Lack of Attitude[1]

Tim Crane and Katalin Farkas

Department of Philosophy, Central European University

1. Attitudes and character traits

It is a commonplace view in contemporary philosophy that commonsense psychology consists in explaining people's behaviour in terms of their beliefs and desires. Familiar examples typically involve people going to the kitchen and getting something from the fridge, because they *desired* water (Zalabardo 2019), beer (Kriegel 2019, Smithies and Weiss 2019), wine (Crane 2003:186), yellow mango (Schroeder 2020) or something to eat (Fiebich and Michael 2015), and they *believed* that it was in the fridge.

It is a further question whether other mental states are required for full explanations of action, in addition to beliefs and desires (e.g. intentions and decisions: Bratman 1987, Holton 2009). It is yet another question whether only *propositional* attitudes — states with truth-evaluable propositional contents — need be used in explanations, or whether non-propositional attitudes are needed as well (Grzankowski and Montague 2018). But even if one held that they were, the general assumption is maintained, that psychological explanation is predominantly in terms of mental states, whether propositional or non-propositional attitudes. (There are some notable exceptions: for example, Andrews 2008 and Machery 2016.)

While it is obvious that we do often talk in this way about people's minds, this standard picture ignores something which is even more obvious: the appeal to character traits in our explanation and prediction of behaviour. Character traits are not propositional attitudes, but we use them all the time in our commonsense psychological talk. We have a rich range of concepts of character traits in terms of which we make sense of what people do. Just think, for example, of the differences between calling someone selfish, self-absorbed, self-important, or self-centred. It is not hard to think of people who are selfish without being self-absorbed — they do everything to suit themselves, but reflect little on themselves — or self-important without being selfish — they attach great significance to their place in the general scheme of things, but are generous and thoughtful with others. And similarly with the other 'self' concepts here.

Perhaps it is easy to overlook the significance of character traits if we focus on examples like getting things from the fridge, since people's personalities rarely make a difference to the execution of such tasks. But on more complex matters, especially in connection with understanding the relationships between people, character traits become much more important.

Relatedly, our concepts of character traits often have a strong ethical aspect, which is why they are discussed in moral philosophy, or in what is known as 'moral psychology' (one of philosophy's more peculiar classifications). The study of traits is also extensive in social psychology. Nonetheless, the absence of discussion of character traits in the philosophy of mind is remarkable — it is, in our view, one of the real lacunae in this area of the subject these days.

Our aim in this paper is not to provide a proper treatment of character traits, but (more modestly) to use the comparison between attitudes and character traits to shed light on the phenomenon of belief. However, we think the view of attitudes we propose here also promises to provide room for a future theory of traits.

When we talk about belief here, we are talking about a kind of standing or persisting state or attitude. A state is an instantiation or an exemplification of a property. In the case of belief, it is the property of believing something, a property that can be instantiated by many subjects. We assume that both people and animals can believe things, and that it makes sense to suppose that two or more people can have the same belief. We call belief a *standing* state because it can persist beyond the subject's awareness of it, or indeed beyond the subject's entire current state of consciousness. People's beliefs persist even through sleep and other forms of unconsciousness. We distinguish therefore between standing attitudes and episodes in the stream of consciousness.

In earlier work (Crane 2017, Crane and Farkas 2022a and 2022b) we compared the ascription of standing attitudes to the use of models in science. In both cases, we accept a large simplification to make sense of a very complex reality. Ascribing standing attitudes is supposed to give an explanation and reliable prediction of behaviour and conscious mental activity in a whole range of attitude-relevant situations. But except for simple cases, behaviour is not in fact organised around such robust attitudes: for there are no such things. Like scientific models, these ascriptions involve a large simplification of a complex underlying reality.

We are happy to call our view a kind of *fictionalism*, since it incorporates two commitments. First, we claim that many ascriptions of standing attitudes are literally false: the robust attitudes we described don't exist. But second, we maintain that these

ascriptions are still useful — indeed essential — for understanding one another. The comparison with scientific models should correct any suggestion that attitude ascriptions are merely frivolous, serving no serious purpose. The aim of these ascriptions is to make sense of one another, in ways we will describe below. So *modelism* might be a more accurate term than *fictionalism*.

