Mental Causation
Proceedings of the Third International Colloquium in Philosophy of Mind
André Leclerc, Giovanni de Queiroz and Michael B. Wrigley (eds.)

CDD: 128.2

MENTALISTIC EXPLANATION AND MENTAL CAUSATION

SANFORD C. GOLDBERG*

Department of Philosophy,
University of Kentucky,
1427 Patterson Office Tower,
LEXINGTON, KY 40506-0027
USA

scgold@pop.uky.edu

Abstract: *In this paper I present an internal difficulty for the hypothesis that mentalistic explanation is causal explanation. My thesis is that intuitively acceptable mentalistic explanations appear to violate constraints imposed by the mental causation hypothesis.*

Key-words: *explanation; mental causation; belief.*

1. Let a mentalistic explanation of subject *S*'s behavior *H* be an explanation that involves the ascription to *S* of propositional attitudes, where *S*'s having those attitudes is (part of) what explains *S*'s *H*-ing. Minimally such an explanation will include at least one ascription of

* I would like to thank the audience at the Third International Colloquium in the Philosophy of Mind, on the topic of Mental Causation, which took place in João Pessoa, Brazil, May 20-23, 2002. And in particular I would like to thank several people for helpful comments and suggestions: Lynne Baker, Pascal Engel, Andre Furhmann, Osvaldo Pessoa Jr., Eduardo Rabossi, and Daniel Vanderveken. A special thank you to André Leclerc for serving as commentator at the session at which I gave this paper.

belief and one ascription of desire, but typically it will include more than one of either or both. Where 'B_j' designates a belief ascription and 'D_k' designates an ascription of desire, we can then say that a mentalistic explanation consists of a set Σ of propositional attitude ascriptions B_1,... B_n and D_1, ... D_m.

In this paper I want to examine the relation between mentalistic explanation and mental causation. Taking a mentalistic explanation to include a set Σ of propositional attitude ascriptions, I want to ask what constraints are placed on Σ (i.e., the particular ascriptions) if we assume that at least some of the members of Σ are supposed to describe the mental causes of the behavior. This is a large issue, one that can be approached from several distinct directions. My aim is to discuss how considerations from mental causation constrain Σ, and then to proceed to argue that these constraints appear to be violated by intuitively acceptable mentalistic explanations. This approach reflects the fact that I have less confidence in the clarity and utility of metaphysical notions such as causation, than I do in our native ability to discern in particular cases whether a given mentalistic explanation is acceptable. Of course, those who disagree with the conclusions I draw on the basis of such an approach are open to treating my argument as the basis for a *reductio* of that approach (or of one of the premises it employs).

Before proceeding to my argument, some additional assumptions and some terminology will be helpful. I take propositional attitudes such as belief to be disposition-like states of a subject. Thus, believing that George W. Bush is President of the USA is a matter of being in a certain complex dispositional state, one that is individuated in terms of its characteristic manifestations (and the propositional content present in at least some of these manifestations). Such manifestations might include utterances of 'George W. Bush is President of the USA', assentings to 'Is George W. Bush President of the USA?', appropriate utterances in other circumstances where what is at issue is the identity of the current President of the USA, and so on. One class of characteristic

manifestations of the belief that George W. Bush is President of the USA is worthy of being singled out: these are the *conscious thinkings* that George W. Bush is President of the USA. These are particular, dated events in the conscious life of a given subject. These are worthy of being singled out, since if you hold that mentalistic explanation is causal explanation, and if you accept further that causes are events that precede their effects, then when it comes to offering mentalistic explanations in particular cases you may find that the best (and perhaps only!) candidate for mental causes are conscious thinkings (see below). Of course, while conscious thinkings that *p* are a manifestation of the belief that *p*, the ascription '*S* believes that *p*' can be true even while *S* is asleep, or thinking of other matters.

One final terminological matter. As I shall use it here, 'thinking that *p*' involves a commitment to the truth of *p*. This is admittedly nonstandard usage, since on this usage '*S* is thinking that *p*' entails '*S* believes that *p*', whereas on standard usage the former is neutral on the question of *S*'s commitment to (belief in) the proposition that *p*. When I want to designate a commitment-neutral thinking-like state, I will use 'entertain'.

