (4′) to (5′) is not uncontroversially valid. Not only is the non-question-begging validity of the argument sacrificed, but (as with (F_{2b})) it seems as if the account of belief is viciously circular: belief that P is defined in terms of a conditional belief whose antecedent is a conjunction, one conjunct of which is 'P'.

Let me summarize. The refined functionalist accounts of belief have problems which are parallel to the problems of the simpler accounts. Construed in a way in which the argument can be seen to be valid, the account of belief becomes unacceptable; and construed in a way which avoids the unacceptable consequences, the account renders the validity of the argument questionable.

IV. CONCLUSION

Some have argued, following Stalnaker, that a plausible functionalist account of belief requires coarse-grained propositions. I have explored a class of functionalist accounts, and my argument has been that, in this class, there is no account which meets *all* of the following conditions: it is plausible, non-circular, and allows for the validity of the argument to coarse-grained propositions. In producing this argument, I believe that I have shown that it might be open to a functionalist to adopt fine-grained propositions; thus, one might be a functionalist without holding that all mathematical beliefs are about strings of symbols (and that the belief that all bachelors are unmarried men is a belief about words).

My project in this paper has been minimal in the following sense. I have *not* argued that *no* functionalist account of belief which meets the three conditions can be produced; rather, I have simply explored the inadequacies of certain sorts of accounts. I think that this is useful insofar as it makes clear the challenges to be met by an account of belief which can play the required role in the argument to coarse-grained propositions. It is compatible with my position that such an account is forthcoming, insofar as I have not produced a functionalist theory of belief which is clearly non-circular, plausible, and which yields fine-grained propositions. Of course, it is also compatible with my position that no plausible, non-circular functionalist account of belief of any sort can be produced. My argument has been that, *if* one construes such mental states as belief as functional states, no convincing argument has yet been produced that they require coarse-grained objects.[14]

NOTES

[1] See [12]: 79–91; for a useful discussion, see [9]: 93–103. Recently, Stalnaker has further developed his position in: [13].

[2] This kind of approach is adopted by Stalnaker at [12]: 87–88; and [13]: 71–78.

[3] I thank Phillip Bricker for bringing this point to my attention.

[4] See [2]: 67–106. I borrow the term, 'compositional plasticity', from Boyd.

[5] Block calls this sort of theory functionalist, though of a different sort from kinds of functionalism which don't posit mental representations. ([1]: 171.) For an elaboration of a theory of belief which involves internal representations with linguistic structure, see [4].

[6] For simplicity's sake, I will be discussing a belief and desire theory which does not incorporate degrees of belief and desire. The necessary refinements, employing the notions of subjective probability and utility, could easily be made.

[7] This point was emphasized to me by Jon Dupre and John Perry.

[8] For the sort of semantics for conditionals which would validate this inference, see [6] and [11].
[9] For a useful presentation of the distinction between a conditional ascription of an unconditional disposition and an unconditional ascription of a conditional disposition, see: [8]: 249–274.
[10] Robert Stalnaker suggested this account to me; he presents a similar account in: [13]: 15.
[11] Robert Stalnaker has developed this kind of approach in correspondence and also (at greater length) in: [13]: 79–99.
[12] In what follows, I shall sometimes suppress mention of the relativization of belief to belief-systems, for the sake of simplicity.
[13] Robert Stalnaker suggested to me this method of avoiding the expansion of belief.
[14] I have benefitted from discussions with Nicholas Asher, Phillip Bricker, Jon Dupre, and John Perry. I am particularly grateful to Anthony Brueckner for his careful comments, and to Robert Stalnaker, who has generously commented on previous versions of this paper. Part of my work on this paper was supported by a Fellowship for Independent Study and Research From the National Endowment for the Humanities.

REFERENCES

[1] Block, Ned: 1980, 'Introduction: What is functionalism?', in Block, N. (ed.), Readings in the Philosophy of Psychology, V. 1, (Harvard University Press, Cambridge).
[2] Boyd, Richard: 1980, 'Materialism without reductionism: What physicalism does not entail', in Block, N. (ed.), Readings in Philosophy of Psychology, v. 1. (Harvard, University Press, Cambridge).
[3] Field, Hartry: 1980, 'Mental representations', Erkenntnis 13 (1978), reprinted in Block, N. (ed.), Readings in Philosophy of Psychology, v. 2, (Harvard University Press, Cambridge).
[4] Fodor, Jerry A.: 1973, The Language of Thought, (Harvard University Press, Cambridge).
[5] Grice, H.P.: 1975, 'Method in philosophical psychology (From the banal to the bizarre)', Proceedings and Addresses of the American Philosophical Association 48.
[6] Lewis, David: 1973, Counterfactuals, (Harvard University Press, Cambridge).
[7] Marcus, Ruth Barcan: 1983, 'Rationality and believing the impossible', The Journal of Philosophy 80.6.
[8] Pears, David: 1971, 'Ifs and cans', Canadian Journal of Philosophy 1.2.
[9] Powers, Lawrence: 1976, 'Comments on 'Propositions' ', in MacKay, A.F. and Merrill, D.D. (eds.), Issues in the Philosophy of Language, (Yale University Press, New Haven).
[10] Shoemaker, Sydney: 1981, 'Varieties of functionalism', Philosophical Topics, 12.1.
[11] Stalnaker, Robert: 1968, 'ATheory of Conditionals', in Rescher, N. (ed.), Studies in Logical Theory (Blackwell, Oxford).
[12] Stalnaker, Robert: 1976, 'Propositions', in MacKay, A.F. and Merrill D.D. (eds.), Issues in the Philosophy of Language, (Yale University Press, New Haven).
[13] Stalnaker, Robert: 1984, Inquiry (MIT Press, Cambridge, Mass.).

Department of Philosophy,
Yale University,
P.O Box 3650,
Yale Station,
New Haven, CT 06520,
U.S.A.