

The Limits of the Doxastic¹

Tim Crane & Katalin Farkas

Central European University

We distinguish between beliefs, the paradigm doxastic state, and the conscious episodes in which we acknowledge, judge or express our beliefs. Beliefs are mental states that govern our actions and are appropriately related to their conscious manifestations. When things go well, there is a kind of harmony between the underlying unconscious state and its conscious manifestations. What we consciously acknowledge or judge conforms to how we behave, and our underlying dispositions to behave and speak change as our interaction with the world changes.

But what about when things do not go well? There are many well-known cases where what people explicitly profess does not match their actual behaviour, which can be grouped into various kinds: cases of self-deception (Bach 1981, McLaughlin and Rorty 1988), prejudice (Munton forthcoming), bias (Brownstein and Saul 2016), as well as simple confusions. What should we say about what subjects 'really believe' when their unconscious dispositions come apart from their conscious judgements or other manifestations? These cases have been widely discussed in the recent philosophical literature, and various approaches have been popular: some philosophers identify the real belief with the underlying disposition and place less importance on the conscious manifestation (e.g Schwitzgebel 2010); others dispute the identification of the belief with the disposition and identify beliefs with stored representations, explaining the disparities through mechanisms of 'fragmentation' (Quilty Dunn & Mandelbaum 2017). Yet others distinguish between what is genuinely believed and some other doxastic-like unconscious state (e.g. 'alief' Gendler 2007).

These approaches have something in common: they assume that the disparities in question are explained by a psychological structure which corresponds in a more or less direct way with the structure of conscious judgement or assertion. It is a commonplace that we explain behaviour by explicitly appealing to beliefs and desires; the assumption behind the views just mentioned is that something like this structure (doxastic/conative states) also exists at the level of the unconscious states which these views are invoking.

In this paper we challenge this assumption. Drawing on work by Crane (2017), we argue that the underlying reality of the unconscious is less determinate than many

¹ We are grateful to support from the CEU Intellectual Themes Initiative.

philosophers today assume. Crane argued that we should think of the the totality of the thinker's doxastic orientation towards the world (what he called the 'Worldview') as having a dispositional structure, but not a structure that corresponds in any direct way with the beliefs we attribute in everyday propositional attitude attributions. There are not individual dispositions corresponding to individual belief attributions. Rather, propositional attitude attributions — e.g. explanations of action in terms of beliefs and desires — should be regarded as models of this underlying structure, which idealise and simplify in certain respects.

We adopt the basic structure of this position, but here we aim to push it further: we keep the term 'Worldview', to apply to the totality of the subject's unconscious psychological dispositions. But we modify it by not restricting this totality to the doxastic. The idea that the unconscious mind must have a structure corresponding to the belief/desire model is something that we here bring into question. Specifically, in this paper we argue that the best way to make sense of many of the cases of mismatch between explicit judgements and patterns of behaviour is to reject the picture that divides the unconscious mind into a doxastic and conative 'areas' (in the well-known Schiffer/Fodor terminology, belief and desire 'boxes'). In particular, we should resist attempts to try and explain all these mismatches in terms of belief, or to use them as crucial evidence in the debate between different views of belief. The truth is rather that many of these cases are better understood without appealing to belief at all — we need to appreciate the limits of the doxastic.

In the first section we introduce the basic distinctions needed to understand these phenomena: between unconscious standing states, dispositions and conscious occurrences (events and processes), and we outline some uncontroversial views about how these fit together in the normal cases. In the second section we consider ways in which the conscious and unconscious phenomena can come apart, and we outline and criticise the popular doxastic and quasi-doxastic approaches to these phenomena, and offer some alternatives. In the final section we sketch our a broader picture of the mind that these alternative explanations suggest.

1. Unconscious standing states and the stream of consciousness

We start with a distinction that is commonly drawn between the two fundamentally different kinds of doxastic attitudes. The first is often called 'belief', and it is understood as a 'standing' state, a state of mind that persists through changes in your conscious awareness, and even in the absence of consciousness. It is a state rather than an

occurrence or event — a belief in this sense is best thought of as a condition or property of a believer, rather than an event or process (something that occupies time by having temporal parts). So if you are like us, you believe that $2+2=4$, you have believed this at least since elementary school, and you very rarely if at all bring this belief to consciousness. This is an example of a persisting or standing belief state. Standing states in this sense are primarily or totally unconscious (Crane 2013).

The second doxastic attitude is the conscious acknowledgement of the truth of something, considered as an episode in the stream of consciousness. As we shall see, sometimes this is also called belief, others use the words 'judgement' or 'assent' instead. When you consciously entertain the proposition that $2+2=4$, and you accept it as true, then this is judgement or assent or acceptance. This, unlike a standing state, is an event, something which takes (maybe a very short amount of) time, and does not itself persist beyond the moment of consciousness.

(The word 'judgement' is not quite right since it can suggest only the act of forming a belief, or coming to an opinion. The type of event we have in mind here includes the formation of belief, but it also includes bringing to consciousness what you already believe. We will talk of acknowledgement and judgement interchangeably, but we ask the reader to bear in mind that both types of mental act are included under these terms.)

What is the nature of these two kinds of doxastic phenomena? How should we analyse or classify them? And what is the relationship between them, in the most general terms? We will sketch two broadly contrasting approaches.