This view of standing states can also be applied to character traits. In using character traits to understand one another, we are using a simplified model of a complex phenomenon. Let's call a character trait 'robust' when it is supposed to give a reliable prediction of behaviour in a whole range of trait-relevant situations. In his well-known book, *Lack of Character* (2002), John Doris argued that there are no robust character traits in this sense. Based on a series of social psychology experiments, as well as evidence gathered from history, anthropology and literature, Doris argued that behaviour is sensitive to small situational variations which, if the traits were robust, should not impact their manifestations.

Both Doris's view and the experiments he relied on have been subject to criticism (see for example some of the contributions in Part III of Fileva 2016). We do not accept Doris's 2002 view as a whole, but we see his work as a serious challenge to anyone who wants to defend the reality of character traits. Here we adopt the following view which takes on board both Doris's points and some of the critiques of his view. Character trait ascriptions have great practical utility, and are undeniably central to our understanding of one another. We constantly make predictions, often very reliable, about how people will behave based on whether they are punctual, trustworthy, mean-spirited and so on. These are part of the common currency of psychological explanation.

But our ascriptions of these traits can also overstate the robustness of the trait, and for this reason we may want to adopt a fictionalism or modelism about character traits. After all, it is barely credible that we might discover that there is a real nature to selfishness as distinct from self-importance — that we might find their underlying psychological structure, let alone their neural basis. In an obvious sense, then, character traits are not real — but they are indispensable to our understanding of one another. They are idealised, simplifying models of our minds.

We see a deep similarity, then, between attitude and trait ascriptions. They should both be seen as models. But models of what? What are our ascriptions attempting to model? Our overall answer is familiar, on the face of it: the totality of a subject's unconscious mental dispositions. This is understood broadly, to include dispositions to have conscious experiences, as well as dispositions to behave, verbally and non-verbally.

These dispositions include those things which are modelled as character traits — and in this sense, attitudes and traits are both dispositions. (How these dispositions are embodied in the brain or body is a further question which we do not address here — except to say that our position is compatible with a range of options on the mind-brain relationship.)

These dispositions are unconscious; they are modelled by standing state ascriptions, and standing states are unconscious, as explained above. The dispositions manifest themselves in consciousness — more precisely, in conscious episodes or temporary states. Thus the belief that *p* may manifest itself in a conscious assertion, or in assent or dissent in response to a question, or simply in a conscious act of judgement. A stored memory of an event may manifest itself in the conscious event of calling to mind the event in the imagination. A generic fear of dogs may manifest itself in an experience of fear in the presence of a particular dog — and so on. In our modelling of standing states, then, we distinguish between standing states, conceived as dispositions, and their manifestations in the stream of consciousness. So far, we hope, so familiar.

What may be less familiar is our conception of the relationship between the dispositions and the ascriptions of attitudes. The dispositions are real enough, and the attitudes explain the behaviour. But we should not think that these explanations succeed because the structure specified in the explanation always corresponds in any straightforward way to the causal structure of the dispositions. And so we do not hold that dispositions can be *identified*, in general, with beliefs or desires. Belief and desire ascriptions model our unconscious dispositions and their manifestations. On this view, to say that a belief is identified with a disposition would be like saying that the relevant structure of the solar system model of the atom, for example, must be found in the atom itself. But obviously this is not the way it is.

Again, the parallel with character traits is instructive. To say that a person is spineless or pusillanimous is to attribute to them certain dispositions. We might make reliable and accurate predictions of their behaviour on the basis of this ascription, thereby summing up a complex collection of dispositions and their manifestations. But this does not mean that we should identify the person's dispositions with their spinelessness. Rather, the character trait is a way of making sense of the behaviour by picking out an open-ended and complex set of psychological dispositions.

Of course, in the case of attitudes, there are simple cases. When an adult is asked 'what is 2+2?' they will immediately answer '4' and this disposition to answer can be identified with the belief that 2+2=4 — so long as other relevant mathematical dispositions

are in place, of course. But in more complex cases, the causal structure of the disposition will invariably exhibit complexities which cannot be lined up in any simple way with the attribution of one belief. The dispositions associated with the belief that bringing up one's children requires a mixture of discipline and kindness, for example, manifest themselves in an open-ended complex set of actions and reactions. Two people can be plausibly ascribed this belief even if their dispositions vary wildly.

Our point here is not just the well-known holism of the attitudes — the plausible idea that what a person believes or wants is connected to so many of their other attitudes that the attitudes themselves must be individuated in terms of these relations. It is rather that the interaction between the dispositions is so complex that in most cases when a belief is ascribed, there is no good reason to literally identify any one of these dispositions with the belief ascribed. The dispositions are connected in the way that holism says; but the idea that these dispositions *literally are attitudes* is at best a projection of a feature of our model, and at worst an illusion.

