

# Imagination, Delusion, and Self-Deception<sup>1</sup>

Andy Egan

Australian National University/University of Michigan

## *Introduction*

Subjects with delusions profess to believe some extremely peculiar things. Patients with Capgras delusion sincerely assert that, for example, their spouses have been replaced by impostors. Patients with Cotard's delusion sincerely assert that they are dead. Many philosophers and psychologists are hesitant to say that delusional subjects genuinely *believe* the contents of their delusions.<sup>2</sup> One way to reinterpret delusional subjects is to say that we've misidentified the content of the problematic belief. So for example, rather than believing that his wife is has been replaced by an impostor, we might say that the victim of Capgras delusion believes that it is, in some respects, *as if* his wife has been replaced by an impostor. Another is to say that we've misidentified the attitude that the delusional subject bears to the content of their delusion. So for example, Gregory Currie and co-authors have suggested that rather than *believing* that his wife has been replaced by an impostor, we should say that the victim of Capgras delusion merely *imagines* that his wife has been replaced by an impostor.<sup>3</sup>

In this paper, I will explore a strategy of this second sort. Saying that the attitude that delusional subjects bear to the contents of their delusions is imagination, though, faces its own problems. I want to suggest that, instead, we ought to say that delusional

---

<sup>1</sup> Thanks to Tim Bayne, Martin Davies, Anne Aimola Davies, Daniel Stoljar, Kendall Walton, Ted Warfield, and Tyler Doggett, and to audiences at the Australian National University, Monash University, and the Macquarie University Workshop on Delusion and Self-Deception.

<sup>2</sup> See, for example, Sass (1994), Berrios (1991), Campbell (2001), and Currie and co-authors in the papers cited below.

<sup>3</sup> See Currie and Ravenscroft (2002), Currie (2002), and Currie and Jureidini (2001)

subjects don't straightforwardly believe the contents of their delusions, nor do they straightforwardly imagine them. Instead, they bear some intermediate attitude – we might call it 'bimagination' – with some of the distinctive features of believing, and some of the distinctive features of imagining.<sup>4</sup>

We might be concerned, though, that there is no room for such peculiar, unfamiliar, fish-nor-fowl propositional attitudes. I'll argue that such concern would be misplaced – there is no principled reason to deny that such attitudes are possible. Finally, I'll suggest that a similar sort of intermediate attitude – this time one that's in some respects belief-like, in others desire-like – can help us to provide an account of certain cases of self-deception.

### *1. Ground Clearing: Assumptions about Representation*

In what follows, I'll be concerned with characterizing (very partially) the attitude that delusional subjects bear to the contents of their delusions. I'm going to assume that the right way to go about this is to characterize the roles that particular, token mental representations play in the subject's cognitive economy. So I'll be assuming that some sort of minimal representational theory of mind is correct – that there is a medium of mental representation, and there are discrete representational items in the head. These representational items are operated on in various ways, and accessed by various systems, in order to regulate both our behavior and the maintenance of other representational items. Believing, desiring, imagining, and the bearing of propositional attitudes in

---

<sup>4</sup> This possibility is suggested in Davies and Coltheart (2002), and is, obviously, a development of the ideas in the Currie et al. papers cited above. In fact, this sort of proposal is consistent with much of what Currie and his co-authors say – perhaps what follows is best understood as a suggestion for how to implement Currie & co.'s proposals.

general is a matter of having a representational item with the right kind of content, which plays the right kind of role in one's cognitive economy.

This sort of view takes no stand on just what the representations are like. So in particular, they needn't be sentencelike. They might be more like maps, or models, or something else altogether. All the view in question says is that mental representation happens in some medium, and that the rich mental lives that we know and love are the products of representations encoded in this medium getting operated on in various ways by various cognitive systems.

On such a view, the way that we come by our beliefs, desires, etc. is by having representational items in our heads that play the right kind of role in our cognitive economies. (I'll have more to say about the details of the roles later, but the following rough sketches should convey enough of the idea for our current purposes.)

So, for example, I believe that there's beer in the fridge because I've got a mental representation that locates the beer in the fridge and that my behavior-planning systems use (or are disposed to use) as the right sort of guide – a guide to where the beer is, right now – when figuring out ways to go get the beer.

I desire that I'm in Tahiti because I've got a mental representation that locates me in Tahiti and that my behavior-planning systems use (or are disposed to use) as a different sort of guide – a guide to where I ought to *wind up*, at the end of the contemplated course of action, rather than to where I am right now.

If we think about behavior-planning systems as systems for figuring out ways to get from some start state (the way things are taken to actually be, right now) to some goal state (the way things are desired to be) we can characterize this difference in the roles of

belief-type and desire-type representations this way: My behavior-planning systems look at belief-type representations in order to determine *start* states, and they look at desire-type representations in order to determine *goal* states.

Obviously there's going to be a lot more to the roles than this. More on this later. For now, the key point is this: The significance of some representational item in our cognitive economy is determined by two things: (a) what it represents, and (b) how it's hooked up to different cognitive systems – what looks at it when, and to what end. So if the beer map had a different role, it would give me desires, not beliefs. If the Tahiti map had a different role, it would give me beliefs, not desires.

I think that such a view is extremely plausible, and quite likely unavoidable, but I won't argue for that – for our purposes, I'm just going to assume that it's right. In particular, I'm going to assume that delusional subjects have their *delusions* in virtue of having particular, token mental representations of the delusional content, which play a distinctive role in their cognitive economies.<sup>5</sup> Not everything that I say below will depend on this assumption, though some of it will. The central idea – that there's no principled reason to deny the possibility of attitudes intermediate between belief and imagination – does not, I think, depend on the assumptions about representational media. The details of just how we ought to understand such attitudes, however, does.