2. The doctrine on which I will focus can be formulated as follows ('EPA' for 'Explicit Propositional Attitudes'):

(EPA) If subject *S*'s behavior *H* (occurring at time *t*) is to be correctly explained by a mentalistic explanation whose ascriptions Σ include $B_1 \ldots B_n$ and $D_1 \ldots D_m$, then for at least one of the B_js in Σ, something in *S*'s conscious life in the interval leading up to *t* must be answerable to the belief ascription B_j.

We will say that some event *e* answers to the belief ascription B_j if (i) B_j is an ascription to *S* of the belief that *p* and (ii) *e*'s occurrence makes B_j true.

Intuitively, the idea behind EPA is that, to be correct, a mentalistic explanation must cite at least one belief that is activated, or consciously held, in the interval leading up to the behavioral explanandum.

EPA might be defended on the grounds that it captures an important insight regarding the conditions under which a subject's having a given propositional attitude can figure among the causes of her behavior. Suppose you think that a mentalistic explanation is acceptable only if it captures the causes of *S*'s behavior.[1] Then, to a first approximation, you can motivate EPA as follows. Since mentalistic explanation is explanation in terms of belief and desire, it follows that if *S*'s behavior *H* at time *t* is susceptible to a mentalistic explanation, then the explanation must cite at least one belief of *S*'s which is such that *S*'s having of that belief is manifested in a *datable event* that figures in the causal etiology of *H*.[2] But the only thing that would appear capable of playing this dual role – it must be both a datable manifestation of *S*'s belief that *p*, but also a cause of *S*'s *H*-ing behavior – would appear to be a conscious thinking that *p*. I will return to the reasoning behind this contention below. For now I want to note simply that it can seem that EPA embodies a constraint imposed on mentalistic explanation by way of considerations of mental causation.

3. Whatever EPA's merits on the score of mental causation, however, there is reason to think that many apparently-acceptable mentalistic explanations violate EPA. In what follows I present two examples in

---

[1] This is not to require that the event is described in terms of the causally relevant features; that is a matter into which I will not enter here. On the notion of causal relevance, in connection with mental states described in terms of their content, see Segal and Sober (1990) and Braun (1991) and (1995).

[2] This is not to say that, if a mental state is to be counted among the causes of an action, it must be treated as an event; for various distinctions are made between types of causes, only some of which require that the cause be an event. For this point in connection with *mental* causation, see below.

which a proposed mentalistic explanation is intuitively acceptable, yet in violation of EPA.

(a) *A* is on a safari, sees a tiger approaching her, and runs off. We might explain that *A* ran off because: she believed that there was a tiger running towards her and she wanted to get out of its way; she believed that there was a vicious creature coming towards her and she wanted to avoid danger; she believed that there was something large and dangerous coming her way and she wanted to get out of its path. Some of these explanations might seem more plausible than others. What is the basis for this plausibility judgement?

A natural speculation is that something like EPA is designed to answer just this question. The idea here would be that explanations are correct only to the extent that they capture what is going on in the conscious life of the subject in the relevant interval leading up to the behavior to be explained. But is this speculation correct? I think not. It is very easy to imagine poor *A*, on seeing the tiger, going into a state of panic in which she does not have any discernible thoughts at all. What comes to her mind: 'Oh, no!' followed by a jumble of fleeting words and images.[3] In such a case should we reject all of the mentalistic explanations above on the grounds that none of them cite any conscious thought she actually had in the relevant interval leading up to her fleeing? I think not. Assuming that this remains a case susceptible to mentalistic explanation (more on which below), it would seem that the mere fact, that no discernible conscious thought went through her mind in the relevant interval leading up to the fleeing, does not by itself undermine the acceptability of any of the explanations proffered above. So we can ask again: how do we determine the correct explanation(s), if any?

---

[3] We will see below that this feature, whereby a subject does not have discernible thoughts going through her mind in the relevant interval leading up to the action, is not unique to cases involving panic; it is actually commonplace.

We can go some distance towards answering this question, and towards seeing why *A*'s behavior remains susceptible to mentalistic explanation despite the conscious jumble, by noting the rationalizing aspect of mentalistic explanation. No matter one's views regarding whether reasons are causes, mentalistic explanations must rationalize the behavior: they must show that the agent's behavior is rational given both the context and the beliefs and desires that (by the lights of the explanation) she had. Clearly, this constraint will effect a significant narrowing on the class of acceptable mentalistic explanations for a given behavior.

