Delusions: A two-level framework *

Keith Frankish

Although well-documented, delusions have proved extremely hard to explain, and many important questions remain open, including the basic one of what kind of mental state a delusion is. The standard position is that delusions are beliefs (the *doxastic conception*); but there are difficulties for this view, and alternative characterizations have been offered. In this chapter I shall propose a new framework for conceptualizing delusions, building on recent work in philosophy of psychology and cognitive science. There are good reasons for thinking that the term 'belief' is commonly used to refer to two different types of mental state, located at different levels. This view harmonizes with work in the psychology of reasoning, where many researchers now endorse some form of *dual system* theory. I shall outline what is, I believe, the most attractive version of this two-level view and show how it offers an account of delusions that explains our competing intuitions about their status.

The chapter is in four sections. The first introduces the doxastic conception and its problems. The second distinguishes the two levels of belief, and argues that delusions, if they are beliefs at all, belong to the second. The third section offers an account of second-level belief, according to which it is a species of a broader mental type, *acceptance*, which is dependent on attitudes at the first level. The fourth section proposes that delusions are acceptances, some of which fall within, and some without, the narrower class of second-level beliefs, and the chapter concludes with some reflections on the implications of this view. Throughout, I shall focus on monothematic delusions, rather than the elaborate polythematic kind, and use simple, schematic examples. This is not because I think it is unimportant to pay attention to the diversity of delusions and the detail of clinical observation (far from it). Rather, it reflects the modest aim of the chapter, which is to propose a hypothesis for subsequent elaboration and evaluation.

1. The doxastic conception and its problems

People suffering from delusions make strange claims. To take three well-known examples, patients with the Cotard delusion claim that they are dead, those with the Capgras delusion claim that a loved one has been replaced by an impostor, and those with the Fregoli delusion claim that someone they know is disguising themselves as other people. Do they believe these things, as the doxastic conception has it? The standard view, enshrined in DSM-IV, is that they do: a delusion is defined as a false

* NOTICE: this is a version of a chapter published in M. Broome and L. Bortolotti (eds.), *Psychiatry as Cognitive Neuroscience: Philosophical Perspectives* (pp. 269-84), Oxford University Press, 2009. This version does not reflect changes resulting from the publishing process, such as copy-editing, corrections, and formatting.

belief which is firmly held despite incontrovertible evidence and which is not shared by others in one's community (American Psychiatric Association 2000, p.821).

Prima facie, this is a plausible view. Deluded patients make their claims firmly and with apparent sincerity, and seem to understand the meanings of the words involved (Bayne and Pacherie 2004a; Hamilton 2007). Speech acts of this kind are standardly taken as expressions of belief, and the subjects themselves appear to take them that way. Moreover, it is precisely because deluded patients seem to believe their claims that we find their condition so bizarre and disturbing. Imagining or wishing that one is dead is not as strange a condition as believing that one is dead (though it might still be worrying). And although it is hard to see how a person could come to believe that they are dead, the content of such a delusion is not much stranger than that of some religious beliefs, which are widely accepted and therefore not considered delusional.

The doxastic conception has its difficulties, however, and many writers have stressed the differences between delusions and beliefs. Four stand out. First, delusions are often unsupported by reasons; deluded patients either offer no reasons for their claims or offer inappropriate ones (Campbell 2001; Hamilton 2007). Second, delusions tend to be relatively cognitively inert; deluded patients often fail to draw obvious conclusions from their delusions and are not bothered by the lack of consistency between them and their other beliefs (Breen et al. 2000; Brett-Jones et al. 1987). Third, delusions may be relatively behaviourally inert; many deluded patients do not consistently act on their delusions, and their assertions are often at odds with their non-verbal behaviour (Bovet and Parnas 1993; Currie 2000; Hamilton 2007; Sass 1994; Young 2000). Some Capgras patients, for example, continue to live on friendly terms with the supposed impostor (for data, see Young ibid. p.53). Indeed, as Hamilton notes, in the case of some bizarre delusions, it will often be *impossible* to act on them (Hamilton ibid.). Finally, delusions often lack the emotional associations the corresponding beliefs would normally have (Sass 1994).

These differences should not be overstated. It is arguable that some deluded patients *do* have evidence for their claims, in the shape of distorted perceptual inputs (e.g. Davies et al. 2001; Maher 1999; Stone and Young 1997). And delusions are certainly not *completely* inert (Buchanan and Wessely 1998). Deluded patients typically report their delusions, and they sometimes act on them in other ways — even with violence. (This is often what brings them to the attention of the medical services.) But it is widely agreed that delusions are somehow compartmentalized and not fully integrated into the subject's belief system. Jaspers, for example, noted that a patient's attitude to the content of their delusion is 'peculiarly inconsequent at times' (Jaspers 1913/1997, p.105), and Young refers to the 'curiously circumscribed quality' of delusions (2000, p.49).

The issue here is not merely taxonomic, but bears on the question of whether we can make sense of delusions — whether deluded patients can be brought within the 'interpretive fold' (Bayne and Pacherie 2004b). Even if we allow that deluded subjects do in some sense believe their claims, it is doubtful that we could form any empathic understanding of their state of mind if these beliefs are not subject to the usual

rational constraints and do not have the usual sensitivity to evidence and influence on action.