## 2. *The Difficulty of Pigeonholing Delusions*

---

<sup>5</sup> Notice that many of the reasons to be concerned about whether *every* belief, desire, etc. that P is the product of a token mental representational item with P as part of its content are not reasons to be concerned about whether *delusions*, in particular, are the products of such token representations. Many of the concerns about the general claim have to do with the possibility of phenomena like tacit beliefs whose contents the agent has never even considered, and which it's implausible to think are explicitly represented. But given the position that delusions occupy in the forefront of their victims' conscious mental lives, it seems as if, if *anything* is represented explicitly in a delusional subject's head, it's the content of their delusion.

So: what *is* the attitude that delusional subjects bear to the contents of their delusions? What attitude do victims of Capgras really bear to the proposition that their spouses, etc. have been replaced by impostors? What attitude do subjects with Fregoli delusion really bear to the proposition that they're being followed around by people who are know to them, but who are in disguise?<sup>6</sup>

The obvious first thing to say – that the relevant attitude is, of course, *belief* – is surprisingly problematic. Delusions, it turns out, display a lot of behavior that doesn't look terribly belief-like.<sup>7</sup>

First off, the relation between a subject's delusion and his or her evidence is interestingly different from the relation of our ordinary (non-delusional) beliefs to our evidence. In many cases, delusional beliefs seem to be formed without the sort of evidential basis that other sorts of beliefs are formed in response to, and they are often resistant to revision in the face of conflicting evidence.<sup>8</sup>

(Note that this isn't, or at least isn't clearly, universal. On some accounts of, for example, Capgras delusion, the subjects' unusual perceptual experiences *do* give them evidence for their delusions, and questions about whether they are responding appropriately to subsequent evidence are difficult to resolve conclusively.<sup>9</sup> Notice, however, that whatever we think about the accounts according to which nothing untoward is happening in the Capgras subject's treatment of his or her evidence, there are other delusions for which this sort of account is far less plausible. It is substantially

---

<sup>6</sup> Capgras and Fregoli are, of course, unusual delusions – they're importantly different from typical schizophrenic delusions, for example. So it may well be that the answer to this question differs for different delusions – perhaps schizophrenic patients bear a different attitude to the contents of their delusions than Capgras patients bear to the contents of theirs.

<sup>7</sup> Bayne and Pacherie (2005) and Currie and Ravenscroft (2002) provide excellent presentations of the troublemaking phenomena that follow.

<sup>8</sup> See for example Campbell (2001),

<sup>9</sup> See for example Maher (1974, 1999), Stone and Young (1997), and Davies et al. (2001).

harder to maintain that delusions like Cotard delusion, whose victims believe that they are dead, are supported by the subjects' perceptual evidence, or that they could withstand the subjects' subsequent evidence if there weren't something extremely peculiar going on.)

It might seem as if it's crucial to something's being a belief that it be formed on the basis of, and sensitive to revision in response to, evidence in the right kinds of ways. If we think this – if we think that a certain sort of evidence-responsiveness is essential to belief – then, in many cases, we'll be reluctant to say that delusional subjects genuinely believe the contents of their delusions. And so, we'll be uncomfortable with characterizing delusions as genuine beliefs.

Another respect in which delusions are puzzling, and which makes categorizing delusions as beliefs problematic, is that delusions, when compared to other beliefs, seem to have an importantly *circumscribed* role in subjects' cognitive economies. Monothematic delusions, such as Capgras, Fregoli, and Cotard delusions, evidence this circumscription particularly strongly.<sup>10</sup>

The first way in which the cognitive role of delusions is circumscribed relative to that of paradigmatic beliefs is *inferential*. Delusional subjects often do not draw the sorts of inferences that one might expect from someone who believed the content of their delusion. A subject with Capgras delusion, for example, who believes that his wife has been replaced by a duplicate, is likely *not* to adopt an overall world view according to which it makes sense that their spouse should have been replaced by an impostor. As

---

<sup>10</sup> Notice that this isn't just to say that monothematic delusions are monothematic. As Davies and Coltheart (2002) point out, monothematicity and circumscription are at least logically independent, even if they do tend to go together in actual cases.

Davies and Coltheart (2002) put it, “Capgras patients do not seem to incorporate the consequences of their belief into their general account of how the world works.”

Notice that this inferential circumscription is a matter of degree. It’s not as if delusional patients never draw any inferences from their delusions – it’s just that their delusions don’t play as rich an inferential role as we’d expect from a full-blooded *belief* with the same content.

Another respect in which the role of delusional beliefs is circumscribed is *behavioral*. Delusional subjects fail, in important ways, to act in ways that we would expect from someone who genuinely believed the things that they profess to believe.<sup>11</sup>

Behavioral circumscription, too, is a matter of degree. Subjects *do* take some of the actions that one might expect a belief with the delusional content to motivate, sometimes with tragic results.<sup>12</sup> The behavioral circumscription of delusions is not a matter of their being completely behaviorally idle, but only of their having an impoverished behavior-guiding role relative to what we would expect from a full-blooded belief with the same content.

Finally, the delusional belief’s role in subjects’ emotional lives seems to be circumscribed, as well. Subjects often do not seem to experience the sorts of affective responses that we would expect from someone who believed that, for example, their spouse had been replaced by an impostor.<sup>13</sup>

So categorizing delusions as straightforward cases of belief faces some pretty serious obstacles. The role that delusions play in their subjects’ cognitive economies differs pretty dramatically from the role that we’d expect beliefs to play.

---

<sup>11</sup> See for example Stone and Young (1997), Sass (1994), Bleuler (1924).

<sup>12</sup> See Stone and Young (1997) for a summary of cases.

<sup>13</sup> Again, see Stone and Young (1997).

Saying, with Currie and his co-authors, that the relevant attitude is not *belief*, but *imagining*, looks potentially promising.

Imagining displays the right kind of evidence-independence. In order to imagine that P, I needn't have any evidence that P. Further, getting evidence that not-P, or noticing that I already had such evidence, needn't interfere at all with my continuing to imagine that P. For example, my imagining that I am a dashing pirate captain, in the course of a daydream during a slow spot in a talk or a faculty meeting, is not a response to any *evidence* that I am a dashing pirate captain. And I will not cease to imagine that I am a pirate captain when confronted with good evidence that I am, in fact, a philosopher.