At the same time, this constraint may not single out a unique explanation; for example, this constraint may not select only one of the three explanations proffered above as uniquely acceptable. Are there then further constraints? As another rule of thumb, we might query the subject herself *ex post facto*: Why did she flee? And we might hold that the agent's own explanation is to be given the benefit of the doubt, with defeat contingent on (i) evidence against the explanatory self-ascriptions she makes or (ii) a better alternative explanation.[4] But there are two reasons why one should not rest comfortably with this suggestion as it stands. First, in the absence of a clear notion of 'better explanation', the suggestion does not have a clear content. In fact, if we leave open what counts as a 'better explanation', the suggestion is consistent with rejecting the speaker's self-assessment more often than not. Second and perhaps more importantly, there would seem to be some arbitrariness in the subject's *ex post facto* explanation. Suppose *A* explained: 'I fled because I wanted to avoid the oncoming tiger, fearing that it was dangerous.' We query: did you fear that it was dangerous (as opposed to hungry or angry or mean or ...)? Now in some cases *A* can answer

[4] This is my attempt to make room for the famous results of Nisbett and Wilson (1977). I am not endorsing their theory; here I aim only not to beg questions against it.

definitively: 'Yes, I feared it was dangerous (as opposed to ...). But do we really want to say that, if she cannot answer this definitively, then no explanation involving any one of the relevant attributions will be acceptable? I think not.[5]

The point can be reinforced by reminding ourselves once again of the rationalizing aspect mentalistic explanation. A given mentalistic explanation involving ascriptions Σ can succeed in rationalizing a piece of behavior so long as the ascriptions in Σ (i) are true and (ii) together make sense of the behavior. Admittedly, 'making sense of behavior' is not an entirely perspicuous notion. However, I would urge that this notion is clear enough to see the acceptability of the claim that we can make sense of a subject's behavior even if none of the ascriptions in Σ correspond to any conscious thinking on *S*'s part in the relevant interval leading up to the behavior. Let me move on to a second example to make this plain.

(b) Thirsty, *C* goes to the fridge and grabs the water pitcher. We explain that *C* went to the fridge because he was thirsty, believed that there would be water (with which to quench his thirst) in the fridge, believed that he could walk there in a matter of moments, and had no countervailing beliefs or desires. Need *C* have thought all, or even any, of these propositional contents in order for our explanation to count as

---

[5] There is a potential worry about realism here. I am claiming that *S* need not have come to consciousness the belief that *p* prior to *H*-ing, in order for the belief that *p* to figure in a correct explanation of *S*'s *H*-ing. So a critic of my proposal might wonder: from the point of view of trying to provide a correct mentalistic explanation, what constraints *are* there on belief-ascriptions? The worry is this: if the constraints do not determine a unique belief-ascription (or perhaps a unique set of several belief-ascriptions, since mentalistic explanation typically requires the ascription of more than one belief), then it seems that there might be many (non-overlapping) belief-ascriptions that could be used to provide a correct mentalistic explanation – in which case there is no 'reality' to the beliefs so ascribed. I will return to this worry below.

correct? I don't think so. Think about the ordinary case: you're thirsty and you go to the fridge. Have you decided that you want water (as opposed to beer or OJ or apple juice or ...)? Have you thought of the object of your desire under the non-specific guise of 'thirst-quenching liquid'? When we 'listen in' on the goings-on in *C*'s mind during the relevant interval leading up to the point where he reaches the refrigerator, all we find (to *C*'s great embarrassment) is following verbalized word salad: "Water ... fridge ... there ... aaahhh!..." In short, not a propositional content to be found. (Is *C* any different from the rest of us in this regard?) I am not saying, of course, that *C* does not believe e.g. that it would be nice to have a glass of water to drink, that there is water in the fridge, and so on. I am merely suggesting that the ascription to *C* of such beliefs, for the purpose of explaining *C*'s fridge-directed behavior, can be warranted even in the absence of *C*'s having explicitly and consciously framed to himself the propositions which the explanation represents him as believing.