Not everyone agrees that these considerations refute the doxastic conception, of course, even if they are granted. It is arguable, for example, that many non-delusional beliefs are also compartmentalized and held in violation of rational norms (Bortolotti 2005). Some theorists, however, regard the doxastic conception as unsustainable, and propose alternatives to it. John Campbell argues that delusions express 'framework propositions', which are immune from questioning and set the context for reasoning and the interpretation of experience (Campbell 2001). Greg Currie and his collaborators propose a *metacognitive* account, according to which delusions are imaginings which those experiencing them misidentify as beliefs: deluded patients believe that they believe their claims (which is why they assert them), but in fact only imagine them (which is why they do not act upon them) (Currie 2000; Currie and Jureidini 2001; Currie and Ravenscroft 2002). Others take a more pessimistic line. Andy Hamilton argues that there is no fact of the matter about whether or not deluded patients believe their claims, since the verbal and non-verbal behavioural criteria for the ascription of belief do not cohere (Hamilton 2007). German Berrios claims that delusional claims are simply 'empty speech acts that disguise themselves as beliefs' (Berrios 1991, p.8).

Now an important thing to note about this debate is that it is as much about belief as it is about delusion. The concept of belief is a folk-psychological one, defined by its role within our everyday practices of psychological explanation and prediction. But there is disagreement about what this role is, and different views on the matter have different implications for the doxastic conception. One basic contrast concerns the function of folk psychology itself. On an 'interpretivist' view, folk psychology has a rationalizing function, and ascriptions of beliefs and desires are made as part of an overall interpretation of a subject which aims to render their behaviour rationally intelligible in the light of their experience (e.g. Davidson 1984; Dennett 1987). Thus, on this view, to regard a person as the subject of mental states is precisely to bring them within the interpretive fold. This view does not harmonize well with the doxastic conception; if there are strict rational constraints on belief ascription, then we will not be warranted in ascribing beliefs that are unsupported, inconsistent, and partially inert. Indeed, given that interpretation is governed by a principle of charity, it is hard to see how we could be justified in interpreting a person as believing a very bizarre content, such that of Cotard's delusion, as opposed to merely believing (say) that it will help to *claim* to believe it.

On the 'theory-theory' view, by contrast, folk psychology is a rudimentary science of the mind, and mental-state concepts refer to internal states of our cognitive systems (e.g. Lewis 1972). This approach is more compatible with the doxastic conception; when we consider the structure and limitations of the human cognitive system, the rational constraints on belief ascription appear weaker and the doxastic conception correspondingly more plausible (Bayne and Pacherie 2005).

It is usually assumed that we must decide between these different views of the role of belief, but this may be a mistake. There is a strong case for thinking that the everyday concept of belief picks out two distinct states with different properties. If so, this may go some way towards explaining, and reconciling, our competing intuitions about the function of folk psychology and, consequently, about the doxastic status of delusions. The next section outlines the case for this view.

2. Two types of belief

There are certain core elements to the concept of belief. Beliefs have propositional content; they have mind-to-world direction of fit (they represent their contents as obtaining, rather than as to be made to obtain); and they guide behaviour in a way that reflects their content and direction of fit. But there are also tensions within the concept of belief which suggest that it may not pick out a unitary state. I shall mention four of these. Note that what follows applies to *non-perceptual* beliefs; perceptual beliefs are a special case, and I shall not discuss them here.

The first tension concerns consciousness. On the one hand, we think of beliefs as part of the furniture of our conscious minds, to which we have direct introspective access. They are things we have thought about and endorsed, which we rely on in our conscious reasoning, and which we would avow if prompted. On the other hand, we also possess a mass of beliefs which we have never thought about, which operate silently, in the background, and which we are only indirectly aware of possessing. Think, for example, of the huge corpus of knowledge about the location and function of household objects which guides your behaviour around the house (where the cupboards are, how the taps work, which way the fridge door opens, and so on). Our behaviour shows that we have this knowledge, but we have no introspective awareness of it, and often the only way to access it is to imagine ourselves performing some relevant action (say, opening the fridge door), and noting how we do it (James 1890 ch.4).

The second tension relates to control. We often speak of ourselves as having a degree of direct control over our beliefs: we talk of *deciding* what to think and *making up* our minds. The idea is that we can consider a proposition, reflect upon the evidence for and against it, and then decide whether to give, or withhold, assent to it. (This is not to say that we can choose to believe anything we like; we can assent to a proposition only if we have good epistemic support for it; see my 2007.) Similarly, we can decide to revise or reject beliefs we no longer regard as warranted. Yet not all beliefs are susceptible to this kind of reflective control; many are formed and modified automatically, by subpersonal processes over which we have no direct control.

The third tension concerns degree. We usually think of belief as a binary, or flat-out, attitude: we either believe something or we don't. Yet we also speak of people having degrees of confidence, which are continuously variable, and much of our behaviour can be interpreted as the upshot of probabilistic reasoning involving these

states (e.g. Oaksford and Chater 2007). Desire, too, can be thought of as either an on-off binary state or a matter of graded preference.

The final tension relates to the ontological status of belief, where, simplifying somewhat, we can distinguish functionalist and dispositionalist views. On the former, beliefs are functionally discrete states of the cognitive system, which are selectively activated in reasoning (e.g. Fodor 1987; Ramsey at al. 1990); on the latter, they are multi-track behavioural dispositions, which are holistically interdependent (e.g. Dennett 1987; Ryle 1949). By focusing on different aspects of folk-psychological practice, a strong case can be made for each view.