The first is Hume's: he applies the name "belief" primarily to the conscious episode, and he treats it as explanatorily prior to the unconscious phenomenon. Hume argued famously that beliefs are not distinguished from other conscious mental acts by their contents ('ideas') but by the attitude involved, and this attitude is a conscious 'feeling':

its true and proper name is belief, which is a term that every one sufficiently understands in common life. And in philosophy we can go no farther, than assert, that it is something felt by the mind, which distinguishes the ideas of the judgment from the fictions of the imagination. It gives them more force and influence; makes them appear of greater importance; infixes them in the mind; and renders them the governing principles of all our actions. (Hume 1739-40/1978: I.3.7)

Hume's focus is the occurrent episode of accepting something as true, and he says little about the persisting state. Still, we can get a glimpse of his picture from the quote above:

'infixing' the belief in the mind is plausibly understood as the inception of a standing state. Being in this standing state seems to mean only that the occurrent episode accompanied by the feeling will arise again. Further facts about belief— for example, that beliefs are the 'governing principles of all our actions' — are explained by the nature of the occurrent episode.

The Humean theory that beliefs are distinguished from other attitudes by a conscious feeling is sometimes contrasted with the second approach we will consider here: that beliefs are the basis of dispositions to act. For example, H.H. Price in his book *Belief* (1969) presents the 'Occurrence Analysis' and the 'Dispositional Analysis' as the two main accounts of belief. The dispositional analysis can be traced in recent centuries at least as far back as Alexander Bain:

The difference between mere conceiving or imagining, with or without strong feeling, and belief, is acting, or being prepared to act, when the occasion arises. The belief that a sovereign is worth twenty shillings, is shown by the readiness to take the sovereign in exchange for the shillings; the belief that a sovereign is light is shown by refusing to take it as the equivalent of twenty shillings. (Bain 1872: 372)

Note that Bain's focus in this passage seems to be the occurrent episode, rather than the standing state. It is the occurrent episode of accepting something as true that is contrasted with 'mere conceiving' — after all, there is no standing state of 'mere conceiving'. Bain's view is that it matters little how we feel when we entertain a proposition; what matters is what we do, or what we are prepared to do next. As he makes clear, a belief alone does not entail (an inclination for) performing a particular action, because the person's other mental states also matter. For example: 'one may have the conviction strongly that abstinence from stimulants would favour health and happiness, and yet go on taking stimulants' (Bain 1872: 372). In this case, a desire for a present pleasure or for a relief of a present pain may be a stronger inducement for action, and doesn't mean that the person lacks the belief in question.

In contrast to the 18-19th century discussion, the contemporary debate on beliefs usually focuses on standing states. Here is how the standard reference work, the *Stanford Encyclopedia of Philosophy* introduces the concept of belief:

To believe something, in this sense, needn't involve actively reflecting on it: of the vast number of things ordinary adults believe, only a few can be at the fore of the mind at any single time. (Schwitzgebel 2019)

Once we have the dispositional analysis, we have something that neatly applies also to standing states. Standing states cannot be beliefs in virtue of a conscious feeling, since they don't involve consciousness. But they can very well be the basis of dispositions to act.

Arguably, all theories of beliefs as standing states all make the possession of appropriate dispositions a necessary condition for having a belief. This can be a bit obscured the terminology: for example, in his *Stanford Encyclopedia* article, Eric Schwitzgebel distinguishes dispositionalism (his own view) from representationalist, functionalist and interpretationist views of beliefs. Dispositionalists hold that 'to believe that P is nothing more than to match to an appropriate degree and in appropriate respects the dispositional stereotype for believing that P' (Schwitzgebel 2002: 253). The other views will appeal to other notions in addition to dispositions in their characterization of belief, but dispositions will still play a central role.

For example, on a functionalist analysis, the nature of a mental state is determined by its functional role: its typical causes and typical effects. The typical effects of the belief need not be actual effects — in fact they are the same as, or they align with, the dispositions associated with the belief. As Schwitzgebel remarks, a dispositionalist can be regarded as a functionalist who places the emphasis on forward-looking functional roles.

Consider next those who are influenced by Fodor's representational theory of the mind, and treat beliefs as inner mental representations ('sentences in a language of thought') which are unconscious and state-like (see Quilty-Dunn and Mandelbaum 2018). Fodor himself accepted a functionalist account of the attitudes: what makes a belief that *p* a *belief*, rather than a desire, is that it is placed in the 'belief-box', whose nature is characterized in terms of a functional role. We have already seen that functionalists share with dispositionalists the view that forward-looking functional role is essential to a mental state. What distinguishes representationalists from mere dispositionalists is that in addition to the relevant dispositions, they impose certain requirements on the mechanism that realizes those dispositions.

It's worth emphasizing two features of the dispositional stereotype for a given belief. First, the stereotype always includes references to other mental states; for example, the same belief *p* may dispose us to act differently in the presence of different desires (as in Bain's example about believing that strong stimulants are bad for one's health). Thee

second feature is that the dispositions can be understood in a somewhat liberal way, to include dispositions to feel something or to have certain conscious episodes: consciously acknowledging the truth of p can be counted as one of the manifestations of the dispositions associated with p .

Hence we see two broadly contrasting approaches. One regards the conscious episode (acknowledging the truth of p) as primary, and captures its essence in its conscious character. This does not have to be a “feeling” in Hume’s sense, but it could be a phenomenal feature in a more general sense, or another feature specific to conscious episodes. The standing state is understood as a mere imprint of the conscious episode, though it is worth noting that an element of dispositionalism remains here too: the standing state could be regarded as a disposition to have the conscious occurrence. The other approach regards the standing state of believing that p as primary, and regards it as the basis of some dispositions (plus possible additional factors: an internal representation, or backward-looking functional role). And one important manifestation of this dispositional state may be the conscious acknowledgment of the truth of p .