I submit that both of these examples are examples in which there is an apparently acceptable mentalistic explanation that violates EPA. At the very least, then, these cases constitute a challenge to EPA.

4. One who wished to preserve the spirit of EPA in the face of the forgoing considerations might respond as follows. Agreed, mentalistic explanation *E* of *S*'s *H*-ing may be acceptable even if it contains a set of ascriptions Σ which is such that none of its belief ascriptions corresponds to any conscious thought *S* had in the relevant interval leading up to *S*'s *H*-ing. But if *E* is acceptable then Σ must contain some ascriptions that stand in some interesting relation to some thought or other that *S* had in the relevant interval leading up to her *H*-ing; otherwise we begin to wonder whether *S*'s *H*-ing admits of a *mentalistic* explanation at all. Perhaps the lesson is that we should accept accounts of the semantics of belief ascription on which the ascription '*S* believes that *p*' can be true even when 'that *p*' does not capture the content of any

belief *S* has (see e.g. Sheir (1996) and Bach (1997)). This would allow that a mentalistic explanation *E* can be true, even when none of the content-specifying portions of any of the ascriptions in Σ correspond to any propositional content consciously thought by *S* in the interval leading up to *S*'s *H*-ing.

Unfortunately, this will not do. In order to be plausible, any approach to the semantics of belief ascription must hold that, in order for '*S* believes that *p*' to be true, *S* must have a belief whose propositional content stands in a semantically interesting relation to the proposition that *p*. (This much is clear in Bach (1996) and Sheir (1997)). Suppose that, in a particular case, the ascription '*S* believes that *p*' is true, and that its truth-maker is the fact that *S* stands in the belief-relation to the proposition that *q*. The trouble is that, given any fact of the form '*S* believes that *q*' which serves as a truth-maker for an ascription of the form '*S* believes that *p*,' we can give examples, parallel to the ones in section 3, supporting the claim that (in the relevant interval leading up to her *H*-ing) *S* need not have consciously thought that *q*, in order for the ascription '*S* believes that *p*' to figure in a correct explanation of *S*'s *H*-ing. The case of thirsty *C* confirmed this: *C*'s mental life in the interval up to her going to the fridge consisted of verbalized word salad, not consciously-entertained propositions. So it would seem that if her behavior can be correctly explained by a mentalistic explanation (as I assume it can), no event involving a fully propositional content before *C*'s consciousness need have preceded *C*'s fridge-directed behavior, in order for there to be a correct mentalistic explanation of that behavior.

Of course, it is a truism that acceptable mentalistic explanations can include only true belief ascriptions. This comes to the following: a candidate mentalistic explanation *E* is acceptable only if the ascription '*S* believes that *p*' (figuring in *E*) was true in the relevant interval leading up to *t*. I am not denying this. Rather, in denying EPA, I am merely suggesting that *S* need not have consciously thought that *p* in the relevant interval up to *t*, in order for an ascription of the form '*S* believes

that *p*' to figure in a correct explanation of *S*'s *H*-ing. The cases I am describing are cases in which, though it is the case that at *t S* believes that *p*, neither this belief nor any other belief manifested itself in the form of a conscious thought in the relevant interval leading up to *S*'s *H*-ing.

5. At this point several other possible rejoinders suggest themselves.

The first is this. So far I have argued that, on the assumption that mentalistic explanation is causal explanation, and given a mentalistic explanation involving the ascription to *S* of the belief that *p* (where this ascription is supposed to pick out part of the cause of the behavior *H*), then there must be something that is both (i) a manifestation of the belief that *p* and (ii) a cause of *H*. And I have argued that the appeal to conscious thought will not in general succeed in satisfying (i) and (ii) in all such cases. So perhaps we ought to look for some other bodily event to fulfill the roles of (i) and (ii). It is here that we must take up the question, introduced in section 2, regarding whether the conscious thought that *p* can be the only thing that can play the dual role of (i) and (ii). Above I suggested that it is; but if I am incorrect about this, then one can modify EPA to insist, not on a conscious event answering to one of the B_is, but rather some bodily event answering to one of the B_is.