Imagining displays the right kind of inferential circumscription. When I imagine that P, I typically do not come to *believe* things that follow from P, or from P together with things that I believe. (Even things that follow very obviously and straightforwardly.) For example, if I imagine that my dog is a bear, and believe that bears have only bears as littermates, I won't infer – and thereby come to believe – that my dog's littermates were bears.

Imagining also displays the right kind of behavioral circumscription. When I imagine that P, I needn't act in ways that would be appropriate if I believed that P. It's perfectly possible for my imagining that P to have no behavioral consequences at all, or none that are like the ones that believing that P would have. For example, I can imagine that I am a pirate, in standard piratical circumstances, without ever saying "arrrr", swinging from a rope, or firing a cannon. I am very likely to *imagine* myself doing these things, but I am unlikely to actually perform any of the piratical actions, or even the play-acting analogues of such actions. (Particularly if the imagining is taking place during a



seminar, or a faculty meeting.) And very many people routinely imagine, in the course of watching movies, that they are being pursued by bears or pirates, shot at, stalked by werewolves, etc., without running away, taking cover, or calling the police.

Finally, the affective impact of imagining is different from the affective impact of belief in the right kind of way. When I imagine that P, I don't typically have the full range of affective responses that believing that P would bring about. For example, we may react to the imaginings of tragic events that accompany watching a disaster movie with excitement rather than sorrow, or with only a faint shadow of the sadness that would accompany a belief that the depicted events were actually occurring. Though it is an interesting feature of fiction and imagination that we often do have quite rich affective responses to them, we certainly do not (standardly) have the *same* emotional responses to imagining and believing that P. When we have an affective response to imagining that P, it is typically not as strong, and sometimes not of the same kind, as the response that we would have to believing that P.<sup>14</sup>

All of this makes an account of delusions as imaginings, rather than beliefs, look quite attractive. Imagination is evidence-independent in the right kind of way, and its role in the imaginer's cognitive economy is circumscribed, relative to that of paradigmatic cases of belief, in the right kind of way. The trouble is that categorizing delusions as straightforward cases of imagining predicts *too much* circumscription, and too much evidence-independence.

First, while delusions don't display the sort of responsiveness to evidence that we'd expect from a belief, they display much more responsiveness to evidence than we'd

---

<sup>14</sup> See for example Currie (1997), Currie and Ravenscroft (2002) and Walton (1990, 1997) for more on this extremely interesting subject.

expect from a paradigmatic imagining. Delusional subjects very often show a keen awareness of, and discomfort with, the tension between their delusions and the other things they believe.<sup>15</sup> Contrast this with my imagining that I am a pirate captain. My imagining that I am a pirate captain and my belief that I am a junior philosopher coexist quite happily with one another – the presence of both produces no tension, discomfort, or sense that one of the two attitudes is in any way inappropriate.

We also do not find the same degree of inferential circumscription in delusions as we find in paradigmatic imaginings. Delusional subjects do report a felt tension between their delusions and many of their ordinary, non-delusional beliefs, and seem to feel a pressure to revise either their delusions or their other beliefs in the face of this tension. Subjects do, in fact, make some revisions in both directions under pressure. And delusions are often elaborated to some extent – the sort of inferential circumscription we find is not the *total* lack of inference to belief that we would expect from a paradigmatic imagining, but only a *much reduced* inferential efficacy, compared to paradigmatic beliefs.

Delusions are also not as behaviorally circumscribed as paradigmatic imaginings. Delusional subjects' verbal behavior, certainly, is routinely and systematically guided by their delusions. While imaginers' verbal behavior might sometimes, within the context of a make-believe game, be guided by their imaginings rather than their beliefs, this is certainly not the usual situation – in ordinary contexts, where sincere assertion is called for, imaginers verbal behavior is guided by what they *believe*, not what they imagine.<sup>16</sup>

---

<sup>15</sup> See for example Stone and Young (1997), Davies and Coltheart (2002), Davies et al. (2001), Young (2002).

<sup>16</sup> This point is due to Bayne and Pacherie (2005).

And as mentioned above, delusional subjects' non-verbal behavior, too, is often guided by their delusions in a way that imaginers' behavior is not guided by their imaginings. Sometimes this is very high-stakes behavior, as in the tragic cases in which a subject with Capgras delusion becomes violent – sometimes homicidally violent – toward the subject of their delusion.<sup>17 18</sup>

The situation with affective circumscription is better for the advocate of the delusion-as-imagining account. Certainly delusions *do* give rise to some affect, but so do (many) imaginings, and so it's not clear that there are any grounds for concern here.

Classifying delusions as straightforward, paradigmatic cases of belief is problematic because it predicts that delusions ought not to display the sorts of circumscription and evidence independence that they in fact display. Classifying them as straightforward, paradigmatic cases of imagination is problematic because it predicts that they should display *more* circumscription and evidence-independence than they in fact display.

---

<sup>17</sup> See Young (2002), Stone and Young (1997), Davies and Coltheart (2002), Davies et al. (2001) for some summaries of cases.

<sup>18</sup> But wait – the proposal wasn't just that delusional subjects imagine, rather than believe, the contents of their delusions. The proposal also said that the subjects mistake these imaginings for beliefs. How does this help? There are two ways that the advocate of the imagination theory can try to exploit this feature of their account in order to answer the charge that the imagination theory predicts too much circumscription. Neither succeeds.

The first response is to say that believing that one believes can play much the same role as actually believing, and so it's the second-order state that's responsible for the belief-like symptoms of delusion. The second is to say that, if one believes that one believes that P for long enough, one will, over time, come to believe that P.

The first of these seems to make a bad prediction – that delusional subjects will act more delusional in their more reflective moments, and in the more reflective aspects of their behavior, when they are likely to be attending to, and acting on the basis of, their beliefs about their beliefs. But this doesn't seem to be the case.