This is a very high altitude picture of some complex terrain, and we do not claim, of course, that these two broad approaches are the only ways to understand the relationship between standing states and conscious occurrences. There are other views. For example, many writers talk freely about belief itself being conscious and unconscious (e.g. Mellor 1978, Kriegel 2014), though in many cases ‘conscious belief’ means simply the episode of being conscious of what you believe (Crane 2013).

However, what these other approaches share with the two approaches we have identified is the idea that there is an unconscious standing state that deserves to be called ‘the belief that p ’ and that this corresponds — in some way — to the conscious acknowledgement of the truth of p . In what follows, we will question this idea. There are conscious episodes of acknowledging the truth of a statement; these do leave an imprint on the unconscious landscape, and in order to make sense of others and ourselves, we can conveniently call them ‘beliefs’. We will agree with dispositionalists that our unconscious mental life — the totality of our standing states — is to be characterized in terms of dispositions; but we will deny that these states can always be compartmentalized into anything that answers to the usual conception of belief (or desire, or any of the other familiar standing states).

We will support this idea by considering some characteristic relationships between standing states, conscious thinking, and behaviour. In much of our mental life, conscious thinking, unconscious standing states and behaviour tend to align quite well; we

acknowledge the truth of a statement like $2+2=4$, and our actions and reactions conform to the dispositional stereotype associated with believing the same. But in many cases, these things can come apart. In the next section we will consider some examples.

2. Actions and professed beliefs

The cases we will discuss are often thought to involve a discrepancy or mismatch between what we believe or profess to believe, and what we do. There are a number of further types of these cases, but the following two categories will be sufficient to make our point.

i) *Implicit bias*. There has been much discussion recently of real and imagined cases of implicit bias and/or prejudice, which often have the form of someone who denies that they are prejudiced (e.g. against women, or against black people) in any way, is well-informed about the facts and the theories, and yet behaves in a way that is contrary to this. Schwitzgebel's well-known story of Juliet, a white college professor who explicitly expresses anti-racist views yet behaves in ways that suggest some racist attitudes, will serve as a useful example below.

(ii) *Emotional responses*. In a well-known paper, Tamar Gendler (2007) introduced a cluster of real-life cases which exhibit a certain pattern in relation to belief and action. For example: many people show signs of terror when trying to walk over something they know to be metre-thick glass, positioned a few hundred metres over a sheer drop in the Grand Canyon, even though they believe it's perfectly safe. Or: some people are revolted by the idea of eating from a newly bought bed-pan, or refuse to eat fudge shaped like dog feces, even though they are in both cases convinced that the food is perfectly clean and harmless.

These cases are often described as if they pointed to a tension between the conscious episode of judging and the standing state of believing. Indeed, one of Schwitzgebel's papers on this topic has the title "Acting Contrary to Our Professed Beliefs, or *The Gulf Between Occurrent Judgment and Dispositional Belief*" (Schwitzgebel 2010, emphasis added). The idea is this: there is something that our protagonists consciously acknowledge as true: Juliet acknowledges that the races are intellectually equal, and the visitor in the Grand Canyon acknowledges that the walkway is entirely safe to walk on. These are their occurrent conscious judgements. Now consider the corresponding standing states, the beliefs that are supposed to be the basis of dispositions to act. First thought: if they had the belief with the same content as their judgement, they would act differently. Their actions don't match the dispositional profile of the corresponding belief.

Second thought: in fact, the actions suggest a belief with a different content. Their actions match the dispositional profile of a belief with the contrary content.

Hence the idea that the above cases illustrate a mismatch between two doxastic states: the standing state and the occurrent episode. A common feature of many discussions of these cases is that they limit their attention to the doxastic features of the situation (see, for example, Keith Frankish's (2016) discussion of the Juliet case). Schwitzgebel's (2010) classification of responses to cases like these is typical. Where P is the proposition the attitude to which is in question ('women are inferior' etc.), Schwitzgebel distinguishes between the judgement view (in our terms: judging P but not believing P, attributed to Gendler 2008 and Zimmerman 2007) the anti-judgment view (not-believing that P and instead believing not-P (Hunter 2009)), the shifting view (shifting between believing P and believing not-P (Rowbottom 2007)), and the contradictory belief view (believing both P and not-P (Gertler 2011)).

But explanations of actions only in terms of doxastic attitudes are mostly incomplete. This is the familiar point mentioned above, made (among many others) by Bain when he claims that someone could believe that strong stimulants are to be avoided and yet take strong stimulants. Our actions are influenced not just by our hypothetical beliefs, but also by the rest of our mental life. If we think that what is going on here must be understood predominantly in terms of doxastic attitudes, then we will miss a lot of the psychological complexity by forcing the explanation in one direction or another: belief or judgement.

For example, someone who betrays an implicit bias against women in a job search, need not do so because they believe women are inferior. In fact, they may be fully convinced, both consciously and dispositionally, that women and men have the same intellectual abilities. But they may rather have different *emotional or affective* reactions to women and men — they may feel more comfortable in the presence of men, or they may prefer the company of men in various ways. To be sure this preference can be criticised, but it need not involve any belief in the inferiority of women.

Similarly, a racist person's behaviour need not spring from the belief that black people are inferior, but from taste or preference: they might just prefer hanging out with white people, and may well discriminate against them because of this preference. This kind of racist could well have the same beliefs about the superiority or inferiority of black people as a non-racist, but their acts of discrimination towards black people are best explained in terms of their tastes and likings rather than their beliefs. Suppose Juliet chooses only white students for her special meetings after class, because she finds them

more appealing or attractive — this attitude is objectionable, and leads to unjust discrimination, but it need not arise from a belief in the inferiority of blacks. (This is not supposed to be an analysis of what it is to be racist, or what racism means, or why it is evil – it is only supposed to be an empirical speculation about what might actually move somebody in a given case.)