The importance of these tensions is often underrated. Typically, theorists assume that there is just one form of belief, and either dismiss one side of each tension or argue that it corresponds to a superficial variant of the core state. This is, I think, a mistake. I have argued elsewhere that the tensions are real and run deep (Frankish 2004, forthcoming(a)). Moreover, the contrasting views of belief line up, suggesting that they correspond to two distinct forms of the state. The controlled/automatic contrast coincides with that between conscious and nonconscious; only our conscious beliefs are susceptible to direct reflective control. The binary/graded contrast also lines up with the conscious/nonconscious one. Conscious belief is typically binary: we either avow belief or withhold it, and our conscious reasoning usually proceeds from categorical premises to categorical conclusions. (Even when we form conscious beliefs about probabilities, such as that there is a 50% chance of rain, the *attitude* is unqualified.) Our degrees of confidence and preference, on the other hand, like our background knowledge, are unavailable to introspection and influence our behaviour at a nonconscious level. Finally, a functionalist view is required for conscious belief, whereas the intuitions supporting a dispositionalist position are best construed as relating to nonconscious belief. Conscious beliefs are functionally discrete (they can be selectively acquired, recalled, and lost), and they require activation in order to influence action (if we fail to recall a belief — if it *slips our mind* — then it has no effect on our behaviour). However, there is no reason to think that nonconscious beliefs have these properties, and indeed it seems unlikely. Implicit beliefs typically come in clusters, and they influence action in a holistic way; think of the mass of background knowledge that is relevant to even very simple everyday actions.

So we can tentatively distinguish two types of belief: one (call it *level 1*) that is a nonconscious, passive, graded, dispositional state, and another (*level 2*) that is a conscious, controlled, binary, functional state. (I shall say more later about why talk of levels is appropriate here.) A formally similar distinction can be made for desire.

There are precedents for this division within the category of belief. Daniel Dennett distinguishes between belief and opinion (Dennett 1978, ch.16), corresponding roughly to level 1 and level 2 belief, and some philosophers of science draw a distinction between belief and acceptance, which I shall discuss in the next section. The division also harmonizes well with recent work in both cognitive and social psychology, where many researchers have converged on some form of *dual-system* theory, according to which humans have two distinct systems for reasoning and

decision-making, often referred to as *System 1* and *System 2* (e.g. Evans 2003; Evans and Over 1996; Kahneman and Frederick 2002; Sloman 1996; Stanovich 2004; see also Evans and Frankish, forthcoming). Theorists differ on the details, but the typical picture is as follows. System 1 is a collection of autonomous subsystems, many shared with animals, whose operations are non-conscious, automatic, fast, parallel, associative, heavily contextualized, and undemanding of working memory. System 2, by contrast, is a uniquely human system, whose processes are conscious, controlled, slow, serial, rule-governed, decontextualized, and demanding of working memory. Such theories were originally developed to account for well-documented conflicts between logical and non-logical processes in deductive reasoning tasks and between avowed attitudes and actual behaviour in social contexts, but they can also explain many other phenomena, including individual differences in reasoning, cross-cultural variation in cognition, and aspects of cognitive development (see the papers by, respectively, Stanovich, Buchtel and Norenzayan, and Klaczynski, in Evans and Frankish, forthcoming).

Dual-system theorists typically assume that the two reasoning systems have separate databases, and it is attractive to link the two types of belief we have distinguished with the two systems — level 1 belief with System 1 and level 2 belief with System 2. (Of course, if level 1 beliefs are behavioural dispositions, then we cannot identify them with the *inputs* to System 1 reasoning; that would be a category mistake. The inputs will be subpersonal informational states of some kind, which lie outside the purview of folk psychology. But level 1 beliefs can be seen as *manifestations* of System 1 activity and the associated subpersonal informational states.) This view is further bolstered by social-psychological work on persuasion and attitude change, where many theorists distinguish implicit and explicit memory systems, the former nonconscious, automatic, and slow-learning, the latter conscious, effortful, and fast-learning (e.g. Wilson et al. 2000; Smith and DeCoster 2000). Again, this distinction aligns well with that between the two types of belief; level 1 beliefs are typically slow to form and change, whereas level 2 beliefs can be formed and revised in one-off episodes. Consider, for example, what happens when normal background conditions suddenly change, as when a light bulb blows. We immediately form the level 2 belief that the bulb is blown and, consequently, that the light switch will not work; but on entering the room we still press the light switch, manifesting a level 1 belief that it will work.

This is, of course, only a sketch of the case for the level 1 / level 2 distinction, and there is much more to be said about the relation between this distinction and related ones in scientific psychology (see my forthcoming(b)). But it is a useful working hypothesis and I now want to consider its implications for the doxastic conception of delusions.

Three points need to be made. First, although the two types of belief will often go together, they are dissociable. We can believe something in a nonconscious way without consciously assenting to it, and we can consciously assent to something without it penetrating to the nonconscious level and affecting our spontaneous

behaviour. (Strictly speaking, if level 1 belief is graded it would be better to say that we can believe something at a conscious level while believing it only *very weakly* at the nonconscious level; on Bayesian principles a rational agent will have some degree of confidence in every relevant proposition of whose falsity they are not certain. For simplicity's sake, however, I shall often speak as if level 1 belief were binary.) Indeed, we can harbour conflicting attitudes at the two levels. The light bulb case is an example, and the social-psychological literature contains abundant evidence of conflict between implicit and explicit evaluative attitudes. In such cases, the behavioural criteria for belief may be seriously out of step with the verbal criteria.

Second, delusions fit the distinctive profile of level 2 belief better than that of level 1 belief. Delusional reports are conscious, and they manifest attitudes that are binary, discrete, and at least available to reflective control, however hard it may be for the subject to exercise it. Indeed, it is hard to make sense of the notion of an *implicit* delusional belief.[1] The relative behavioural inertia of delusions reinforces the classification of delusions at level 2, since level 2 beliefs have a more restricted influence on action than level 1 beliefs. In order to influence action, level 2 beliefs must be activated in episodes of conscious thought, whereas level 1 beliefs manifest themselves in spontaneous, unreflective behaviour.