The second response – that subjects who believe that they believe that P are likely to eventually come to believe that P – doesn't actually seem to help. Unless we postulate a range of states intermediate between belief and imagination that the delusional subject moves through on their way to believing, what we should see is that, until the subject comes to believe, they should just act like imaginers. Then once they come to believe, they should act just like believers. At no stage should we expect to see the peculiar, somewhere-between-belief-and-imagination pattern of circumscription that we actually see.

What would be nice would be to be able to say that the attitude is something *in between* paradigmatic belief and paradigmatic imagination – that delusional subjects are in states that play a role in their cognitive economies that is in some respects like that of a standard-issue, stereotypical belief that P, and in other respects like that of a standard-issue, stereotypical imagining that P.

Can we say this? One might think that we can't, because one might think that the various parts of the belief and imagination roles are a package deal – you can't have the origin of an imagining and the behavior-guiding role of a belief, or a belief-like behavior guiding role *here* and an imagination-like behavior guiding role *there*, or a belief-like origin and an imagination-like updating policy, or...

We'll be particularly likely to think this if we think about token beliefs as particular representational items whose role in a subject's cognitive economy is described by some boxological-functionalist accounts of cognition of the sort provided in, e.g., Nichols and Stich (2003). This sort of account, understood in a certain way, suggests that the belief and imagination roles are all-or-nothing package deals. After all, there's a belief-box and there's an imagination-box, but there isn't any halfway-in-the-middle box. (I should note that this does not seem to be the way that Nichols and Stich themselves understand their account.)

I think that this is a mistake. There are independent reasons to think that the various parts of the paradigmatic belief- and imagination- roles *aren't* a package deal, and there are independent reasons not to understand boxological-functionalist theories in the way that makes them incompatible with the existence of the sort of intermediate<sup>19</sup>

---

<sup>19</sup> "Intermediate" is actually not quite right. It suggests that there is some nice, linear continuum of cases between belief and imagination, along which attitudes will fall into some reasonably straightforward order

attitudes that we'd like to have available in order to accommodate the peculiar patterns of circumscription that we find in delusions.

### 3. *Why the roles aren't package deals: Fragmentation*

Let P be some proposition – some candidate object of belief, imagining, desire, hope, or fear – such as *that snow is white, that sheep eat grass, that my wife has been replaced by an impostor*, etc. Suppose that some representational item R represents P. What role does R have to play in my cognitive economy in order to give rise to a belief that P? Call this role the *belief-role*, and call a representational item that plays the belief-role a *belief-like* representation. Sometimes I'll use 'belief' to talk about particular belief-like representations. I'll use similar terminology for the other propositional attitudes.

We might subscribe to either of two views about the sorts of roles that are available for representations to play. On the *restrictive* view, there are only a few roles available – there are a small number of representation-types, corresponding to the standard propositional attitudes or something like them, and every representation of a given type has the same functional profile. On the *permissive* view, there are very many roles available – particular representations might play any of a number of different functional roles in a subject's cognitive economy, some of which might look very different from those that fit nicely with the standard propositional attitudes.

Here is a way to think about the contrast: on the restrictive view, being a belief comes first, and you get your functional role because you're a belief-type representation.

---

of increasing belief-likeness or imagination-likeness. In fact I think that there are very many dimensions along which an attitude can be more or less paradigmatically belief-like or paradigmatically imagination-like. Since I haven't been able to find a better word, however, I'll continue to use "intermediate".

Representations get tokened in the belief-box, and because they're in the belief-box, they've got the distinctive functional role of a belief. On the permissive view, you get your functional role first, and then you get to be a belief-type representation because your functional role is of the right kind. Representations have whatever functional role they have, and then they count as "being in the belief box" because they have the right kind of functional role.<sup>20,21</sup>

In the rest of this section I'll argue that we ought to endorse the second picture rather than the first. I'll spend most of the section arguing that the belief-role is not nearly as unified as one might have expected. There is enough heterogeneity even within belief to make the availability of intermediate states plausible – it's certainly not the case that there are only a few functional profiles available. The postulation of intermediate states will be plausible if, and to the extent that, paradigmatically belief-like behavior-guidance *here* is separable from paradigmatically belief-like behavior-guidance *there*, and belief-like inferential connections to *these* beliefs is separable from belief-like inferential connections to *those* beliefs.

Before we get started, a word about how the argument will proceed. My goal here will be to undermine the view that I take to be the source of potential principled opposition to the possibility of intermediate, bimagination-type states and attitudes: the view that the belief role and the imagination role are all-or-nothing package deals. I'll spend most of my time when I'm talking about belief arguing that perfectly ordinary, commonplace phenomena give us good reason to think that given representation can play

---

<sup>20</sup> One reason to interpret Currie et al. as advocating some sort of intermediate-state view rather than a pure-imagination view is that Currie and Juredeini (2001) say just this sort of thing about the relative priority of *having a certain functional role* and *being a belief*.

<sup>21</sup> Nichols (2004) appears to endorse the second sort of view, saying that the talk about belief-boxes is meant to be understood as a way of talking about classes of functionally similar representations.

a paradigmatically belief-like role in guiding behavior or generating inferences sometimes and in some respects without playing it all the time in every respect. If I can show that, I will have undermined the claim that the belief-role is an all-or-nothing package deal, and taken off of the table the view that I take to be the source of principled opposition to states with intermediate functional roles.

I *won't* be arguing that perfectly ordinary, commonplace phenomena show us that a given representation can play a paradigmatically belief-like role sometimes and a paradigmatically imagination-like role at other times. (Well, I'll do a little bit of this when I'm talking about imagination. But it won't be the main argumentative goal.) The phenomena that I think provide the best evidence for this are delusions. The goal here is just to open up a space for saying that there are such representations, and such representational states, by undermining the all-or-nothing-roles view which, if we endorsed it, would rule out such intermediate roles, states, and representations.