The case of Juliet is a an invented one of course, and Schwitzgebel is free to describe it as he likes. But many real-life cases are so relevantly similar to this case that it is important that ‘belief/non-belief’ does not become our only philosophical model for thinking about bias and prejudice. A point from Brownjohn and Saul (2016) is worth noting here:

in psychology, attitudes are understood as likings or dislikings, or, more formally, as associations between a concept and an evaluation ... This conceptualization of attitudes is importantly different from the typical usage in philosophy, which is much more expansive (including beliefs, desires, intentions, and so on).

In other words, if we use the word ‘attitude’ as a label for the source of our biases and prejudices, then the relevant attitudes should not just comprise beliefs. Munton (2020) acknowledges too that prejudice might involve an affective element, while going on herself to give an account in terms of misperceived salience.

Consider now the cases in our group (ii): emotional reactions which are supposed to be in conflict with our professed beliefs. Why should someone’s refusal to eat fudge that looks like dog feces even raise the suggestion that they believe the fudge is filthy? We can easily find something disgusting even if we don’t believe it’s bad for us. There are many visceral reactions which are independent from our knowledge of what is and isn’t harmful. Gendler, in her discussion of these cases, fully acknowledges this point: she mentions the affective patterns that are activated by certain visual impressions (eg. something that looks like feces). Yet she thinks that something else needs to be added to explain the case. She thinks we need to hypothesise a previously unnoticed mental state to explain the disparities between belief and behaviour. This hypothesised attitude is “alief”, which is spontaneous, not sensitive to reason, and whose effect on action is immediate and hard to modify. Although you do not believe that the glass in the Grand Canyon walkway will collapse, you do ‘alieve’ it. You do not believe that the sterilised toilet bowl is dirty, but you do ‘alieve’ it.

But do we need to do this? Can't we use existing concepts to explain what's going on? If so, which concepts should they be? If we take Gendler's examples one at a time, it is easy to find resources from our commonsense psychology to account for these cases. We feel some of the bodily effects characteristic of fear when approaching the glass pavement; we feel disgust at the thought of what was in the toilet bowl; and so on. True: fear, disgust and some other emotional/affective reactions are invulnerable to revision in the light of reasoning and belief. But this is a familiar point which does not need the invention of a new psychological category.

Gendler's theory is a mixture of the usual problem that we detect in these discussions, and the way we propose to overcome this problem. We see it problematic that many discussions focus on the doxastic aspect of these cases, and we propose to overcome this problem by asking to appreciate the non-doxastic factors. Although Gendler's explanations do not appeal only to doxastic attitudes to explain the mismatches, it is surely significant that the word she coins 'alief', is close to the word 'belief'. It is as if she is looking for something which is close to belief without actually being belief.

Schwitzgebel has a different response. He treats 'believes that P' as a vague predicate admitting of borderline cases. In his discussion of cases similar to the ones discussed here, Schwitzgebel says that it might be best to describe me as being in a 'in-betweenish state' (Schwitzgebel 2010: 335). His reason for saying this is that the totality of a subject's dispositions (to behave and to bring about certain mental events) do not always determine whether the subject falls into the extension or the anti-extension of 'believes that p'. But there can be nothing more which will settle the question than the totality of these dispositions: 'once the dispositions are fully characterized the question of what the subject believes is closed' (Schwitzgebel 2002: 273).

For the reasons given in the next section, we agree with the spirit of this diagnosis: we agree with Schwitzgebel that there is nothing more to the subject's unconscious psychological organisation than the totality of their dispositions. In fact, we urge to go further than Schwitzgebel. His discussion seems to suggest that we can separate the subject *belief-relevant* dispositions, and then place subjects on a scale to measure their closeness or distance to the dispositional stereotype of a belief.

What happens then is that we too easily classify a piece of behaviour as discordant with a certain belief. Someone regularly choses men over women to invite to conferences or hire for jobs – that must be because they believe women are less smart! Not necessarily – they may just prefer men in some other ways. Their relevant dispositions are determined not only doxastically, but also affectively and conatively. However, and this is a crucial

point, we are not saying that thereby we can restore the attribution of a full-blown belief, now alongside with a full-blown preference and a consequent desire. Rather, we think that in many of these cases, something like “in-betweenness” will concern not only belief and disbelief, but also belief and other mental attitudes.

Returning now to Juliet’s case, the totality of her dispositions may lend itself to an explanation in terms of a belief in the equal intelligence of the races but a preference for white student – she is simply the kind of person whose evaluations are easily influenced by her likes and dislikes (we could try find further evidence for this in her non-race related evaluational biases). Alternatively, she may be a person who successfully separates evaluations from her likes and dislikes, and bases them strictly on intellectual merit, but she is in two minds about the intellectual merits of the races. We can experiment with these explanations and sometimes find one more helpful than the other; but there may not be anything in the totality of Juliet’s disposition that would settle which one is correct.

3. The limits of the doxastic

We have argued that a straightforward belief/non-belief view will not give an adequate explanation of everything that counts as a mismatch between professed beliefs and actions. On a more psychologically realistic view of the typical cases of mismatch, we need to consider how non-doxastic psychological factors influence the mismatch, one way or another — either by influencing the conscious manifestations, or by influencing the behaviour arising from the unconscious dispositions.