Taking the general idea further, we hold that the very classification of attitudes into beliefs and desires is also a feature of the model. One standard view is that beliefs and desires are distinguished by their 'direction of fit'. The role of beliefs is to 'fit' how the world is, and the role of desires is to make the world fit them.

The 'direction of fit' idea can be fruitful, but it would be wrong to think that this distinction is a distinction in the causal structure of the dispositions themselves. A familiar causal understanding of direction of fit — that beliefs are disposed to change in response to interaction with conflicting evidence; desires disposed to vanish when the propositional content of the desire is represented as coming true — is simply not true of our minds. Our beliefs often persist in the face of conflicting evidence. Our desires do not always disappear when their propositional content becomes true.

To this it could be said that the causal conception of direction of fit is really a normative standard to which the empirical reality can fail to live up to. We would rather say that it is part of a model of subjects' psychological dispositions, which like all models, is accurate in some ways and not in others. The truth is rather that collections of dispositions don't need to have a unique direction of fit, and models can incorporate attitudes with multiple directions. Such ideas lie behind the notions of *alief* (Gendler 2007) or *pushme-pullyu representations* (Millikan 1995). Our theory treats the introduction of these notions and other psychological concepts — even widely accepted ideas like Freudian unconscious desires, or mental health conditions like personality disorders — as further

conceptual innovations intended to offer a model of the hugely complex reality of our mental dispositions.

Elsewhere we have called the totality of these dispositions the subject's *Worldview* (Crane and Farkas 2022b). The Worldview is the part of unconscious mentality that is constituted by our cognitive, conative and affective dispositions. (There are other parts of unconscious mentality which are not part of the Worldview: for example, unconscious processing of information in the visual system.)

We also introduced the notion of the *Habitus*, which encompasses dispositions we normally associate with personality and character traits (Crane and Farkas 2022b). We would make the distinction between attitudes and traits, then, via the distinction between the Worldview and the Habitus.

Attitudes and traits work in tandem in our attempt to make sense of this unconscious psychological reality. We can explain the same behaviour by the competing models of ascribing someone a strong desire and a cautious character, or a more impulsive character and a weaker desire. This inevitably raises the question of whether there will always be a fact of the matter, in such cases, about which attribution is correct? We say no. The best description will be the one that makes the best sense of the subject's behaviour; but there may be more than one appropriate description in these terms, and each will bring some features to light, while putting others in the background.

So far we have been expounding our theory of the attitudes in general. We must now turn to the specific case of belief, and what it means to model an attitude with the concept of belief.

2. What are beliefs?

Central to our view is the sharp distinction we make between conscious episodes and standing states, and we place beliefs firmly in the category of the standing states. This might be partly terminological, and of course philosophers in the history of philosophy have occasionally talked in a different way. Hume, for example, treated belief as a matter of 'feeling'. But other disputes do seem to be more substantial. For example, some contemporary philosophers have assigned belief to the category of 'act' rather than state. According to Matthew Boyle, for example, 'belief is an enduring, non-occurrent act of assenting to a proposition' (Boyle 2011: 142).

Of course, we agree that there is such a thing as assenting to a proposition, but we think of this as a conscious occurrence, and not something that endures beyond the moment of conscious awareness. So we would question the idea of an 'enduring, non-

occurrent act'. Assenting to a proposition is a conscious occurrence, and if the idea of the belief 'enduring' is supposed to indicate the lasting commitment involved in believing, then it is not the occurrent act that endures, but the persisting state. (See Hunter 2018 for further criticism of Boyle here.)

The conscious assent to, or conscious acceptance of, the proposition that p is of course connected to the belief that p. Indeed, this kind of occurrence or episode is one of the main cues in how to interpret ourselves (and others) as having the belief that p. Assent is characterised by a conscious experience of commitment or conviction, in some cases associated with verbal expression. Commitment comes in degrees, and we do not want to claim that all entertained propositions come either with an experience of clear commitment or the lack of it. There will be cases where we are inclined towards accepting a proposition, but we are not entirely convinced.

But these cases are quite different from those where a person has a mixed dispositional profile: some behaviour and conscious episodes point towards believing that p, while others point towards lacking that same belief. The borderline cases of conscious commitment are comparable to the borderline cases of being tall. In the case of tallness, there is a single determinate measure which is relevant: a person's height. Similarly, in the case of a conscious episode, there is determinate measure that is relevant: the strength of felt commitment. In contrast, in the case of a belief, there is a multidimensional and holistically co-dependent nexus of dispositions, which in the 'in-between' cases, can be interpreted in different ways, depending on other features of the model.