Third, the distinction between level 1 and level 2 belief aligns with the two views of the function of folk psychology mentioned earlier. If we focus on level 1 belief, then an interpretivist perspective is called for. To have a level 1 belief with content *p* is to be disposed to behave in ways that would be rational if *p* were true, so being consistently interpretable as possessing a given level 1 belief is both necessary and sufficient for possessing it. When we focus on level 2 belief, on the other hand, a theory-theory perspective is appropriate. These beliefs are functional states of the cognitive system, which can be selectively activated and which are subject to selective failures of activation (slips of mind). It follows, then, that the rational constraints on the ascription of level 2 beliefs are much weaker than those on the ascription of the level 1 kind. Thus the relative inertia of delusions tells strongly against their being level 1 beliefs, but much less strongly against their being level 2 beliefs.

The moral of this for the doxastic conception is obvious. Delusions may be beliefs of one type but not the other, and the evidence suggests they are not level 1 beliefs. So if delusional claims are believed, it will be at level 2 only. (Or at least, this is plausible for the problematic monothematic delusions we have focused on; I do not wish to claim that *no* delusions are also believed at level 1.)[2] Of course, even if this is right, it

---

1 Matthew Broome has pointed out to me that when patients are very ill, they may become mute and posturing, so that it is no longer easy to attribute delusions to them (though they themselves may later offer post-hoc delusion-related explanations for their behaviour). In so far as such behaviour is intentional, it is probably best seen as spontaneous and unreflective, manifesting non-delusional level 1 attitudes of some kind. This again suggests that delusions involve a higher level of cognitive activity, and that a patient can become, as it were, *too ill to be deluded*.

2 It is tempting to speculate that it is because delusions are believed only at level 2 that they lack the normal emotional associations. Level 2 belief is an intellectual form of belief, which is less intimately

does not end the debate; the objections to the doxastic conception may extend to the claim that delusions are level 2 beliefs. In order to make progress here, I want to outline a hypothesis about the nature of level 2 belief, which will open a further option for the classification of delusions, and may have far-reaching implications for our thinking about them.

The hypothesis is designed both to account for the distinctive features of level 2 belief and to address a problem raised by the level 1 / level 2 distinction. The problem concerns the role of level 2 belief in the guidance of action. If level 1 beliefs are behavioural dispositions, which are ascribed on the basis of an overall interpretation of the agent's behaviour, then there seems no room for level 2 belief to influence action. All behaviour that is interpretable as intentional will be interpretable as a manifestation of graded nonconscious level 1 beliefs. (I have referred to this as the *Bayesian challenge*; Frankish 2004.) Conscious, level 2 belief threatens to be behaviourally inert. Some writers are tempted by this conclusion. Dennett suggests that our opinions (roughly equivalent to level 2 beliefs) influence our verbal behaviour only (Dennett 1978, ch.16), and some cognitive scientists hold that conscious decision-making is largely confabulatory, serving merely to rationalize intuitive responses generated by non-conscious processes (e.g. Gazzaniga 1998; Wegner 2002; Wilson 2002). This view would also explain the inertia of delusions, on the assumption that they are level 2 beliefs. Nevertheless, it should be resisted. It may be that conscious thought has far *less* effect on behaviour than we ordinarily suppose, but it is highly implausible to hold that it has none, and in what follows I shall outline an account of level 2 belief which does not have this consequence. We shall then need a further explanation for the relative inertia of delusions.

3. Level 2 belief and acceptance

The account I propose builds on, and substantially modifies, existing accounts of what some philosophers of science refer to as *acceptance* — a psychological state that is often contrasted with belief. There are a number of independent versions of the belief/acceptance distinction, each addressing different concerns and fostering different conceptions of the two states (for a survey, see Engel 2000). I shall focus on a version developed by Jonathan Cohen (Cohen 1992). According to Cohen, belief is a disposition, which is involuntary, graded, and truth-directed; to believe that *p* is to be disposed to feel it true that *p* when you consider the matter. Acceptance, on the other hand, is a policy, which can be actively adopted in response to pragmatic considerations. To accept *p* is to decide to *treat it as true* — to take it as a premise in one's conscious reasoning and decision-making. For example, I might decide that, for the purposes of deciding what food to buy, I shall take it as a premise that beef is unsafe to eat, even if I am not completely convinced that it is. Cohen also identifies a parallel conative state, goal adoption, which involves committing oneself to taking

---

connected with behaviour and bodily responses, and it may be that the affective component of a level 2 belief depends on the existence of a corresponding level 1 belief.

some outcome as a goal. (For convenience I shall use the term 'premising policy' for both acceptance and goal adoption.) Cohen's distinction between belief and acceptance clearly overlaps with that between level 1 and level 2 belief, particularly in the passive/active contrast, and I think his conception of acceptance offers a fruitful model for level 2 belief.