Our target is not the dispositional view of belief as such, but the belief-based understanding of the dispositions. Our question is: why think that this characterisation of the subject’s dispositions must always be a characterisation of what the subject *believes*? Schwitzgebel thinks his dispositional (and ‘in-between’) view of belief is the best way to account for the mismatch cases, as well as some other cases where we have no clear intuition about whether someone believes something (Schwitzgebel 2001). We have argued that dropping the requirement that the relevant psychological dispositions must be thought of wholly in terms of belief will leave room for a more psychologically realistic picture of the relationship between the dispositions and the relevant conscious occurrences.

What might this picture look like? In the final section of this paper we propose a development of a view of the unconscious which one of us has sketched elsewhere (Crane 2016). We call this the ‘Worldview theory’. The Worldview theory can be introduced by contrasting it with the dominant view of the relationship between the conscious and the

unconscious. This dominant view has been nicely described by John Searle (who opposes it too):

Perhaps at its most naive, our picture is something like this: unconscious mental states in the mind are like fish deep in the sea. The fish that we can't see underneath the surface have exactly the same shapes they have when they surface. The fish don't lose their shapes by going under water. Another simile: unconscious mental states are like objects stored in the dark attic of the mind. These objects have their shapes all along, even when you can't see them. We are tempted to smile at these simple models, but I think something like these pictures underlies our conception of unconscious mental states; and it is important to try to see what is right and what wrong about that conception. (Searle 1989: 195)

Searle's critique of this picture is based on his idea that the conscious mental states are the only fundamentally mental ones, and unconscious states have no genuine mental characterisation independently of their disposition to produce conscious states. We reject this view since we believe that the unconscious states are properly mental, but Searle's picture of the usual view is vivid and accurate.

The Worldview theory holds that the ways that the world is represented in the conscious and the unconscious mind are very different. In the conscious mind, the content of intentional attitudes is relatively determinate, attitudes are distinguished from one another more or less sharply, and the complexity of a content can be exhibited in its linguistic expression. The view is not that all conscious content is linguistically expressed, but that expression in language is one of the ways in which a more sharply defined shape can be put on the contents of our attitudes.

Imposing shape on our attitudes is what happens when we bring what is in our unconscious to consciousness. When we ask ourselves what we think about a certain subject, sometimes the answer comes quickly and easily: what is the German word for hope? It comes to mind: *Hoffnung*. This knowledge was there, I could easily retrieve it. My belief is made conscious and explicit. How this exactly happens is a difficult question, but that it can happen in this way should not be questioned.

But other cases are more difficult. Supposing someone (call her Sally) is trying to figure out what she really believes about, say, the moral status of animals. This is a complex question, and her attitude to it affects Sally's behaviour every day. On the one hand, Sally is aware of the way that dairy cows are treated, and she is certain that cows

are capable of suffering. On the other hand, she cannot bring themselves to accept that the life of a cow matters as the life of a person does, and she needs to buy milk for their children, and they enjoy eating cheese and other dairy products. When she reflects on this, she attempts to uncover some firm opinion which she can claim as her own. But she may find conflict and confusion, and puts the question to one side for consideration on another day. Sometimes she buys milk which has been described by the producer as having come from farms that treat animals well — but she never looks into the details. And sometimes she is too busy, or forgets, and doesn't bother.

But how do Sally's dispositions relate to what she believes? Does Sally believe that it is wrong to buy and consume dairy products? A simple action-based view might say no, since she clearly acts as if it were not wrong. But as we argued above, we also have to consider her tastes and preferences — maybe she considers that it is indeed wrong, but likes cheese and milk too much, or does not want to disturb her life's routines, or does not really care too much about animals (though she might care a little). Or it may be that she does believe it is wrong, but she also believes it is 'sort of OK' to eat dairy products, and she has not resolved the tension in her mind between these beliefs and the other things she believes about animals, suffering, pleasure, convenience and the right thing to do. In other words, maybe her total unconscious state and the actions based on it are a mixture of vague commitments, confusion, unclarity and weakness of the will.

We suspect that many people are like Sally when it comes to large and complex questions like this. And many of us try and resolve these complexities and confusions in our own minds. How do we — or how should we — go about doing this? When we reflect on our beliefs about such questions (which as many have pointed out, is often the same as reflecting on the questions themselves) it can sometimes seem as if we are finding out what we already believed about something. We can even be surprised to discover what we really believe. But in other cases it can also seem as if we are settling a question for ourselves, making up our mind.

It is natural to suppose that these are different things — finding out what you think, and making up your mind. The usual view of the relationship between the conscious and the unconscious — the one captured in Searle's image — maintains a sharp distinction between these things, because beliefs are unconscious states of mind which are 'there anyway', no matter what you think about them. This is true whether beliefs are explicit representations (Mandelbaum and Quilty-Dunn 2017) or dispositions (Schwitzgebel 2010). Even on Schwitzgebel's view, whether or not you 'sort-of believe' something is a fact about you which you can discover by reflecting on your dispositions; but changing your mind is

something very different — it is changing your dispositions. On the usual view, then, discovering what you believe is one thing, but making up your mind is a very different thing.

This is what the Worldview theory denies. It is the essence of the Worldview theory that there is no sharp distinction between finding out what you believe about some subject matter, and making up your mind about it. There are facts about what your psychological dispositions are, and facts about how you behave; when reflecting on all these facts about yourself you form a conscious judgement one way or the other. But if the Worldview theory is right, then there are cases where there is no fact of the matter about whether the content of the judgement was something you already were committed to, or whether you were creating a new belief from the dispositions you already have.