This applies to other standing states too. Most standing states — though not all — have a privileged characteristic conscious manifestation. Desires, in the unproblematic cases (like the desire for the beer or the yellow mango the fridge), may manifest themselves in episodes of conscious feeling, or even yearning. Intentions may be manifested in conscious awareness of a commitment — not to a proposition, but to a course of action. These states are also manifested in other conscious episodes: in a feeling of joy when a desire is fulfilled, or a feeling of disappointment when a desire is frustrated. But there is a clear sense in which conscious yearning is the most intimate conscious relative of desire, just in the way that conscious assent is the clearest expression of belief.

The overall metaphysical picture here is that our standing states are dispositions which manifest themselves in behaviour and in conscious episodes. Some of these dispositions have their own distinctive kinds of conscious episodes associated with them — for example, we claim that the episode of judging or assenting is a distinctive

manifestation of belief. But some do not — there are no central or specific conscious manifestation of the complex emotion of love, for example. Someone's love for their children can manifest itself in countless ways — there need be nothing that stands to love as assent stands to belief.

This raises the question of how this metaphysical picture of belief differs from other contemporary views of belief. Is our view just a variant on functionalism or dispositionalism? These terms have meant many things in the tradition, so we first have to clarify how we understand them.

A disposition, as we understand it, is a state or property of an object whose nature involves certain characteristic effects (manifestations) in response to triggering conditions and other background circumstances. An essential feature of dispositions is that objects can have their dispositions whether or not they manifest them. To call a belief a disposition, then, is to say that beliefs have certain characteristic effects in certain circumstances (e.g. the presence of other attitudes) in response to triggering conditions (e.g. the act of asking oneself what one believes). Eric Schwitzgebel (2002, 2010) calls this the 'dispositional stereotype' of the belief — notice that a subject can have such a dispositional stereotype even if it is never manifested.

In a number of writings, Schwitzgebel has defended his version of dispositionalism about beliefs (e.g. Schwitzgebel 2002, 2010). His overall claim is 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' (2002: 253). It is an important part of his view that these dispositional stereotypes include phenomenal experiences as well as behaviour.

Our view has some similarities with Schwitzgebel's — in particular, in including phenomenal experiences in the dispositional stereotype — but it differs in two ways. First, we do not identify (all) beliefs with individual dispositional stereotypes. Rather, attributions of beliefs (like attributions of character traits) are idealising models which do not always directly track in detail our actual psychological dispositions. On our view it can make good sense to attribute beliefs in many of the 'in between' cases, since these ascriptions function only as models. (For more on these cases, see section 3 below.)

The second way in which we differ from Schwitzgebel is in terms of how we draw the contrast between dispositionalism and other views. Although Schwitzgebel sometimes presents his dispositionalism as a new and radical view, there is a sense in which (almost) all theories of belief treat beliefs as dispositions, or as having a dispositional component. Behaviourism treats beliefs as dispositions to behave, of course. Like Schwitzgebel we reject behaviourism, but not because it dispositionally connects belief and behaviour.

Functionalism understands beliefs in terms of causal or functional roles; but these 'roles' are best understood as dispositions too. For the claim that the belief that 2+4=4 is characterised by its typical pattern of relations (to its inputs, outputs and other mental states) cannot require that the these relations are continually being activated whenever someone has that belief. The relations must be dispositional, if they are relations at all. In this way *functional role* might just be another word for *dispositional stereotype*.

Of course, different types of functionalism may involve further commitments which go beyond this mere claim. And functionalism tends to emphasise more strongly the typical inputs which individuate psychological states, whereas Schwitzgebel's theory is largely concerned with their typical outputs. So we are not saying that functionalism is identical to Schwitzgebel's dispositionalism; only that they have an important common element.

The same can be said of other variants on causal/functional views of belief, such as Aaron Zimmermann's 'pragmatic' view, according to which to believe something 'is to be so disposed that you would use that information to guide those relatively attentive and self-controlled activities you might engage in' (Zimmermann 2019: 1). This plausible claim clearly identifies beliefs with dispositions of the believer.