Cohen does not say much about what is involved in executing premising policies or what role such policies have in the guidance of action. In other work I have addressed these questions and developed a detailed account of the nature and function of acceptance (see my 2004, ch.4). The key idea is that there is a level of human reasoning which is conscious and under intentional control — that is, which involves actively applying learned inferential procedures and problem-solving strategies, motivated by a desire to find a solution to some problem and a belief that the procedures employed may generate one. The activities involved — which will often employ inner speech — might include constructing deductive arguments, applying heuristics, running thought-experiments, imagining scenarios, or simply questioning oneself in order to stimulate spontaneous inference. I call this type of reasoning *personal reasoning*, in contrast to reasoning that is nonconscious and subpersonally controlled, and I have argued that System 2 reasoning, as described by dual-system theorists, is best understood as personal in this sense (see my forthcoming(b)). (We might think of System 2 as a *virtual* system, in the sense described in Dennett 1991.) Because personal reasoning is under intentional control, we can commit ourselves to regulating it in various ways, and premising policies, I suggest, are just such commitments. In accepting a proposition or adopting a goal, we commit ourselves to taking it as an input to our personal reasoning.

Now it is attractive to think of level 2 beliefs as acceptances in this sense. Acceptance is by definition conscious and controlled. It is also a binary state: for any proposition, p, one either has or has not adopted a policy of premising that p. (It is true that we can have varying degrees of attachment to our premising policies, but so long as we hold on to a given set, our commitment to each of them will be the same.) Moreover, acceptance states are functionally discrete, in the way that level 2 beliefs are. Premising policies can be individually adopted and abandoned, and they can be selectively recalled and executed in personal reasoning (see my 2004, ch.6). It is true that premising policies are not *brain* states — it is people who have policies, not their brains — but from a functional perspective this is irrelevant.[3] Finally, this account offers a solution to the problem of how level 2 beliefs can influence action. This requires a little explanation, but will be important later.

The actions involved in forming and executing a policy are intentional ones, motivated by the desirability of adhering to the policy. We adopt policies because we think that following them will bring some long-term benefit, and we perform the actions they dictate because we want to secure this benefit. (In the case of premising

[3] To underscore the point that level 2 beliefs are formed and processed at a personal level, I have elsewhere dubbed them 'virtual beliefs' and 'superbeliefs'; see Frankish 1998, 2004.

policies, I assume this benefit is that of having a settled knowledge base and goal structure to draw on in one's personal reasoning. The advantages of forming a premising policy are precisely those of *making up one's mind*.) Thus, when we take our premises and goals as inputs to personal reasoning, we do so because we believe our premising policies require us to do this. Of course, these beliefs will not normally be conscious ones. The attitudes that drive our premising activities will typically be of the nonconscious level 1 kind, which reveal themselves in our attitudes to our conscious premises and goals.

How does this explain how level 2 beliefs influence action? Well, suppose that premising policies also involve a tacit commitment to overt action. That is, suppose that in adopting a premise for use in personal reasoning we also commit ourselves to acting upon the results of that reasoning — adopting any derived conclusions as further premises or goals, and performing, or forming intentions to perform, any dictated actions. Then if we believe that our premises and goals mandate a certain action, we shall be motivated to perform the action precisely because our policies dictate it — again the motivating attitudes being of the level 1 kind. Of course, we would not normally explain the resulting action by citing these metacognitive level 1 attitudes; we would simply cite the attitudes involved in our conscious personal reasoning. And this explanation would not be wrong, since those attitudes did play an important role; but it would be underpinned by another, more basic, one, citing level 1 attitudes. This view again harmonizes with work by dual-system theorists, several of whom argue that conscious reasoning has only an indirect effect on behaviour, mediated by nonconscious reasoning processes (see Carruthers 2006; Evans forthcoming). Thus, if level 2 beliefs are acceptances, there is no conflict between the claim that all intentional actions have level 1 explanations and the commonsense view that level 2 attitudes influence behaviour; level 2 beliefs influence behaviour *in virtue of* our level 1 beliefs and desires about them. This dependency makes the terminology of levels particularly appropriate.

For all this, we cannot simply identify level 2 belief with acceptance. For as writers on the subject stress, acceptance possesses properties that are alien to belief. Two in particular stand out. First, acceptance is responsive to prudential considerations — professional, ethical, religious, and so on. For example, a lawyer may accept that their client is innocent for the purposes of defending them, even though they do not believe it. Second, acceptance can be context-relative; we can accept something for reasoning in certain situations but not in others, as the lawyer does. Now, the first contrast here is not so clear. It is arguable that beliefs *can* be formed for pragmatic reasons, provided one also has good evidence for their truth (see my 2007). It is true, however, that belief formation is typically *sensitive* to evidential considerations in a way that acceptance is not. The second contrast is more straightforward: belief, as we commonly conceive of it, plays an open-ended role in deliberation. (This is connected with the first contrast; if rational, one will not be willing to rely on a proposition in an open-ended range of deliberations unless one has good evidence for its truth.) However, it remains possible that level 2 beliefs are a subset of acceptances, and this is

the view I want to propose. Specifically, I suggest that level 2 beliefs are acceptances that are not restricted to particular contexts, and which are, therefore, typically evidence-sensitive. Again, there is more to be said in defence of this view (see my 2004, ch.5), but I propose to adopt it as a working hypothesis and move on to consider how delusions might fit into this framework.

4. Delusions and acceptances

How does the hypothesis that delusions are level 2 beliefs fare on this model? Are delusions unrestricted acceptances? I think this is a possible view, though we still need to explain their relative inertia. In fact, an explanation of this falls out naturally from the proposed account of how level 2 beliefs influence action. For we should expect the mechanism involved to break down in the case of delusions. Let me explain.

