Against fragmentation

Aaron Norby

A number of philosophers have suggested that we should adopt theories of belief on which individuals' sets of beliefs are 'compartmentalized' or 'fragmented' in ways that reflect their cognitive limitations (Davidson, 2004; Egan, 2008; Greco, forthcoming; Lewis, 1982; Stalnaker, 1984).

The advantage of building allowances for our cognitive limitations into our theory of belief is that we can hope to end up with a picture that is more descriptively accurate and more explanatory than we otherwise would. Nonetheless, I want to argue that invoking fragmented belief states achieves neither of these goals. My claim is that, in the contexts in which they are commonly invoked, talk of fragments does not designate anything psychologically real. Although fragmentation talk might seem explanatorily useful, there is reason to think that in reality it is merely a figurative re-description of the phenomena we want to explain and that any explanations it seems to offer are at best trivial.

1 Magellan

Imagine that a man called Magellan has the following three beliefs about the layout of the town in which he lives. First, he believes that Apple Street runs north-south. Second, he believes that Banana Street runs east-west. Third, he believes that Apple Street and Banana Street are parallel to each other. He hasn't noticed the inconsistency though, and this is because, while he sometimes reasons and behaves in accord with the first two beliefs, and sometimes in accord with the third, he has never as a matter of fact ended up bringing to mind all of them at once.[1] Sometimes, for instance, hell be giving

---

[1] A caveat. Im going to talk about the relation that belief has to reasoning *and* behavior, and often, for convenience, I'll mention only one or the other. Nothing in what follows, however, turns on whether it is only behavior, or only reasoning, or both, that belief is ultimately tied to.

directions, and the directions he gives would be correct given the truth of his first two beliefs but not given the third, and sometimes the directions he gives would be correct given the truth of the third of his beliefs but not the first and second. Since the situations in which the first two are manifested do not overlap with those in which the third is, there are situations in which, it seems, Magellan's disposition is to act in accord with the first two beliefs and fail to be so disposed with respect to the third; and there are situations in which the reverse is the case.[2]

Situations like this are fairly common and ought to be familiar, but they complicate the following question: what is the connection between what one believes and how one reasons and behaves? The standard view is that what one believes is connected to how one reasons and behaves, or to how one is disposed to reason and behave, by one's reasoning and behaving somehow *as if* what one believes were true, or in ways that would be successful or that would satisfy one's desires in worlds in which what one believes is true. I'll refer to this as the *classical view*. It's simple, intuitive, and if correct allows us to predict and explain behavior in terms of ordinary notions like belief and desire.

But do cases like that of Magellan force us to rethink the classical conception of the connection between belief and behavior? After all, when Magellan is giving directions that would be correct were his first two beliefs true, he's giving directions that would be incorrect were his third belief true—and vice-versa. Assuming that he wants to give correct directions in each case, it follows that when he's behaving in ways that would satisfy his desires were his first two beliefs true, he's *failing* to behave in ways that would satisfy his desires were his third belief true. And when he's behaving in ways that would satisfy his desires were his third belief true, he's failing to behave in ways that would satisfy his desires were his first two beliefs true. Moreover, there seems to be no way for him to behave such that behaving in that way

---

[2] This example is derived, with some minor modifications, from Lewis (1982); it is further discussed by Egan (2008) and by Elga & Rayo (2012) in reference to present issues. See those, as well as Stalnaker (1991), for examples of underutilization of belief not involving inconsistency.

would satisfy his desires were *all* of his beliefs true, since his beliefs do not all together form a coherent set. What we might say about him is that Magellan is prone to *underutilizing* some of his beliefs: he believes certain things that, at times, he fails to utilize (i.e., to guide himself by) in his reasoning and behavior. And that possibility doesn't seem to gel with the classical view of things, on which one is said to simply behave, or be disposed to behave, as if one's beliefs were true.

Since such cases are fairly common, they pose a problem for the classical view of belief: we can't capture them simply by stating that agents will behave as if their beliefs were true.