Now, on to some arguments.

Given the sort of representational theory of mind discussed above, we should expect a certain amount of fragmentation in subjects' beliefs: we shouldn't expect all of our beliefs to be stored in a single, monolithic representation. On the plausible assumption that the cost (in terms of allocation of cognitive and computational resources) of accessing some representational item is proportional to its informational richness, we shouldn't expect all of our beliefs to be brought to bear all of the time. Suppose that geographical information is stored in something like mental maps. Then there's no reason to refer to my whole mental map of the world when I'm figuring out how to get to the deli around the corner, and there's no reason to call up a fully detailed representation

of downtown Ann Arbor, in all of its glorious detail, when what I'm concerned about is just the distance from here to Canberra.

If this is right, then we shouldn't be surprised to find that not every fragment plays precisely the same functional role. It's often said that a belief that P disposes one to act in ways that would be likely to be successful – that would be likely to satisfy one's desires – if P. The paradigmatic case is one where the belief that P disposes the believer to act, *all the time and in every respect* (so long as they are acting intentionally) in ways that would be likely to be successful if P.<sup>22</sup> On the sort of fragmented-belief picture under consideration now, this isn't what we should expect to see. What we should expect instead is that a belief that P disposes the believer to act, *when it's active in regulating the subject's behavior*, in ways that would be likely to be successful if P. Some of our beliefs will be active in guiding certain aspects of our behavior in certain contexts, and others will be active in guiding other aspects of our behavior in other contexts. None of our beliefs will have the full, wide-spectrum behavior-guiding role that characterizes paradigmatic, stereotypical belief. Instead, each will play some *part* of that role – each will have some narrower role, of guiding some behavior, some of the time.

If *this* is right – and we are about to look at some evidence that it is – then at least the behavior-guiding part of the belief-role isn't all-or-nothing. A single representation can play the behavior-guiding role that's distinctive of belief *sometimes* without playing it *all* the time. The behavior-guiding role of belief isn't monolithic.

---

<sup>22</sup> Here “belief that P” means, “belief-like representational item according to which P”, not “belief-like representational item whose whole content is P”. So, one could have a single representational item, whose whole content is, say, that snow is white and cold, which counts both as a belief that snow is white and a belief that snow is cold.



In order to for delusional subjects to bear the sorts of intermediate attitudes that I'm suggesting they bear to the contents of their delusions, I need it to be the case that their representations of the delusional content play a belief-like behavior guiding role sometimes, with respect to some of their behavior, but do not play such a role most of the time, with respect to most of their behavior. The opponent that I'm arguing against here maintains that there can't be any such intermediate attitudes, because behavior-guidance is a package deal. If a certain representation ever plays a belief-like behavior-guiding role, then it always does.

One kind of case that supports a view according to which behavior-guidance is *not* a package deal is the sort of case of inconsistent belief that David Lewis discusses in "Logic for Equivocators" (Lewis 1982). One of his central examples is someone who believes that a certain street – we'll call it A street – runs North/South, believes that another street – we'll call it B street – runs East/West, and believes that A and B streets are parallel. Such cases clearly do occur – we have all, I'm sure, had the uncomfortable experience of discovering an inconsistency in our beliefs at some point or other. Lewis proposes that the right way to understand what's going on in such cases is that the agent is in two distinct belief-states, one according to which A and B streets are parallel and they both run North/South, and another according to which A and B streets are parallel and both run East/West. Sometimes, the agent's behavior is guided by the first belief-state, and sometimes it's guided by the other.

We see the same kind of phenomenon – of some, but not all, of a subject's beliefs being involved in regulating their behavior – in other cases, where what's happening is not that the agent has inconsistent beliefs. The following case is illustrative (names have

been changed to protect the innocent): James is watching television during a thunderstorm. Suddenly, the power goes out. James thinks, “Drat – now I can’t watch TV anymore. Well, I guess I’ll go upstairs and check my email.” Now, James knows full well that computers need electricity, and that the power is out. It’s not that he’s *forgotten* – in the sense of *ceasing to believe* – either of these things. It’s just that he’s not bringing both of these beliefs to bear – one or the other of them is not being used to guide his behavior. (We also have an inference failure here: James is not bringing both of these beliefs to bear on each other in order to draw the inference that the computer isn’t going to work.) I take it that this kind of case, too, is perfectly commonplace.<sup>23</sup>

So, the behavior-guiding role of belief is not an all-or-nothing package deal. The different bits of the stereotypical role – the dispositions to guide certain aspects of behavior in certain circumstances – are separable from one another. A single belief can be disposed to guide one sort of behavior in one sort of context, but not disposed to guide another sort of behavior in a different context.

We also see evidence of the same sort of fragmentation in the role of belief in inference and belief-formation – the inferential role of belief also isn’t monolithic. In order for delusional subjects to bear the intermediate attitudes that I’m suggesting they bear to the contents of their delusions, I need it to be the case that representations of delusional contents play a belief-like role in generating inferences sometimes, in combination with some of the subject's beliefs, but not other times, and not in

---

<sup>23</sup> So Stalnaker (1984) says, for example: “a person may be disposed, in one kind of context, or with respect to one kind of action, to behave in ways that are correctly explained by one belief state, and at the same time be disposed in another kind of context or with respect to another kind of action to behave in ways that would be explained by a different belief state. This need not be a matter of shifting from one state to another or vacillating between states; the agent might, at the same time, be in two stable belief states, be in two different dispositional states which are displayed in different kinds of situations.” (p83)

combination with other beliefs. The opponent I'm arguing against here maintains that there can't be any such intermediate attitudes, because the inference-generating role of belief is a package deal. If a given representation used in generating inferences sometimes, in combination with some beliefs, then it's used in generating inferences all the time, in combo with all of the subject's beliefs.