Consider for example the belief box account of belief. Suppose that *S* believes that *p* if and only if *S* has in her 'belief box' some mental token *M*, where *M* means (or has as its semantic content) that *p*. Presumably *S* can have *M* in her belief box even in intervals during which the semantic content that *p* fails to reach her consciousness. In line with this, the proposal might be made that we can now reformulate EPA, replacing EPA's insistence on 'conscious' states of *S* with a more non-committal description of 'mental' states of *S*, as follows:

EPA2 If subject *S*'s behavior *H* (occurring at time *t*) is to be correctly explained by a mentalistic explanation whose ascriptions Σ include B_1 ... B_n, then for at least one of the

> B_js something in *S*'s mental life in the interval leading up to *t* must be answerable to that ascription.

However, EPA2 has two potential problems.

The first, which I mention only in passing, is this. Unless there are grounds for assuming that *S* has in her belief box a mental symbol whose meaning relates in the appropriate way to one of the ascriptions in B_1 ... B_n, where the grounds in question are other than the need to explain *S*'s *H*-ing, this proposal will seem *ad hoc*. However, developing this criticism would take me too far afield (into such matters as how mental symbols are identified, and how they are ascribed content), and would require speculation regarding matters about which present is little known at present (see e.g. Goldberg (2002)), so I will move on to my second criticism of EPA2.

The second criticism is that in at least one respect, EPA2 is worse off than EPA. As I noted at the outset, for a great many semantic contents that *p*, we can ascribe the belief that *p* to a subject *S* who is asleep, thinking of other matters, etc. So if we accept that

S believes that *p* iff (*S* has *M* in her belief box and *M* means that *p*), then for any belief that can be ascribed to *S* in the absence of *S*'s conscious thinking of it, there will be a corresponding mental symbol in the belief box. This will be so for all times during which *S* can be ascribed the belief in question. But then a question arises: if the mental symbol (whose meaning is that *p*) is in the belief box all along, and if the subject's believing that *p* is part of the explanation of her *H*-ing at *t*, why did she not *H* prior to *t*? It seems that the only answer is this: she *H*-ed at *t* because at least one of the beliefs cited in the explanation was manifested in the interval prior to *t*, in a way that it was not manifested previously. To take an example: Although thirsty *C* has believed for some time that water is typically in his fridge, it wasn't until he found himself to be both thirsty and near the fridge that he moved towards the

fridge. In short, in order to explain why a subject acts when she does, in those cases in which the purported explanation cites some beliefs that are long-standing, we will need to cite some manifestation of one of the cited beliefs in the relevant interval leading up to the action. For all its difficulties, EPA has this over EPA2.

I just considered and rejected one possible move to make in defense of EPA against the would-be counterexamples above. The move consisted of modifying EPA to EPA2, and then using some account of belief (such as the belief box account) indicate how beliefs can 'manifest themselves' as a bodily event in a way other than that of a conscious thought. This move was seen to fail to be a way to preserve the core of EPA on the grounds that we need to cite an event in the relevant interval leading up to t if we are to explain why S acted when she did, given that at least some of the beliefs in the explanation are long-standing beliefs.

We can see the same point in connection with another proposal, to the effect that EPA should give way to EPA3:

> EPA3 If subject S's behavior H (occurring at time t) is to be correctly explained by a mentalistic explanation whose ascriptions Σ include B_1 ... B_n, then for at least one of the B_is something in S's conscious life *at some point* prior to t must be answerable to that ascription.

EPA3 too has problems similar to those facing EPA2. Suppose S exhibits behavior H at time t. Supposed an explanation E of S's H-ing is offered, where E includes the belief ascriptions B_1 ... B_n but no others. Do we really want to imply that if prior to t S has not consciously thought any of the thought contents ascribed in the explanation's belief ascriptions (or any other thought-content that would warrant those ascriptions), then E cannot explain S's H-ing? This suggestion seems wrong for a principled reason: mentalistic explanations rationalize a person's behavior, and it would appear that an explanation can succeed

in rationalizing a person's behavior even in the absence of the person's ever having had the conscious thoughts which this implication would require of her. This is borne out by the case of thirsty *C*. Like most people who have not reflected on such examples in the course of thinking about mental causation, *C* has never explicitly consciously thought any of the following propositions: water is typically found in the refrigerator; water quenches thirst; from his living room he (*C*) can walk to the refrigerator in seconds; etc. No doubt, if *C* were ever presented with sentences expressing such propositions, he would immediately assent to them; but he hasn't framed them for himself prior to his walking to the fridge. Since these facts do not appear to cast doubt on the claim that *C* went to the fridge because he believed that he would find water there etc., the conclusion is that even the weakened EPA3 appears too strong.