2   Fragmentation

Fragmentationalism attempts to account for cases like Magellan by thinking of some subjects' total set of beliefs as 'divided up' into fragments or compartments, where on any particular occasion only some of these fragments will be available for utilization in reasoning and behavior. We can think of those beliefs, or those belief-fragments, as the ones that are 'activated', with the others inactive and so unavailable for guiding reasoning and behavior. One would be disposed to guide one's reasoning/behavior by a given belief only when that given belief's fragment is activated. In this way one could believe something at one time even if one weren't, at that time, disposed to behave like it (Egan, 2008; Lewis, 1982; Stalnaker, 1984).[3]

Thus, for example, we could say that Magellan's belief that Apple Street and Banana Street are parallel is in a different doxastic fragment from his beliefs that Apple Street runs north-south and that Banana Street runs east-west. The former fragment is activated in different contexts than is the latter, and so he ends up failing to put the pieces together and remove the

---

[3] Daniel Greco (forthcoming) points out that one might take belief to be fragmented in a different sense, in which belief has many different sub-species that all contribute differently to behavior. Amongst these might be alief (Gendler, 2008) or in-between belief (Schwitzgebel, 2010). I set these cases aside, in part because of space constraints and in part because I think these are notions best used to think about different kinds of cases than those we're focused on here.

inconsistency (though in the right circumstances, perhaps, in which both fragments are simultaneously activated, he might).

The first thing to consider is what these fragments, psychologically speaking, might be. In the works I've cited there is little indication. And although empirical psychology provides a number of ways in which we might divide beliefs up by psychological type, it is unlikely that any of them are suitable for fragmentationalism's purposes. The general reason for doubt is that amongst the factors that determine the underutilization of belief—that is, that give rise to Magellan-like cases—are variables that would appear to allow beliefs *within* psychological divisions to manifest 'fragmentational' phenomena. If that's right, then doxastic fragments would have to cut across rather than conform to divisions drawn by empirical psychology.

We can see the issue most clearly if we allow ourselves, at least for the time being, to think of beliefs as stored memories. Included amongst the variables that influence how readily a particular memory will be retrieved on a particular occasion are such factors as how recently it was last retrieved, the frequency with which it's been rehearsed, and whether there are any salient environmental cues that might facilitate its retrieval (Anderson, 2007). Whether an appropriate, salient cue or *prime* is present can often determine whether or not a particular memory will be retrieved in response to a task (Förster & Liberman, 2007, for review). For example, you might be able to recall the name of Oregon's capital if you're presented with the first few letters ("Sal_"), but unable to recall it if given no cue at all.

Factors like the presence of a relevant cue, however, are ones that appear to be able to vary across items *within* psychological divisions, and so make it likely that beliefs within single divisions will give rise to Magellan-like phenomena. In general, the set of cues that elicit one memory will be distinct from those that elicit a second memory, even though there may be some cues that elicit both, and, critically, *even though both memories may be stored in the same memory subsystem* (Eichenbaum & Cohen, 2004).[4]

---

[4] This is also why networks of *associated beliefs*, as found in spreading activation models Collins & Loftus (1975), cannot be identified with fragments. For just as two memories can be elicited by distinct but partially overlapping sets of cues, two beliefs in an

For example, it is widely accepted that semantic memories for things in the *animate* domain—e.g., memories about the identities of different kinds of animals—are subserved by a system that is functionally and anatomically distinct from that which subserves semantic memories for things in the *inanimate* domain—e.g., memories about the identities of different kinds of tools (Caramazza & Shelton, 1998).[5] But one animate-memory can be cued, and so retrieved and then utilized, by stimuli that don't cue other animate-memories. For example, seeing a picture of an octopus might facilitate the retrieval of information about octopi while allowing one's ideas about spiders to remain dormant. If we imagine someone whose beliefs about octopi are inconsistent with her beliefs about spiders,[6] we can see how the cued elicitation of her octopi-memories can lead, Magellan-like, to a failure to behave as if some of her beliefs about spiders are true. This is possible even though these beliefs may all be stored in the same memory subsystem. Thus, fragmentationalism would have to count some of this agent's octopi-beliefs as fragmented from some of her spider beliefs, and this boundary would cut across a psychological division.

This sort of cue-dependent retrieval shows how beliefs on the same side of a psychological division can be utilized under distinct conditions, making it difficult to identify fragments with such divisions. Since this is a feature shared by a variety of memory subsystems (Eichenbaum & Cohen, 2004), it suggests that fragmentationalism may not find motivation from the perspective of psychological science.