If can we only think of beliefs as determinate, discrete states (whether dispositions or representations), and of our conscious judgements as the outputs of these states — in the way a printer outputs the text stored on a computer — then this will seem thoroughly obscure. How can it be that there is no fact of the matter about whether you are deciding you believe something, or finding out that you believe it?

Our view is not that there are no clear cases in which you find out what you already believe; or that there are no clear cases in which you make up your mind. It is that there is not a sharp distinction. The way to see the point is to focus on what happens when you make the conscious judgement. The conscious judgement is not just a mechanical production of an unconscious state, like a computer's output. Rather, it is an attempt to make sense of what you know about yourself. We compare the relationship between the conscious judgement and the underlying state to 'modelling' — in the sense used in the philosophy of science (see Crane 2013). The conscious act of acknowledging or judging is a way of modelling your unconscious.

When you assert something, you express your belief by expressing a proposition, which in many cases will be a simplification or an idealisation of the underlying unconscious reality, and in this sense a 'model' of it. But it is a model which makes sense of this reality, just as scientific models make sense of the messy reality by simplifying. One way in which a judgement can make sense of unconscious reality is to impose a determinacy on what was not determinate or fixed. The 'fixation' of belief is not just a matter of responding to perceptual or other evidence; it can also be a matter of making up your mind. Sally may reflect on all her dispositions and then conclude that no, she does not really believe that farm animals are of sufficient moral status for her to change her habits. Or she may conclude after reflecting on the very same facts that since she does

have strong views about the suffering of dairy cows, she will now become a vegan. To insist that one of these cases is a discovery and the other a decision is to insist on a distinction which need not exist in the unconscious.

The Worldview theory was introduced by Crane (2017) originally in connection with beliefs. But we hold that the same structure applies to desires and motivations. In fact, it might be easier to see the general point, from the first person perspective, for the case of motivational states. Take desire: we are all familiar with the phenomenon of wanting something (say, something to drink or something to eat) and yet not being quite sure what it is that we want or need. You try one thing, then another — sometimes you fail to be satisfied by any of it. But sometimes having tried one drink, you try another and then feel immediately that *this* is what you had wanted all along. Is this because in your unconscious you had really desired this drink rather than that one? Or is it that you realised retrospectively that this one satisfied some more general, less specific desire which you had before you drank it? There need not be always be a fact of the matter about which of these is the case.

Something similar can happen in the case of making a decision. One may think one is making a decision for a certain reason, but ‘the true reason for deciding only becomes apparent once the decision has been taken’ in John Forrester’s words (Forrester 1990: 198). In fact, Forrester’s ‘apparent’, here is ambiguous between the case in which one has already decided but one is unaware of this at the time, and the case in which one reconstructs a rationalisation retrospectively, maybe out of material that is already there, or maybe created after the event. Again, this need not be a sharp distinction in every case.

In all these cases — belief, desire, decision — our understanding of ourselves should be thought of as a kind of modelling of our psychological dispositions. Bringing what we believe to consciousness can be immediate and without obstacle, but it can also involve a process of reflection and interpretation. Crane’s (2017) statement of the Worldview theory leaves in place the idea that there is a clear distinction in the unconscious between the doxastic and the non-doxastic — for all that was said there, the Worldview could be just a messy version of Fodor’s ‘belief box’. The considerations in the previous section give us one reason for modifying this idea. The Worldview, as we now think of it, is the entirety of the subject’s psychological dispositions, whether these are classified as doxastic, conative, motivational, emotional and so on.

On this development of the theory, the acknowledgement by a subject of one of their states *as a belief* is also an interpretation: we are imposing an order on something which need not have this order in itself. Our unconscious psychological dispositions are

genuinely representational — perhaps because of their aetiology or for some other reason — but they do not all neatly fall into the category of ‘belief’ or ‘desire’. The assigning of these dispositions to these categories of attitude is also part of the modelling that we do of our own mental states.

The idea that beliefs and desires are clearly distinct types of mental states, with very different features, is, of course, widely accepted in philosophy. But some have been dissatisfied with aspects of this idea — for example, in moral philosophy there has been substantial debate about whether moral beliefs themselves can motivate (Smith 1994). But even those who accept that beliefs can motivate typically maintain the picture of beliefs as discrete representational states. Others have tried to break the exhaustive belief-desire dichotomy by hypothesising that there are other mental states which have belief-like and desire-like features. Ruth Millikan’s ‘pushmi-pullyu’ representations, which have what she calls ‘descriptive and directive content’ is perhaps the most familiar (Millikan 2005: 167; see also Shea 2018 Chapter 7). And Gendler’s ‘aliefs’, too, have representational and motivational elements.

We mention these things merely to draw attention to the idea that the motivational and the ‘purely descriptive’ states might, in some cases, be blended or indistinguishable at some level should not in itself be a wholly unusual idea. Our suggestion here pushes this further and rejects the idea that our unconscious psychological dispositions are necessarily divided into states that have the characteristics of beliefs and desires, as those are conceived in commonsense psychological explanation. It is rather that psychological explanation (of others) and interpretation of ourselves involves an element of construction and idealisation — what we are describing as modelling. In modelling our own states of mind, we classify aspects of our Worldviews as beliefs or as desires or as preferences or likings and so on. So just as the proposition which expresses the content of a belief is a model of the real content of that underlying state, the classification of the state as a belief or as a desire is also part of the modelling of our states of mind. Modelling takes place both at the attitude level and the content level.