Perhaps what the proponent of EPA needs here is Audi's 1994 category of dispositions to believe. We can say that *S* has a disposition to believe that *p* at time *t* if (1) at no time prior to *t* has *S* had the belief that *p* but (2) at *t S* has a disposition which is such that, if e.g. she were to consider whether *p*, she would acquire the belief that *p*. The proposal is that we can reject the weakened EPA3 in favor of a related doctrine regarding *dispositions to believe*.

> DB If subject *S*'s behavior *H* (occurring at time *t*) is to be correctly explained by a mentalistic explanation whose ascriptions Σ include B_1 ... B_n, then at *t S* needs to have a disposition answerable to one or more of B_1 ... B_n.

Whatever virtues this proposal has, it has one very important drawback: the state one is in when one has a certain disposition, as opposed to the state one is in when one of one's dispositions is manifested, is not the sort of state that can be a cause. So no one who hopes to retain the causal relevance of propositional attitudes such as belief should try to do

so by appeal to dispositions to believe. (Audi (1994) himself is perfectly aware of this, pp. 425-426.) I should add that this goes for any other belief-type state, such as states of implicit belief or that of tacit belief, which are understood dispositionally.

I have been asking how a proponent of EPA might try to defend EPA by appeal to the causal relevance of propositional attitudes. However, I have suggested that the cost of doing so is to surrender the correctness of explanations that would otherwise seem perfectly correct. Can the proponent of EPA preserve the doctrine by suggesting that not all mentalistic explanation is causal explanation? Such a proponent can still insist that EPA will be true in those cases in which the mentalistic explanation is also a causal one. However, such a position seems *ad hoc*. Take a subject *S* whose action we find it natural to explain as due in part to *S*'s belief that *p*. Do we really want to say that there is a difference in type of explanation between the case in which *S* was consciously thinking that *p* prior to acting, and the case in which she was not (but where we still find it natural to explain her behavior in this way)? I think not. Take two subjects S_1 and S_2, both of whom go to the fridge, but only one of whom (say, S_1) consciously thinks 'I would like some water to drink' etc. But both retrieve and drink a glass of water. This prompts us to offer the same explanation for both: both S_1 and S_2 went to the fridge because they wanted water, believed that there was water in the fridge, etc. Assuming that such explanations are correct in both cases, do we really want to say that in the one case but not the other the explanation was causal? This seems *ad hoc* in the extreme. Surely the desire to preserve EPA is not so great as to convince us to make such a move?

6. There is one final objection that must be dealt with, before I can conclude my case against EPA. This is the objection from realism, which runs as follows. According to the thesis on offer, a mentalistic explanation *E* whose belief ascriptions are B_1 ... B_n can succeed in

explaining *S*'s *H*-ing, even if nothing in *S*'s conscious awareness in the relevant interval leading up to *S*'s *H*-ing answers to any of B_1 ... B_n. But then a question arises: from the point of view of trying to provide a correct mentalistic explanation, what constraints *are* there on belief-ascriptions? The worry is this: if the constraints do not determine a unique belief-ascription (or perhaps a unique bundle of belief-ascriptions which together constitute the belief component of the uniquely correct mentalistic explanation), then it seems that there might be many (distinct) belief-ascription bundles, each one of which could be used to provide a correct mentalistic explanation – in which case there would appear to be no 'reality' to the beliefs so ascribed.

In response, we should take care to distinguish indeterminacy at the level of mentalistic explanation from indeterminacy at the level of belief. In particular, the former does not entail the latter. For even if the facts of a given situation are such that they are indeterminate between various mentalistic explanations E_1 ... E_p, in the sense that those facts do not determine a unique E_i as the correct one, still it may be perfectly determinate which belief ascriptions are true of the subject *S*. The only question is which true belief ascriptions are those relevant to explaining *S*'s behavior! This suggests that even those who insist that there are determinate facts about what a subject believes can embrace the thesis that in given cases it may be indeterminate which mentalistic explanation is correct.