For now, however, lets set these considerations aside. It may very well be that the best and most charitable way to understand the fragmentation-

---

associative network can be closely associated along one dimension but not along another. In general, excitatory and inhibitory links in associative networks can be transitory, tied to particular contexts, and asymmetric (Förster & Liberman, 2007), making them ill-suited to fragmentationalism's purposes.

[5] Thanks to an anonymous referee for raising the importance of discussing this sort of psychological division. It's worth noting that the animate and inanimate are not the only domains that are subserved by dissociable subsystems. See Caramazza & Mahon (2003) for review.

[6] E.g., we can imagine someone who believes that octopi have eight legs, believes that octopi aren't arachnids, and also believes that any creature with eight legs is an arachnid.

alist's suggestion is that the notion of a doxastic fragment is a highly useful way to model subjects' belief states, however those fragments are psychologically realized. Thus, although there might not really be 'fragments' between beliefs, all it takes for two beliefs to be in different fragments is for them to be activated and utilized under different circumstances, as we see in Magellan's case. In what follows, this is what I will take the fragmentationlist thesis to be in the first instance.

What I want to focus on is the fact that the predictive power of the fragmentational theory would come from our ability to say *when* a given doxastic fragment or belief will be activated to guide behavior. We can't say how, in all circumstances, Magellan is going to be disposed to behave so far as the arrangement of Apple Street and Banana Street are concerned; rather, what we know is that sometimes he'll be disposed to act one way, and other times in a different, incompatible, way. We can predict his behavior only by knowing under what circumstances hell be disposed in the one or the other way, i.e., in what circumstances each of the relevant fragments will be activated. This is important, I think, because one of the things we want from a theory of belief is an understanding of how the beliefs that a person has relate to her thought and behavior. This predictive limitation is also tied, as we will see, to a bigger worry, which is that fragmentational models may not be able to say anything substantive about belief.

One might be tempted to suggest that the predictive limitation here is no problem because the fragmentationalist has recourse to a perfectly good generalization linking belief to behavior. Specifically, she can incorporate the notion of an 'active' belief or belief-fragment into the generalization in the following sort of way, which we can call the fragmented belief law:

> **(Fragmented Belief Law)** If you believe that p, then, if you are in a circumstance in which the fragment containing that belief is active, you will be disposed to reason/behave as if p is true.

The fragmentationalist can hold that specifying someone's overall doxastic state requires specifying not only what they believe and how those beliefs

are divided into fragments, but also the circumstances in which each fragment is activated. The sort of law described incorporates this consideration. The fragmented belief law, it might be suggested, takes on not much more baggage than the standard classical approach. For with classical sorts of generalizations already, in order to predict someone's behavior we needed to know what that person believes; what the fragmentational law adds, it seems, is only that we must also know the circumstances in which particular fragments are activated.

The sort of fragmentational 'law' suggested, however, which reserves a place for specifying the circumstances of the activation of doxastic fragments, is not able to do the work that it is needed to do. Indeed, it serves to highlight just the difficulty of the problem that we face. For there simply is no notion of what the *activation* of a belief or belief-fragment is other than that it consists in that belief (or the beliefs in the relevant fragment) guiding one's behavior.[7] And, as we've said, what it is for a belief or set of beliefs to guide one's behavior includes that one is disposed to reason/behave as if they were true. What this means, though, is that the fragmented belief law amounts to this triviality: if you believe that p, then, if you are in a circumstance in which you are disposed to reason/behave as if the beliefs in the fragment containing p are true, you will be disposed (in those circumstances) to reason/behave as if p is true. That is, it says that when you are disposed to reason/behave as if p is true, you are disposed to reason/behave as if p is true.