The familiar cases of inconsistent belief and failures of closure show that this isn't how things actually work. People with inconsistent beliefs don't just infer everything, and it often happens that we find ourselves failing to believe some of the consequences of what we believe. (This is why many deductive arguments are interesting – we believed the premises all along, but didn't believe the conclusion.) Closure takes cognitive work, and some of the consequences of our beliefs are easier to spot than others.

Here is an explanation of this fact: not every belief-type representation is equally tied to every other. There are coordination constraints between belief-type representations. These sorts of connections encourage consistency among belief-type representations, and the elaboration of belief-type representations, or the production of new ones, that represent the consequences of things represented in one's various beliefs. But these connections are not equally strong everywhere. Some of our beliefs are very closely tied to one another, so the elimination of inconsistency and drawing of inferences comes easily or automatically, but there are also pairs of belief-type representations which are not so closely tied to one another, where elimination of inconsistency and drawing of inferences is difficult and/or unlikely.

Another sort of case in which subjects fail to bring to bear some of the things that they believe when forming or updating their beliefs has been investigated by Peter Ditto

and coauthors (in Ditto and Lopez 1992, Ditto et al. 1998, Ditto et al. 2003). In one of Ditto's experiments (Ditto et al. 1998), subjects read a written report of a confederate's supposed impression of the subject. Some subjects are told that the report was written under constraints – that the confederate was instructed to write only positive things about the subject, or only negative things. In the forced-negative condition, the subjects appropriately discounted the evidential value of the report. In their assessments of the confederate's actual opinion of them, subjects who were told that the confederate's report-writing was constrained gave more positive responses than did subjects who were told that the confederate was free to write whatever they pleased. In the forced-positive condition, however, subjects did *not* discount the evidential value of the report. Subjects who were told that the confederate was required to write only positive things and those who were told that the confederate was free to write whatever they pleased gave equally positive assessments of the confederate's actual opinion of them.

Subjects don't appear to *forget* – in the sense of *ceasing to believe* – that the report-writers were constrained. (This was tested for in the experiment.) Instead, what seems to happen is that, in cases where the news is good, the subjects simply fail to bring to bear all of their inferentially relevant beliefs.

So: the belief-role isn't monolithic. Within belief, there's variation in the sort of behavior-guiding role that's played by different beliefs, and there's variation in the sort of inference-generating role that's played by different beliefs.

It's also worth having a look at the diversity of the roles played by actual imaginings.

First off, imaginings do sometimes seem to play a belief-like behavior guiding role. In a cops and robbers game, what seems to happen is that my real beliefs run some aspects of my behavior, while my make-beliefs about cops and robbers run other parts. It's not the case that – or at least it's not obviously the case that – my belief-type representations are hooked up to behavior-guidance, but imagination-type representations never are. Quite plausibly, my beliefs are hooked up to some kinds of behavior, imaginings to others. So it's not as if it's deeply incoherent or wildly revisionary for imaginings to have some bits of the belief-like behavior guiding role. And clearly, if imaginings *are* sometimes behavior-guiding, they're *only sometimes* behavior-guiding – which is further reason to believe that the belief-like behavior-guiding role isn't all or nothing.

We also see inference, or something quite like it, within imagination, and between imagination and belief. First off, there are inferential-type connections across imagination-type representations – we sometimes, as a result of imagining that P, and that if P then Q, come to imagine that Q. So we should think that there are, between imagination-type representations, the same (or very similar) kinds of inferential connections as there are between belief-type representations.

Second, there are coordination constraints between belief and imagination. Walton (1993) discusses a number of cases of imagining and make-believe games that are guided by such constraints – principles of the form, *if it's actually the case that P, pretend that it's the case that Q*. (For example: if someone holds their hand *so*, pretend that they're holding a gun; if there's a lump of mud at a certain location, pretend that there's a pie at that location; etc.) These license revision in both directions – from what's

believed to what's imagined, and from what's to be imagined to what's believed. (If we're playing the mud pies game and I tell you that I left a pie in the oven, you'll come to imagine that there's a pie in the oven, and because of the coordination constraints, you will – quite appropriately – come to *believe* that there's a lump of mud in the hollow stump.)

So, imagination-type representations can stand in the sort of inference-generating connections to other sorts of representations that beliefs seem to stand in to one another, and imagination-type representations can stand in the sorts of relations to belief-type representations that can bring about changes in *belief*. And they can do this without standing in precisely the same sorts of relations to each other, and to beliefs, that beliefs stand in to one another.<sup>24</sup>

Here, I think, is the moral: The belief-role and the imagination-role are a lot more complicated, and a lot less unified, than one might have thought. It's not just a matter of a given representation being hooked up like a belief, or hooked up like an imagining. A given belief-type representation will have a whole range of different connections to different behavior-planning mechanisms (or to different bits of the one mechanism, or different kinds of connections to the same bit of the same mechanism, or...) and a whole range of different kinds of connections to different representations of various types. There are no necessary connections between these various connections – it's not the case that anything that's got one element of a certain package has also got to have all of the rest, since we see a variety of mix & match patterns even within belief. No belief-type representation plays the whole stereotypical belief-role – regulating all behavior all the

---

<sup>24</sup> Another interesting class of phenomena to look at would be the sorts of 'contagion' phenomena discussed in Gendler (2003). These, I think, provide more evidence that the belief/imagination divide is not as sharp as we might have thought, and that delusional subjects are not the only bimaginers.

time, and being equally and perfectly coordinated with all of one's other beliefs. The different bits of the stereotypical role – for example, regulating *this* bit of behavior in *these* circumstances, and combining with *these* sorts of beliefs to generate inference – are separable.

So there seems to be no principled reason to think that we can't get a spectrum of cases, from clear, totally non-belief-like imaginings to clear, full-blooded, paradigmatic beliefs, with intermediate, hard-to-classify states in the middle.