It is perhaps worth adding here that we have methods of delimiting the class of acceptable mentalistic explanations. The methods in question typically operate *ex post facto*: we ask the person why she acted as she did, and we provisionally accept her self-directed mentalistic explanations. We do so provisionally, since it may come to pass on a particular occasion that we have what we consider to be strong evidence for rejecting her self-directed mentalistic explanation. ("*S* can't believe what, in her explanation of her own actions, she claims to believe"; or "*S* does believe what in her explanation she represents herself as believing,

but something other than the beliefs she cites explains her behavior"; etc.). What is more, we allow for some indeterminacy of explanation without this undermining our confidence in the reality of belief.

This last point can be illustrated with reference to thirsty *C*. When asked why she went to the fridge, she told us that she walked to the fridge because she was thirsty and thought water would be in it, but she might equally well have told us that she was thirsty and thought (i) that water would be on the top shelf of the fridge; (ii) that the water pitcher in the fridge would be full; (iii) that there would be water in the fridge; etc. At this level of fine-grainedness, worries about realism are ill-placed. True, there were a variety of distinct mentalistic explanations she might have given us (corresponding to the variety of distinct belief self-ascriptions she might have made); but even if none of the propositional contents self-ascribed in these explanations were consciously thought by her prior to walking to the fridge, this would not invalidate her explanation, nor would it make us worry about the 'reality' of the beliefs so ascribed. In fact, it is precisely *because* she believes all of these things – i.e., it is because the facts regarding what she believes are assumed to be perfectly determinate – that we get the explanatory indeterminacy in the first place! I conclude, then, that the repudiation of EPA adds no worries on the score of the reality of the beliefs figuring in such explanations. And so concludes my case against EPA: none of the modifications proposed in an attempt to salvage something of EPA – EPA2, EPA3, and DB – is without difficulty.

7. In this paper I have argued that there are cases in which there is a conflict between (on the one hand) our intuitive criteria for offering mentalistic explanations in particular cases and (on the other) the principle that any acceptable mentalistic explanation must cite the mental causes of the subject's behavioral effects. My argument can be seen as having the form of a *reductio:*

(1) To be acceptable, mentalistic explanations must cite mental causes [assumption made for the purpose of the *reductio*].

(2) In cases in which mentalistic explanations involve belief ascriptions, the acceptability of such explanations requires that there be something that plays the dual role of (i) a manifestation of one of the beliefs and (ii) a cause of behavior [from (1)].

(3) The only candidate to play the dual role of (i) and (ii) is conscious thought [argued for in section 5].

(4) But there are acceptable mentalistic explanations involving belief ascriptions in which there is no conscious thought that is answerable to any of the belief ascriptions in the explanation [argued for in 3]. So:

(5) Assumption (1) must be rejected.

In reaction, our options are essentially two. One might try to show that something is amiss in the argument given; or one can accept the argument and use it as an occasion to make what I see as the forced choice between embracing the intuitions behind (4) or the principle of (1).

REFERENCES

AUDI, R. (1994). "Dispositional Beliefs and Dispositions to Believe", *Nous* 28:4, pp. 419-434.

BACH, K. (1997). "Do Belief Reports Report Beliefs?", *Pacific Philosophical Quarterly* 78, pp. 215-241.

BRAUN. D. (1991). "Content, Causation, and Cognitive Science", *Australasian Journal of Philosophy* 69:4, pp. 375-389.

———. (1995). "Causally Relevant Properties", *Philosophical Perspectives* 9, pp. 447-475.

GOLDBERG, S. (2002). "Belief and Its Linguistic Expression", *Philosophical Psychology* 15:1, pp. 65-76.

NISBETT, R.E. & WILSON, T.D. (1977). "Telling more than we can know: Verbal reports on mental processes", *Psychological Review* 84, pp. 231-259.

SEGAL, G. & SOBER, E. (1990). "The Causal Efficacy of Content", *Philosophical Studies* 63, pp. 1-30.

SHIER, D. (1996). "Direct Reference for the Narrow Minded", *Pacific Philosophical Quarterly* 77, pp. 225-248.