I suggested that our motivation for acting on our acceptances is essentially *metacognitive* — we act on them because we have a level 1 desire to adhere to our premising policies. But in some cases other level 1 desires may outweigh that desire. We may consciously judge that, given our premises and goals, we ought to perform action *A*, and so become motivated to perform it by our level 1 desire to adhere to our premising policies. Yet we may nonetheless refrain from performing *A* because we have a stronger level 1 desire not to do *A* or to do something else. For example, I might conclude, on the basis of personal reasoning involving my chosen premises and goals, that I should refuse another slice of chocolate cake, yet take one anyway since my level 1 desire for chocolate cake is stronger than my level 1 desire to act in line with my premises and goals.

I have argued elsewhere that this is what occurs in cases of akrasia (Frankish 2004, ch.8), and something similar may happen with delusions. If we have — for whatever reason — adopted a premise which tends to dictate extreme or highly unusual actions, then our desire to adhere to the policy may be overridden by other level 1 desires making for caution and conformity. So, for example, a Capgras patient may conclude, from the premise that their spouse has been replaced by an impostor, that they should shun their current partner and search for their true spouse. Yet they may fail to follow this course since their level 1 desire to act on their premises is weaker than their level 1 desire to remain with their current partner (perhaps because they have a strong level 1 belief that they are *not* in fact an impostor). (It is worth stressing that the claim here is not that level 2 attitudes are in *direct* competition with level 1 attitudes; if they belong to different systems this will not be the case. Rather, the conflict is between level 1 desires of different *orders*: a metacognitive desire to act on one's level 2 beliefs and a first-order desire to do something else.) Likewise, a Cotard patient has a level 2 belief (that is, a premise) which dictates that there is no point in their seeking food, shelter, companionship, and so on; but their level 1 desires for these things may outweigh their level 1 desire to act on their premises, rendering this particular level 2 belief largely inert. In such cases a level 2 belief might continue to influence what a person *says* without affecting their non-verbal behaviour at all. In a similar way, level 1

desires may inhibit a person from using a premise freely in their personal reasoning, rendering it cognitively as well as behaviourally inert. This might happen, for example, because they fear the conclusions the premise may warrant, or because they are concerned that their adherence to it may be undermined by the discovery of inconsistencies with other important premises.

One option, then, is that delusions are atypical level 2 beliefs, whose atypicality is explicable in terms of their unusual content. Note that this view involves no specific commitments as to why delusional beliefs are formed and sustained, and it is compatible with most existing versions of the doxastic conception. It is, for example, neutral on the questions of whether delusions are grounded in abnormal experiences, whether they are the product of abnormal fixation processes, and whether they involve reasoning deficits. It does, however, require that these questions be recast as ones about *personal activities* — forming and executing premising policies — rather than subpersonal processes. In assigning a key role to metacognitive attitudes, the account also has commonalities with metacognitive accounts of delusions, such as that proposed by Currie.

This is not all, however. The proposed two-level framework also offers another option for conceptualizing delusions. This is that delusions fall into the class of acceptances that are not beliefs — *non-doxastic acceptances*. This would also explain our competing intuitions about them. On the one hand, non-doxastic acceptances are very like beliefs; they possess the core properties of propositional content, mind-world fit, and an action-guiding role (in certain contexts), together with all the distinctive properties of level 2 states (conscious, binary, etc.). And, like beliefs, we would — in the right context — avow them and defend them (they are our premises, to which we are committed). On the other hand, non-doxastic acceptances lack the sensitivity to evidence that is characteristic of belief (and characteristically absent in delusions), are effective only in certain restricted contexts, and may not be consistent with each other. (We can accept something for reasoning in one set of circumstances, and something quite different for reasoning in another.) All of this fits the profile of delusions very well.

We have, then, two models of delusions, one doxastic and other non-doxastic. On the former, delusions are atypical doxastic acceptances (level 2 beliefs), on the latter, they are typical non-doxastic ones. The models are broadly similar, and the differences between them concern motivation and sensitivity to evidence. Doxastic acceptances are typically formed for epistemic reasons and are sensitive to evidence, whereas non-doxastic acceptances are formed for pragmatic reasons and without reference to evidence. The distinction between the two models thus corresponds roughly to that between bottom-up and top-down models of delusions (Bayne and Pacherie 2004a). (On bottom-up models delusional beliefs are grounded in experience, whereas on top-down ones, they are the product of central, non-perceptual processes of some kind.) Indeed, non-doxastic acceptances are akin to Campbell's framework propositions, which structure reasoning but are immune from empirical scrutiny.

Of course, we may need both models; some delusions may belong to one class and some to the other, and some may shift from one category to the other over time. Moreover, from a theoretical perspective the similarities between the models are more important than their differences; the key claim on both views is that delusions are acceptances. It is this inclusive position that I propose, and I shall conclude by looking at two important implications of it.

The first implication is that the search for the explanation of particular delusions should take a different route. If the present hypothesis is correct, then delusions are premising policies, whose adoption is motivated by level 1 desires. So the question is not simply 'Why is the patient in this mental state?', but 'Why are they doing this? What are their motives for pursuing this policy?' The *immediate* explanation will lie, not in abnormal psychological or neurological processes, but in the level 1 desire that motivates the pursuit of the policy. Of course, this desire may itself be the product of abnormal processes of some kind, but to skip straight to these would be to miss an important explanatory step. (I should stress that in claiming that delusions are the product of nonconscious mental states and processes I am not endorsing a traditional psychodynamic theory. The level 1 mind is not the Freudian dynamic unconscious, but the modern cognitive unconscious; see, e.g., Hassin et al. 2005; Wilson 2002.)