It says this and nothing more, that is, unless we include in the law the

---

[7] This is not to impugn the use of the notion of 'activation' in cognitive science generally. First, cognitive scientific work on activation aims, in no small part, at discovering the variables that determine when certain processes or states will become activated (e.g., Bargh & Williams 2006). Thus, even on the understanding of 'activation' glossed above, the generalizations produced by this work are non-trivial. Second, insofar as 'activation' in cognitive science picks out a particular psychological process—or, say, some neural process—it does not seem that this is the sense of 'activation' that fragmentational theories of belief should invoke. For this sense of would require that there be the right sort of connection between activation and belief, which there may not be. For example, in the literature on knowledge activation, it is common to point out that when a piece of knowledge is activated it nevertheless may not be used in subsequent behavior (Förster & Liberman, 2007; Higgins, 1996), which would mean that the generalization offered in the text (linking the activation of belief to the utilization of belief) would be false. Thanks to an anonymous referee for raising this problem.

actual, empirically discovered, conditions under which particular beliefs and belief-fragments are activated. But then there would be no need to invoke doxastic fragments, since these generalizations tell us about the variables that determine when our various beliefs are and are not likely to be utilized, and will be able to do so without appeal to the psychologically dubious notion of a doxastic fragment.

At this point, it might be admitted by the fragmentationalist that, after all, if you want generalizations with very robust, detailed predictive power, then you will in the end need to fill in the generalization schema as outlined above. But, the thought goes, this does not mean that there is not still a more general and important connection to be drawn between belief and behavior, one that captures the crucial features of that relationship without having to delve into chauvinistic psychology.

As weve said, fragmentational theories cannot on their own say *when* you'll behave as if p, should you believe it. But they can say this: if you believe that p, then you will in some circumstances (those in which the belief or its fragment is activated) behave as if p. Let's call this the *modest generalization.*

> **(Modest Generalization)** If you believe that p, then you will in some circumstances (those in which the belief or its fragment is activated) behave as if p.

A fragmentational theory that is committed to producing a truly general theory can say nothing stronger,[8] but perhaps, one might argue, the modest generalization captures something important about belief nonetheless.

This line of defense, however, will drive fragmentationalism into triviality. For the modest generalization cannot do any real work. The problem facing it is that it is not strong enough to distinguish the connection that *belief* has to reasoning and behavior from the connection that other, wildly different, states have. (As I'll point out below, this is not a weakness of the classical conception of the relation.) If one thinks, as many (including myself) do, that it is part of the job of a theory of belief to pick out *belief's* role

---

[8] Egan (2008), for one, is explicit about this fact.

in reasoning/behavior, then this would be enough of a problem to justify a rejection of this version of the view.

Ill illustrate the difficulty with an example, from which the basic recipe to generate more should be obvious.[9]

Stalnaker (1984) introduces a propositional attitude he calls *acceptance*, a category of attitude broader than that of belief. On a theory that acknowledges acceptance, every belief is an acceptance, but not every acceptance is a belief; there is mere acceptance. For us the important difference between acceptance and belief is that you can accept things in a circumscribed context for a particular purpose without believing them to be true, and also while believing them to be false. This is especially clear in conversational contexts, in which a conversational participant might accept that p, and thus go on as though p were true, solely for the sake of keeping the conversation moving along. To use an example very close to one of Stalnaker's, suppose that you say to me, "The man with the gold tooth stole my hat." I might know that the man's tooth isn't gold, but for the sake of keeping the conversation moving *accept* that it is, and go on talking as if you'd said nothing false (Stalnaker, 2002, pp.718–9). This sort of thing should be familiar to most of us, and, importantly, we can readily recognize that mere acceptance is distinct from belief.

But, looking at our example, we can see that if you merely accept that the man's tooth is gold, you will in *some* circumstances—namely, the circumstance of the immediate conversational context—behave as though the man's tooth is gold. That is, because you merely accept, but don't believe, that the man's tooth is gold, you're disposed in some circumstances to behave as if the man's tooth is gold. So, we have another state—acceptance in a context—that is distinct from belief and that is, intuitively, behaviorally distinct from belief, but that meets the condition described by the modest generalization.

At the very least, what this example shows is that on the sort of frag-

---

[9] In addition to the following example, one could appeal to cases of assuming something for the sake of argument, as well as to cases of involved pretense and imagination, like those described in Velleman (2000).