#### *4. Two Ways to Organize a Library*

Here is a metaphor that I find helpful for thinking about theories of the functional roles of mental representations: Suppose you have a library, and you want to file the books in a way that allows the right people to look at the right books for the right purposes at the right times. One way to do it is to just have two kinds of labels: *fiction* and *non-fiction*, and encourage those people who need books that correctly characterize the world to look at the non-fiction, and those people who need books for some other purpose to look at the fiction. This system has the advantage of making the distinction between fiction and non-fiction extremely sharp and easy to maintain, and so making it extremely unlikely that anyone whose needs would be best served by a non-fiction book will mistakenly wind up looking at a fiction book instead. It has the disadvantage of forcing people to sift through a lot of superfluous information.

Another option would be to have a more complicated filing system, where each book might receive a number of labels, each of which indicates that the book is to be referred to by some particular group of consumers in some particular circumstance. This

sort of system has the advantage of allowing consumers to receive only the information that's relevant to their purposes. It has the disadvantage of making mislabeling easier – once we move away from the binary fiction/non-fiction labeling system, we open up the possibility of a book that's meant only to be sent to the fiction-consumers accumulating, through confusion or clerical error, a couple of labels that occasionally send it out to the consumers who would be better served by non-fiction.

The phenomena discussed in the previous section motivate a move away from the two-tag, “take seriously”/“don't” tagging picture to something more complicated. That allows for flexibility, among the seriously-taken things, about just what they're taken seriously for and when, and similarly for the non-seriously-taken things. But it also leaves room for the taken-seriously/not distinction to go blurry – the same thing could be treated seriously in one respect and not in another. If that happens, then we'll get hybrid states, intermediate between belief and imagining. And those would be the right sorts of things to display a lot of the peculiar behavior that's distinctive of delusions.

##### *5. An Overgeneration Problem?*

Here is a possible source of concern about this sort of picture: If we get some mixing and matching, why don't we just get it willy-nilly? Why do we so often see people bearing attitudes that are very close to stereotypical belief to various propositions, and so rarely see anyone bearing these peculiar intermediate attitudes to anything? Why do we see the kind of clustering of functional roles of actual people's representations – where, in people's actual representations, the distinctive features of the stereotypical belief-role tend to go together, and the distinctive features of the stereotypical



imagination-role tend to go together – that would make the package-deal picture (and folk psychology in general) attractive in the first place?

The answer, very briefly, is that agents whose representations tend to have something like the stereotypical roles will tend to do much better than those whose representations tend to have intermediate, mix-and-match roles. We shouldn't expect to see the peculiar hybrid roles all over the place because they're pretty maladaptive. The stereotypical features of belief go together nicely – we want the features that guide our behavior to be ones that originate and get revised in belief-like ways. Letting your serious behavior be guided by representations that have the sorts of connections to evidence that are distinctive of imagination, rather than those distinctive of belief, is likely to get you in trouble. So is allowing *too* much of a difference in the contents of the representations that guide the different aspects of your behavior in different circumstances.

It's a bad plan – maladaptive and irrational – to treat your representations whose origins and updating policies aren't truth-directed origins as sources of start-states for your planner. The bits of the paradigmatic belief- and imagination- roles complement each other well, and don't undermine each other in the way that some of the mixing and matching would. So there's good reason to expect to see the kind of clustering that we do – deviations from it are signs that something's gone wrong.

#### *6. Self Deception and Attitudes Intermediate between Belief and Desire*

I'll close with a very brief discussion of another place where it might be explanatorily helpful to postulate an attitude that's intermediate between two of the

familiar propositional attitudes. Some cases of self-deception seem to display the same sort of peculiar circumscription and insensitivity to evidence that's characteristic of delusions. In these cases, we may have the same sort of reluctance to say that the self-deceiver genuinely *believes* that the relevant proposition is true, but it also doesn't seem right to say that they merely *desire* that it's true, either. Instead, they seem to be in a state intermediate between belief and desire. We might, to appropriate some terminology from metaethics,<sup>25</sup> call such states 'besires'.

This would allow for the self-deceiver's "belief" to be insensitive to evidence for its falsity in the same way as a desire, and yet to play some part of the behavior-guiding role of belief. It would also allow us to account for cases in which the self-deceiver's "belief" has an impoverished behavior-guiding and inferential role, as seems to sometimes be the case. I don't want to suggest that this is the right account of self-deception *in general*. I suspect that self-deception is a many-splendored thing, and that there won't be any single, unified account of self-deception in general to be found. Instead, I want to suggest that this sort of intermediate-state account is the right way to describe what's going on in some restricted class of cases that might plausibly fall under the heading of "self-deception" – the ones that display the same peculiar sort of evidence-independence and circumscription that we see in delusions.

There are, broadly speaking, two sorts of accounts of the origin of intermediate attitudes. This sort of account of self-deception seems to fit better with one than the other. On the first account, the representations are peculiar from the get-go: something goes wrong in the original construction of the representation, such that it winds up with a

---

<sup>25</sup> In particular, from Altham (1986). One reason to be cautious about appropriating this terminology, though, is that the sorts of intermediate states I'm discussing here are not the same sorts of things that go by the name 'besire' in metaethics.

non-standard role in the subject's cognitive economy. On the second sort of account, the problematic representations start off with a fairly standard functional role, and then they drift into some intermediate area. (Possibly eventually drifting all the way over to some other paradigmatic role. This would be the case of the imaginer who eventually comes to believe – not all at once, but gradually, as the representation's functional role drifts from a paradigmatically imagination-like role into an intermediate area, and then eventually into a paradigmatically belief-like role.)

The self-deception account, I think, fits best with this second kind of origin. The most plausible sort of story seems to be one on which some representations start life as desires, but eventually acquire some aspects of the functional role of a belief.