The motives for forming delusional beliefs might be epistemic ones — desires to adopt true premises for use in personal reasoning. An agent might accept a delusional claim because it reflects or explains the content of their experience, as on bottom-up approaches. Equally, however, the motives might be pragmatic, and in non-doxastic cases they typically will be. The nature of these motives is a matter for empirical investigation, but as an illustration it may help to consider self-deception, which can be regarded as a type of mild, non-psychotic delusion. I have argued elsewhere that self-deception involves a form of general acceptance, borderline between doxastic and non-doxastic, which is motivated by a desire to shield oneself from disturbing truths (Frankish 2004, ch.8). The self-deceiver finds a certain view probable but highly disturbing, and has a strong level 1 desire to avoid consciously accepting its truth. This desire leads them to accept that the view is in fact false, thereby ending conscious deliberation on the matter and committing themselves to a view they find comforting. Psychotic delusions, too, may be formed and sustained in response to emotional influences of various kinds, perhaps pathological in origin. Even in cases where epistemic motives are operative, it is likely that emotional factors will have to be invoked to explain the unusual fixity of delusions, which on the proposed view will be a consequence of the strength of the supporting level 1 desires.

The second implication is that it will be possible to bring delusional patients within the interpretative fold, at least at level 1. Irrational and inconsistent level 2 attitudes can be rational manifestations of consistent level 1 attitudes. (Indeed, if we adopt a dispositionalist view of level 1, as I have proposed, then the attribution of the supporting level 1 attitudes will require an assumption of rationality at that level.) This may seem an unattractive consequence, given the notorious difficulties in understanding delusional patients. However, the sense in which delusions are

understandable will not be the usual one. The claim is not that the actions of deluded patients are understandable as manifestations of their delusions qua level 2 beliefs; all the problems for that view remain. Rather, it is that they are understandable qua manifestations of the premising policies in which those states consist, and interpreting a deluded patient will involve forming and testing hypotheses as to the character of these policies. And this may be a difficult task. Much of the activity involved in executing a premising policy is covert; it consists in pursuing a certain strategy in one's conscious personal reasoning, and although such reasoning is (I would maintain) squarely intentional, it is rarely publicly observable. Interpretivist theory usually ignores such covert behaviour, but if the present view is correct it will be crucial to take it into account if we are to understand delusions (for defence of the claim that interpretivists can legitimately recognize the existence of covert behaviour, see my 2004, ch.5). Fresh approaches may therefore be needed in order to uncover the relevant evidence. For example, 'talk-aloud' and 'think-aloud' protocols might be employed, in which subjects verbalize or explain their thought processes.

Conclusion

This chapter has offered a new hypothesis about the nature of delusions, which locates them within an expanded version of our commonsense mental taxonomy. I sketched a two-level view of the mind, supported by both analysis of folk psychology and experimental work on reasoning, and argued that delusions belong to the second level. I then outlined a model of level 2 belief based around the concept of acceptance developed by some philosophers of science, and proposed that delusions are acceptances — some doxastic, some non-doxastic. I argued that this view can reconcile our competing intuitions about delusions and explain their puzzling features. This does not, of course, amount to a conclusive case for the view, but I hope it establishes it as a serious hypothesis. Its fate will, of course, depend on whether it proves theoretically fruitful and clinically valuable.

Acknowledgements

I thank Lisa Bortolotti, Matthew Broome, and Maria Kasmirli for their comments on an earlier draft of this chapter. Work on the chapter was supported by a research leave award from the UK's Arts and Humanities Research Council.

References

American Psychiatric Association (2000). *Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition, Text Revision*. Washington D.C.: American Psychiatric Association Pub. Inc.

Bayne, T. and Pacherie, E. (2004a). Bottom-up or top-down? Campbell's rationalist account of monothematic delusions. *Philosophy, Psychiatry, & Psychology*, 11, 1-11.

Bayne, T. and Pacherie, E. (2004b). Experience, belief, and the interpretive fold. *Philosophy, Psychiatry, & Psychology*, 11, 81-6.

Bayne, T. and Pacherie, E. (2005). In defence of the doxastic conception of delusions. *Mind and Language*, 20, 163-88.

Berrios, G.E. (1991). Delusions as 'wrong beliefs': A conceptual history. *British Journal of Psychiatry*, 159, 6-13.

Bortolotti, L. (2005). Delusions and the background of rationality. *Mind and Language*, 20, 189-208.

Bovet, P. and Parnas, J. (1993). Schizophrenic delusions: A phenomenological approach. *Schizophrenia Bulletin*, 19, 579-97.

Breen, N., Caine, D., Coltheart, M., Hendy, J., and Roberts, C. (2000). Towards an understanding of delusions of misidentification: Four case studies. *Mind and Language*, 15, 74-110.

Brett-Jones, J., Garety, P., and Hemsley, D. (1987). Measuring delusional experiences: A method and its application. *British Journal of Clinical Psychology*, 26, 257-65.

Buchanan, A. and Wessely, S. (1998). Delusions, action, and insight, in X.F. Amador and A.S. David (eds.) *Insight and Psychosis*, pp. 241-68 (New York: Oxford University Press).

Campbell, J. (2001). Rationality, meaning, and the analysis of delusion. *Philosophy, Psychiatry, & Psychology*, 8, 89-100.

Carruthers, P. (2006). *The Architecture of the Mind*. Oxford: Oxford University Press.

Cohen, L.J. (1992). *An Essay on Belief and Acceptance*. Oxford: Oxford University Press.

Currie, G. (2000). Imagination, delusion and hallucinations. *Mind and Language*, 15, 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: Oxford University Press.

Davidson, D. (1984). *Inquiries into Truth and Interpretation*. Oxford: Oxford University Press.