This is the beginning of our general account of the mismatches between professed beliefs and actions. Our starting point is the observation, in section 2 below, we should not infer that just because someone’s dispositions lead them to act in a way that is not usually associated with the belief that *p*, that they must therefore believe not-*p*. A full account of our actions will rarely pinpoint one belief content rather than a collection of states that we model as beliefs, desires, preferences and so on. Schwitzgebel is right to say that once we have fully characterised someone’s dispositions, there is nothing more to say (about the

content of the unconscious mind, at any rate). But we extend this to the non-doxastic states too. We can certainly describe some parts of their 'dispositional profile' as beliefs, but there is no reason to think that all parts can be described in this way; indeed, even identifying one part of this profile in terms of a proposition believed is an abstraction from the actual psychological complexity.

This inevitably raises the question of realism, a topic which was at the focus of the philosophy of mind in the 1980s and 90s but has somewhat faded from view these days. To what extent is the distinction between beliefs and desires a real distinction in the mind? It is true that our practice of psychological explanation invokes beliefs and desires to explain why people do what they do, and explanatory schemes like decision theory (rational choice theory) have constructed a more rigorous, quantitative version of this practice. Our claim is that there is no straight route or simple correspondence between the descriptions of states which we give in our commonsense psychological explanations, and the psychological reality that makes them true.

The similarities between the view developed here and Daniel Dennett's views of psychological reality and intentionality may suggest that we are taking an 'instrumentalist' view of the intentional as opposed to a realist view. Although we willingly acknowledge the influence of Dennett's work here, we prefer to outline the similarities and differences rather than first adopt a label like 'instrumentalism' which has been the subject of controversy in the past. We share Dennett's opposition to what we could call the hyper-realism of Fodor's language of thought hypothesis, opposition he has developed since late 1970s; we share his commitment to the importance of understanding how to make sense of our attribution of beliefs to young children and animals. We agree with him that attributions of beliefs and other intentional states are often idealisations and that our attribution of beliefs is an attempt to discern real patterns in the behavioural and other dispositions of subjects (Dennett 1991b).

One difference between Dennett's approach and ours is that he focuses exclusively on third-person attributions of intentionality ('heterophenomenology'), whereas our view is intended to be applied to both third person and first person attributions. In particular, we hold that when we bring our beliefs to consciousness we often impose a kind of determinacy on what we think (what we acknowledge to be true) which Dennett's account of consciousness denies (Dennett 1991). Where consciousness is concerned, we depart from Dennett's interpretationism: consciousness is the place from where a certain kind of order is imposed upon the chaos of the unconscious. It is not the 'Cartesian Theatre' as defined by Dennett, since that is supposed to be a *place in the brain* where it all comes

together (1991: 39). We are not committed to there being a place in the brain where things all come together: where things do come together is in the person or the subject, and in its conscious point of view.

This difference from Dennett enables us to attempt to answer directly the question: are there really such things as beliefs? The simple answer to the question is yes — but the more complex answer depends on whether we are talking about the unconscious or the conscious.

When we are considering the unconscious, it is true that people and animals have beliefs — just as people have character traits, for example — but that neither truth requires a structure in the unconscious mind which corresponds exactly to the structure of our belief (and other attitude, or even character trait) ascriptions. What is in the unconscious mind is a complex of related dispositions whose manifestations can be interpreted in terms of beliefs, desires, decisions, emotions and all the other attitude concepts. Claims about what people believe are attempts to identify patterns among the manifestations of these dispositions; as such they are incomplete and idealised. If that counts as ‘instrumentalism’, then so be it.

On the other hand, there clearly are conscious episodes of acknowledging the truth of a proposition, or judging something to be true. These episodes are not themselves beliefs, since they do not persist beyond the conscious moment, but they are occasions when we commit to the truth or falsehood of something that we understand well, and in many cases we understand the consequences of the propositions too. And these occurrences normally lead to a firming up of the unconscious dispositional connections in relation to the proposition judged — judgement can bring what you believe closer to the surface of your mind and enable it to be poised to harmoniously affect your behaviour. In this way, our view is somewhat close to Hume’s: while we do not identify a belief with a conscious feeling, the conscious episode can imprint the dispositions into our unconscious, and they become the ‘governing principles of our actions’. We are as ‘realistic’ as it is possible to be about these episodes.

We can illustrate our view with an image. Since Quine’s influential discussions of belief-systems, it has become common to talk about the ‘web’ of belief. But a web is actually a rather well-organised structure, with nodes and connections. For the view we are trying to sketch here, the ‘swamp’ of belief would be a better image than the web. A swamp is an expanse of mud and water, with ill-defined boundaries which change over time. Things near the surface of the water are visible and may sometimes be easily extracted; but sometimes not, things connect them to the dark, smelly mud at the bottom

of the swamp, where roots and plants wind themselves around garbage that can only be extracted through long and careful digging.... And there is no sharp distinction, within the swamp of the unconscious, between different areas of the swamp. Objects in the swamp — roots, creepers, weeds, discarded garbage— can spread themselves all over the swamp, and the criteria of individuation of areas of the swamp are not at all precise. Things near the surface of the swamp are those things which we have no difficulty removing. But many things are closer to the bottom, with the result that they are harder to remove. Sometimes trying to remove them breaks or modifies them in some way, or disrupts what is already there. In this sense, perhaps, there is no sharp distinction between what is already in the swamp and what you create by trying to remove something.

Analogies like this, of course, only go so far. But if the unconscious does really have a structure more analogous to a swamp than to a web, then it would explain both why we sometimes find it hard to identify exactly whether we are being driven by a belief, a desire, or some other affective or emotional state; and why we sometimes find ourselves with a mismatch between our professed beliefs and our actions.