Even representationalism, which Schwitzgebel likes to contrast with dispositionalism, involves some dispositional elements. Jerry Fodor, historically the leading representationalist, held that the difference between belief and desire can be functionally (and therefore dispositionally) characterised, even if the difference between the belief that *p* and the belief that *q* cannot (Fodor 1985). But some representationalists even treat the representations themselves as having a dispositional character. Jake Quilty-Dunn and Eric Mandelbaum, two contemporary defenders of the view, describe representationalism as follows:

> According to representationalism, to have a belief is to stand in a particular relation to a mental representation. The mental representation is poised to perform certain (typically computational) functions within the mind that often bear only remote connections to stimuli, behavior, and phenomenology. (Quilty-Dunn and Mandelbaum 2017: 2354)

Being 'poised' to perform certain functions in the mind is clearly a dispositional notion: the representations are there, ready to interact with others, whether they do so or not. Quilty-Dunn and Mandelbaum go on to insist that the 'mental representation is poised to perform certain (typically computational) functions within the mind that often bear only remote

connections to stimuli, behavior, and phenomenology' (2017: 2354), and they also would deny that a belief is *nothing more than* its dispositional profile. In these ways they distinguish their view from views like Schwitzgebel's. Our point here is not that they are really dispositionalists, but is just to draw attention to the ubiquity of dispositional notions in all or most contemporary theories of belief.

We join Schwitzgebel in rejecting representationalism, though we can only summarise our reason here. Representationalism holds that these representations in the brain can be individuated independently of these dispositional elements; for example, as sentences in a language of thought (see Quilty-Dunn, Porot and Mandelbaum 2022). Taking this literally, as it is intended, implies that there is a separate representation for each belief you have. And this means that there must be an answer to the question of how many beliefs you have, even if we can never find this out. We believe that Daniel Dennett's critique of this idea has never been properly answered (Dennett 1978; and cf. Crane 2017).

It may be useful to point out at this point that although we accept the essence of Dennett's critique of the language of thought, this is not because we accept the whole of Dennett's interpretationist view. The main difference is that we hold that to the extent to which it is correct, interpretationism only applies to standing states. These states are, in a sense, the 'products' of interpretation — or to put it more literally, the concepts of standing states like beliefs belong to the idealised models we use to make sense of one another. The reality being modelled is a collection of dispositions which do not map neatly on to the model, no matter how useful it is. This is very much in harmony with Dennett's views. But unlike Dennett, we do not think that this 'interpretationism' applies to consciousness itself. We hold that the stream of consciousness has a very different kind of reality from the unconscious (see Crane 2017).

This helps us to avoid a familiar problem that Dennett's view, and certain kinds of fictionalism about the mental, both face. Adopting the interpretative/intentional stance, and treating some piece of reality as a fiction, are plausibly themselves mental attitudes. So how can these specific mental attitudes themselves be explained by adopting the stance, or treating something as a fiction? In other words, since taking the intentional stance is itself a mental attitude, how can all attitudes be explained by the idea of taking the intentional stance? We avoid this problem by insisting that the relevant attitudes of the interpretationist are conscious states or occurrences which straightforwardly exist and have a nature in themselves. This provides a foundation for the models of the unconscious traits and attitudes.

Another place where we do agree with Dennett (and others) is in the idea that belief is an ideal. This ideal type in the 'belief model' is a state which is responsive to reasons and evidence. This is the conception summed up in Pamela Hieronymi's remark that 'believing is an activity done for reasons' (Hieronymi, 2009: 174). Although we think the psychological reality which we model by using the notion of belief is actually much more messy and irrational than this — we often believe for no reason at all — we nonetheless need the idea of a state which attempts to track reality and which is in some complex way responsive to reasons and evidence. We need it because we need to distinguish between the notion of belief and other psychological notions — the notion of a habit of behaviour, for example. Beliefs are rational, habits are not.

That is the ideal, anyway, and in many cases we live up to it. But what happens when we don't? What happens when belief fails to confirm to this rational standard? The last section of this paper will answer this question.

3. Borderline cases

An important theoretical role played by the notion of belief is in the standard analysis of knowledge. Many philosophers agree that belief — understood in the standard sense of the attitude of holding a proposition true — is necessary for knowledge. However, there are some dissenters to this view. Schwitzgebel and Blake Myers-Schulz (2013) designed a range of scenarios where they argued that agents have knowledge without belief. They supported their view by asking groups of US undergraduate students (not necessarily students of philosophy) whether they thought the protagonists of the scenarios knew or believed a certain proposition. These scenarios are a useful starting point to show what our proposal entails for cases where the presence of belief is subject to debate.