(Why would this happen? I don't have very much of a story to offer here, but here are some speculations: perhaps it sometimes happens that a desire that's often brought consciously to mind, fantasized about, etc. comes to acquire a certain belief-like vividness, which causes it to be treated, in some respects, like a belief. Or alternatively, perhaps sometimes a strong desire that P leads to frequent and vivid *imagining* that P, and it's the imagining, rather than the desire, which acquires a belief-like role due to its increasing prominence in the subject's mental life.)<sup>26</sup>

### *Conclusion*

Delusions are not happily classified as either straightforward cases of belief or straightforward cases of imagining. It would be convenient, then, if those weren't the only classificatory options in the neighborhood. I have argued that there is no principled

---

<sup>26</sup> Here is another place where there are connections with ideas in Gendler (2003), which unfortunately I will not be able to explore here.

reason to deny that other sorts of mental states, intermediate between belief and imagination are possible, and that it's plausible that having a delusional "belief" is such a state. I've also suggested that such intermediate states might be helpful in categorizing, and perhaps in explaining, certain kinds of self-deception.

The claim that this is what's happening in any actual cases of delusion and/or self-deception is, of course, very speculative, and subject to empirical refutation. It could very easily be completely wrong. What is much less speculative, and much less likely to be refuted empirically, is that postulating such intermediate states is not conceptually incoherent, and not incompatible with any obvious facts about our actual psychologies. So while I don't take myself to have established that some intermediate-state account of delusion or self-deception is *true*, I do take myself to have established that it's not *crazy*. If it is wrong, it's not for conceptual reasons, and not because it's incompatible with a lot of obvious stuff that we know about the mind. If it's wrong, it's because we happen not to be wired that way, not because we *couldn't* have been wired that way and still have roughly the kind of psychology that we actually have.

## References:

- Altham, J. (1986) "The Legacy of Emotivism", in *Fact, Science, and Morality: Essays on A.J. Ayer's Language, Truth and Logic*, Macdonald and Wright, eds., Blackwell.
- Bayne, T. and Pacherie, E. (2004a) "Bottom-up or Top-down?: Campbell's Rationalist Account of Monothematic Delusions", *Philosophy, Psychiatry, & Psychology*, 11/1: 1-11.
- Bayne, T. and Pacherie, E. (2004b) "Experience, Belief, and the Interpretive Fold", *Philosophy, Psychiatry & Psychology* 11/1, 81-86
- Bayne, T. and Pacherie, E. (2005) "In Defense of the Doxastic Conception of Delusions", *Mind and Language*, 20/2: 163-88.
- Berrios, G. (1991) "Delusions as 'Wrong Beliefs': A Conceptual History", *British Journal of Psychiatry* 167, 153-5.
- Bleuler, E. (1924). *Textbook of Psychiatry (4th ed)*, trans A.A. Brill, Macmillan.
- Campbell, J. (2001) "Rationality, Meaning, and the Analysis of Delusion", *Philosophy, Psychiatry & Psychology* 8, 89-100.
- Currie, G. (1997) "The Paradox of Caring", in M. Hjort and S. Laver, *Emotion and the Arts* (Oxford University Press).
- Currie, G. (2002) "Imagination, Delusion, and Halucinations", *Mind and Language* 15:1, 168-83
- Currie, G. and Jureidini, J. (2001) "Delusion, Rationality, Empathy: Commentary on Martin Davies et al", *Philosophy, Psychiatry & Psychology* 8, 159-62.
- Currie, G. and Ravenscroft, I. (2002) *Recreative Minds*, Oxford University Press.
- Davies, M. and Coltheart, M. (2002) "Pathologies of Belief", *Mind and Language* 15:1, 1-46.
- Davies, M., Coltheart, M, Langdon, R, and Breen, N. (2001) "Monothematic Delusions: Toward a Two-Factor Account", *Philosophy, Psychiatry & Psychology* 8, 133-158.
- Ditto, P., & Lopez, D. (1992) "Motivated Skepticism: The Use of Differential Decision Criteria for Preferred and Nonpreferred Conclusions" *Journal of Personality and Social Psychology* 63, 568-584.
- Ditto, P., Scepansky, J., Munro, G., Apanovitch, A., & Lockhart, L. (1998) "Motivated Sensitivity to Preference-Inconsistent Information" *Journal of Personality and Social Psychology* 75, 53-69.
- Ditto, P., Munro, G., Apanovich, A., Scepansky, J., & Lockhart, L., (2003). Spontaneous skepticism: The interplay of motivation and expectation in responses to favorable and unfavorable medical diagnoses. *Personality and Social Psychology Bulletin* 29.
- Gendler, T. (2003) "On the Relation Between Pretense and Belief", *Imagination, Philosophy and the Arts*, Routledge.
- Lewis, D. (1982) "Logic for Equivocators" *Nous* 16, pp 431-41.
- Maher, B. (1974) "Delusional Thinking and Perceptual Disorder." *Journal of Individual Psychology*, 30: 98-113.
- Maher, B. (1999) "Anomalous Experience in Everyday Life: Its Significance for

- Psychopathology”, *The Monist* 82, 547-70.
- Nichols, S. (2004) “Imagining and Believing: The Promise of a Single Code”, *Journal of Aesthetics and Art Criticism*.
- Nichols, S. and Stich, S. (2003) *Mindreading*, Oxford University Press.
- Sass, L. (1994) *The Paradoxes of Delusion: Wittgenstein, Shreber, and the Schizophrenic Mind*, Cornell University Press.
- Stalnaker, R. (1984) *Inquiry*, MIT Press.
- Stone, T. and Young, A. (1997) “Delusions and Brain Injury: The Philosophy and Psychology of Belief”, *Mind and Language* 12:3/4, 327-64.
- Walton, K. (1990) *Mimesis as Make-Believe*, Harvard University Press.
- Walton, K. (1993) “Metaphor and Prop Oriented Make-Believe,” *The European Journal of Philosophy* 1/1, 39–57.
- Walton, K. (1997) “Spelunking, Simulation, and Slime”, In M. Hjort and S. Laver, *Emotion and the Arts* (Oxford University Press).
- Young, A. (2002) “Wondrous Strange: The Neuropsychology of Delusional Beliefs” *Mind and Language* 15:1, 47-73.