mentational theory we're looking at the *cognitive role* of belief cannot be distinguished from that of mere acceptance; and that in itself is a critical problem for a theory of belief. This is so not least of all because many philosophers think that it is the role that a belief plays in thought and behavior that makes it a belief. The point of these examples is not to show definitively that on the sort of fragmentational theory we're looking at there is no way that one might be able to distinguish belief from, say, mere acceptance or, alternatively, from pretense. Perhaps that distinction could be made with further conditions on belief, such as how the state must be caused. But such a move still seems to be missing something important, for on the face of it there really is a behavioral difference—or, more broadly, a difference in cognitive role—between belief and mere acceptance. Moreover, it—s perhaps not unreasonable to think that a large part of the reason that we're interested in belief is that it plays an important characteristic role in cognition and behavior. A theory that is unable to say what this role is—or to distinguish it from that of other states—appears not to be providing a large part of what we want out of our theory of belief. If the example of mere acceptance is successful, what it shows is that the modest generalization *cannot* be what describes the connection that belief has to thought and behavior, for the reason that belief is supposed to have a privileged, unique connection to them.

Note that this is not a problem that the classical approach faces. For on the standard way of thinking of the relation between belief and behavior, you will behave as though what you believe is true across a wide range of circumstances, and so the fact that you fail to do so in normal sorts of situations would discount you from genuinely believing, which is enough to distinguish the behavioral role of belief from that of mere acceptance. Of course, that doesn't change the fact that it seems to have nothing to say about Magellan's case.

The fragmentational theory won't work, then, and it won't work for the simple reason that the theory makes it too easy for a mental state to have the connection to thought and behavior that belief is supposed to have. That's setting the bar too low. It is far from clear that the notion of a 'fragment'

serves any purpose except as an aid to imagination, a way to more simply re-state the phenomenon that we're trying to understand: that an agent can manifest a belief in one set of circumstances and not another.[10]

References

Anderson, John R. 2007. *How can the human mind occur in the physical universe?* Oxford University Press.

Bargh, John A., & Williams, Erin L. 2006. The automaticity of social life. *Current Directions in Psychological Science*, **15**(1), 1–4.

Caramazza, Alfonso, & Mahon, Bradford Z. 2003. The organization of conceptual knowledge: The evidence from category-specific semantic deficits. *Trends in Cognitive Sciences*, **7**(8), 354–61.

Caramazza, Alfonso, & Shelton, Jennifer R. 1998. Domain-specific knowledge systems in the brain: The animate-inanimate distinction. *Journal of Cognitive Neuroscience*, **10**(1), 1–34.

Collins, Allan M., & Loftus, Elizabeth F. 1975. A spreading activation theory of semantic processing. *Psychological Review*, **82**(6), 407–28.

Davidson, Donald. 2004. Paradoxes of irrationality. *In: Problems of rationality.* Oxford University Press Clarendon.

Egan, Andy. 2008. Seeing and believing: Perception, belief formation, and the divided mind. *Philosophical Studies*, **140**(1), 47–63.

Eichenbaum, Howard, & Cohen, Neal J. 2004. *From conditioning to conscious recollection.* Oxford University Press.

Elga, Adam, & Rayo, Agustín. 2012. Belief fragmentation. *Unpublished draft of November 20, 2012.*

Förster, Jens, & Liberman, Nira. 2007. Knowledge activation. *Pages 201–31 of:* Kruglanski, Arie W., & Higgins, E. Tory (eds), *Social psychology: Handbook of basic principles, 2nd ed.* New York: Guilford Press.

Gendler, Tamar Szabó. 2008. Alief in action (and reaction). *Mind and Language*, **23**(5), 552–85.

Greco, Daniel. forthcoming. Iteration and fragmentation. *Philosophy and Phenomenological Research.*

Higgins, E. Tory. 1996. Knowledge activation: Accessibility, applicability, and salience. *In:* Higgins, E. Tory, & Kruglanski, Arie W. (eds), *Social psychology: Handbook of basic principles.* Guilford Press.

Lewis, David. 1982. Logic for equivocators. *Noûs*, **16**(3), 431–441.

Schwitzgebel, Eric. 2010. Acting contrary to our professed beliefs. *Pacific Philosophical Quarterly*, **91**, 531–53.

Stalnaker, Robert. 1984. *Inquiry*. Bradford MIT Press.

Stalnaker, Robert. 1991. The problem of logical omniscience, I. *Synthese*, **89**, 425–40.

Stalnaker, Robert. 2002. Common ground. *Linguistics and Philosophy*, **25**(5), 701–21.

Velleman, J. David. 2000. The aim of belief. *In: The Possibility of Practical Reason*. Oxford University Press.