Davies, M., Coltheart, M., Langdon, R., and Breen, N. (2001). Monothematic delusions: Towards a two-factor account. *Philosophy, Psychiatry, & Psychology*, 8, 133-58.

Dennett, D.C. (1978). Brainstorms: Philosophical Essays on Mind and Psychology. Montgomery, Vt.: Bradford Books.

Dennett, D.C. (1987). *The Intentional Stance*. Cambridge, Mass.: MIT Press,.

Dennett, D.C. (1991). *Consciousness Explained*. Boston: Little Brown and Co.

Engel, P. (2000). Introduction: The varieties of belief and acceptance, in P. Engel (ed.) *Believing and Accepting*, pp. 1-30 (Dordrecht: Kluwer).

Evans, J.S.B.T. (2003). In two minds: Dual-process accounts of reasoning. *TRENDS in Cognitive Sciences*, 7, 454-9.

Evans, J.S.B.T. (forthcoming). How many dual process theories do we need? One, two or many?, in J.S.B.T. Evans and K. Frankish (eds.) *In Two Minds: Dual Processes and Beyond* (Oxford: Oxford University Press).

Evans, J.S.B.T. and Frankish, K. (forthcoming). *In Two Minds: Dual Processes and Beyond*. Oxford: Oxford University Press.

Evans, J.S.B.T. and Over, D.E. (1996). *Rationality and Reasoning*. Hove: Psychology Press.

Fodor, J.A. (1987). Psychosemantics: The Problem of Meaning in the Philosophy of Mind. Cambridge, Mass.: MIT Press.

Frankish, K. (1998). Natural language and virtual belief, in P. Carruthers and J. Boucher (eds.) *Language and Thought: Interdisciplinary Themes*, pp. 248-69 (Cambridge: Cambridge University Press).

Frankish, K. (2004). *Mind and Supermind*. Cambridge: Cambridge University Press.

Frankish, K. (2007). Deciding to believe again. *Mind*, 116, 523-47.

Frankish, K. (forthcoming(a)). Partial belief and flat-out belief, in F. Huber and C. Schmidt-Petri (eds.) *Degrees of Belief: An Anthology* (Dordrecht: Springer).

Frankish, K. (forthcoming(b)). Systems and levels: Dual-system theories and the personal-subpersonal distinction, in J.S.B.T. Evans and K. Frankish (eds.) *In Two Minds: Dual Processes and Beyond* (Oxford: Oxford University Press).

Gazzaniga, M.S. (1998). *The Mind's Past*. Berkeley: University of California Press.

Hamilton, A. (2007). Against the belief model of delusion, in M.C. Chung, K.W.M. Fulford, and G. Graham (eds.) *Reconceiving Schizophrenia*, pp. 217-34 (Oxford: Oxford University Press).

Hassin, R.R., Uleman, J.S., and Bargh, J.A. (2005). *The New Unconscious*. New York: Oxford University Press.

James, W. (1890). *The Principles of Psychology*. New York: Henry Holt and Co.

Jaspers, K. (1913/1997). *General Psychopathology: Volume 1* (J. Hoenig and M.W. Hamilton, Trans.). Baltimore: Johns Hopkins University Press.

Kahneman, D. and Frederick, S. (2002). Representativeness revisited: Attribute substitution in intuitive judgement, in T. Gilovich, D. Griffin, and D. Kahneman (eds.) *Heuristics and Biases: The Psychology of Intuitive Judgment*, pp. 49-81 (Cambridge: Cambridge University Press).

Lewis, D. (1972). Psychophysical and theoretical identifications. *Australasian Journal of Philosophy*, 50, 249-58.

Maher, B.A. (1999). Anomalous experience in everyday life: Its significance for psychopathology. *The Monist*, 82, 547-70.

Oaksford, M. and Chater, N. (2007). *Bayesian Rationality: The Probabilistic Approach to Human Reasoning*. Oxford: Oxford University Press.

Ramsey, W., Stich, S.P., and Garon, J. (1990). Connectionism, eliminativism and the future of folk psychology, in J.E. Tomberlin (ed.) *Philosophical Perspectives, 4: Action Theory and Philosophy of Mind*, pp. 499-533 (Atascadero, CA.: Ridgeview Publishing Company).

Ryle, G. (1949). *The Concept of Mind*. London: Hutchinson.

Sass, L.A. (1994). The Paradoxes of Delusion: Wittgenstein, Schreber, and the Schizophrenic Mind. Ithaca, N.Y.: Cornell University Press.

Sloman, S.A. (1996). The empirical case for two systems of reasoning. *Psychological Bulletin*, 119, 3-22.

Smith, E.R. and DeCoster, J. (2000). Dual-process models in social and cognitive psychology: Conceptual integration and links to underlying memory systems. *Personality and Social Psychology Review*, 4, 108-31.

Stanovich, K.E. (2004). *The Robot's Rebellion: Finding Meaning in the Age of Darwin*. Chicago: The University of Chicago Press.

Stone, T. and Young, A.W. (1997). Delusions and brain injury: The philosophy and psychology of belief. *Mind and Language*, 12, 327-64.

Wegner, D.M. (2002). *The Illusion of Conscious Will*. Cambridge Mass.: MIT Press.

Wilson, T.D. (2002). *Strangers to Ourselves.* Cambridge, Mass.: Belknap Press.

Wilson, T.D., Lindsey, S., and Schooler, T.Y. (2000). A model of dual attitudes. *Psychological Review*, 107, 101-26.

Young, A.W. (2000). Wondrous strange: The neuropsychology of abnormal beliefs. *Mind and Language*, 15, 47-73.