We will end with a comment on the significance of the question of how we give an overall explanation of mismatches between professions of belief and actions. Does it matter whether we say that the explanation is in terms of belief, or in terms of other states and events?

In so far as the mismatches are failures of some kind, then it will matter, since your answer will affect how you try to remedy these failures. If someone's actions are genuinely a result of a belief which they willingly acknowledge (for example, the prejudiced belief that all Roma people are criminals) then the only way to change their dispositions to bring more evidence and argument to dislodge this belief. But if their actions are a result of preference or taste, then there may be very little anyone else can do to change them, other than to argue about the appropriateness or morality of having certain tastes. However, since what matters in the moral sphere is what people do, there will be an advance even if people do not act on their preferences — whether or not they are able to modify them.

But in other cases that have been discussed in this paper, there is nothing that people need to change. No one needs to learn to walk over the glass floor at the Grand Canyon, and no-one needs to learn to drink from toilet bowls.

REFERENCES

Bach, Kent (1981) 'An Analysis of Self-Deception' *Philosophy and Phenomenological Research* 41: 351-70

- Bain, Alexander (1872) *Mental and Moral Science* (London: Longmans, Green & Co)
- Braithwaite, R.B. (1932-33) 'The Nature of Believing' *Proceedings of the Aristotelian Society* 33: 129-46
- Brownstein, Michael and Jennifer Saul (eds) *Implicit Bias and Philosophy* (Oxford University Press)
- Crane, Tim (2013), 'Unconscious Belief and Conscious Thought' in *Phenomenal Intentionality: New Essays* edited by Uriah Kriegel (Oxford: Oxford University Press 2013) 156-73
- Crane, Tim (2015) 'The Mental States of Persons and their Brains' in Anthony O'Hear (ed.) *Mind, Self and Person, Royal Institute of Philosophy Supplement* 76 (Cambridge: Cambridge University Press 2015) 253 - 270
- Crane, Tim (2017) 'The Unity of Unconsciousness' *Proceedings of the Aristotelian Society* CXVII (2017) 1-22
- Dennett, Daniel C. (1975), 'Brain Writing and Mind Reading' in Keith Gunderson (ed.) *Minnesota Studies in the Philosophy of Science* 7:403-15, reprinted in Dennett, *Brainstorms* (Hassocks: Harvester 1981)
- Dennett, Daniel C. (1982), 'Beyond Belief' in Andrew Woodfield (ed.) *Thought and Object* (Oxford: Oxford University Press)
- Dennett, Daniel C. (1991a) *Consciousness Explained* (Little, Brown)
- Dennett, Daniel C. (1991b) 'Real Patterns' *Journal of Philosophy* 88: 27-51
- Forrester, John (1990) *The Seductions of Psychoanalysis* (Cambridge: Cambridge University Press)
- Frankish, Keith (2016) 'Playing Double: Implicit Bias, Dual Levels, and Self-Control' in Brownstein and Saul (eds) (2016)
- Geach, P.T. (1970) *Mental Acts* (Routledge and Kegan Paul originally published 1957)
- Gendler, Tamar Szabó (2007) 'Alief and Belief' *Journal of Philosophy* 105: 634-663
- Gertler, Brie (2011) 'Self-Knowledge and the Transparency of Belief', in A. Hatzimoysis (ed.) *Self-Knowledge*. Oxford: Oxford University Press
- Hume, David (1739/40) *A Treatise of Human Nature*, various editions
- Kriegel, Uriah (2004) 'Moore's Paradox and the Structure of Conscious Belief' *Erkenntnis* 61: 99-121
- Mandelbaum, Eric and Quilty-Dunn (2018) 'Against Dispositionalism', *Philosophical Studies* 175: 2353–2372
- McLaughlin, Brian and Rorty, Amelie (eds.) (1988) *Perspectives on Self-Deception* (Berkeley: University of California Press)

- Mellor, D.H. (1978) 'Conscious Belief' *Proceedings of the Aristotelian Society* 78: 87-101
- Munton, Jessie (forthcoming) 'Prejudice as the Misattribution of Salience' *Analytic Philosophy*
- Myers-Schulz, Blake, and Eric Schwitzgebel, 2013, 'Knowing that P without Believing that P', *Noûs*, 47: 371–384
- Millikan, Ruth (2005) *Language: a Biological Model* (Oxford: Oxford University Press)
- Price, H.H. (1969) *Belief* (London: George Allen and Unwin)
- Schwitzgebel, Eric, (2001) 'In-between believing', *Philosophical Quarterly*, 51: 76–82
- Schwitzgebel Eric (2002) 'A phenomenal, dispositional account of belief', *Noûs*, 36: 249–275.
- Schwitzgebel Eric (2010), 'Acting contrary to our professed beliefs, or the gulf between occurrent judgment and dispositional belief', *Pacific Philosophical Quarterly* 91: 531–553
- Schwitzgebel, Eric (2019) 'Belief', *The Stanford Encyclopedia of Philosophy* (Fall 2019 Edition), Edward N. Zalta (ed.), URL = <<https://plato.stanford.edu/archives/fall2019/entries/belief/>>
- Searle, John R. (1989) 'Consciousness, Unconsciousness and Intentionality' *Philosophical Topics* XVII: 193-209
- Shea, Nick (2018) *Representation in Cognitive Science* (Oxford: Oxford University Press)
- Smith, Michael (1994) *The Moral Problem* (Oxford: Blackwell)
- Zimmerman, Aaron Z. (2018) *Belief: A Pragmatic Picture* (Oxford: Oxford University Press)