The first scenario discussed by Myers-Schulz and Schwitzgebel is based on Colin Radford's well-known example of the unconfident examinee (Radford 1966): someone who gives the right answers in an exam but feels uncertain when she gives them. In the second scenario, a person does not update a routine even though he receives contrary information: when he executes the routine, the new information is inert. The third scenario involves implicit bias: a professor who openly and sincerely repudiates any kind of bias in judging her students' intellectual abilities, yet many of her actions suggests that she treats student athletes as less intelligent than her other students. In the fourth scenario, after watching a horror movie, someone has a fearful reaction to an ordinary event that he is very unlikely to regard as dangerous. The fifth scenario is a classic case of self-deception: a husband who refuses to openly acknowledge various clear signs that his wife is cheating

on him, but he still undergoes anxiety and unwelcome images as a result of encountering these signs.

The result that Myers-Schulz and Schwitzgebel were interested was that in each case, people were more inclined to attribute knowledge than belief (77% vs 41%, on average for the 5 cases). For our present purposes, the interesting feature is that for each scenario, people were more or less divided on the question of whether an agent had a belief or not.

Our theory thrives on this kind of case. Each scenario presents a complex set of dispositions which do not lend themselves to a straightforward generalisation. People have different strategies of grouping together dispositions. In a particular case, two people who disagree on the ascription of a belief are offering different models of the situation. Since a single attitude rarely makes sense of a complex situation, their strategies are probably complemented by ascribing other attitudes. When we hear their competing stories, we can decide which provides a better model of the situation.

A special case of hearing these competing stories is to learn from philosophers about the philosophical theory that backs their belief attributions. In each scenario, the protagonists act, or feel, or think both in ways that is suggestive of a belief, and in ways that *prima facie* suggest the lack of the same belief. Myers-Schulz and Schwitzgebel, in accordance with Schwitzgebel's dispositionalist view of beliefs, propose that these 'either are not cases of belief or are, at most, "in-between" cases of belief in which the subject is on the vague border between believing and failing to believe' (2013: 371). We have already expressed our agreement with some elements of the dispositionalist view, but again, we would also like to emphasise here the differences.

Recall that there is a variety of mental features that we can use to model the causes of an action or a reaction. In addition to beliefs and desires, we can appeal to a whole range of emotions, preferences, traits, habits and so on. Once we recognise this, it is possible to attribute full blown beliefs to people, and at the same time explain their seemingly discordant actions or reactions by appealing to the other states.

A much-discussed version of this strategy is proposed by Tamar Gendler (2007). In cases that are relevantly similar to the freaked-out horror film watcher and the person who acts out of habit, Gendler states that there is no doubt about what the subjects believe:

> Ask the subject directly and she will show no hesitation in endorsing such claims as true. Ask her to bet, and this is where she will place her money. Ask her to

think about what her other beliefs imply and this is what she will conclude. Look at her overarching behavior and this is what it will point to. (2007: 638)

Hence the movie-watcher firmly believes that there is no danger, and the person acting out of habit believes the new information. But Gendler acknowledges that there is a mismatch between the belief and the subject's behaviour. Part of her explanation is that the situations evoke a set of affective and cognitive response patterns and motor routines, and these conflict with the behaviour that would be entailed by the belief. Gendler proposes that there is a separate mental state she calls "alief" which is the basis of these response patterns and routines. We have, for example, aliefs about dangerous looking creatures that are normally depicted in horror movies, and the content of the alief is something like "Dangerous creature! Activate flight mode!".

Gendler's idea of alief is especially useful in expounding our own view, since she claims we need a *new concept* to group together elements of a subject's Worldview. If Gendler's proposal were to catch on, then one day we will perhaps explain or predict people's behavior in terms of what they alieve — rather like the way people started to make sense of behaviour in terms of repressed desires under Freud's influence. Alternatively, we might remain sceptical that the concept of alief gives us a good model of Worldviews. It's not clear that we need this new concept above what we have already and what was clearly described by Gendler: affective response patterns and motor routines. (For doubts on whether "alief" is a legitimate psychological category, see also Muller and Bashour 2011).

Schwitzgebel identifies an important difference between his and Gendler's approach: he says that according to Gendler, 'beliefs, by their nature, are meant to track the truth and to change in response to evidence. Aliefs … do not change in this way' (2011: 539). Schwitzgebel thinks that the contrast between evidence-responsive beliefs and evidence-unresponsive habits won't stand in this sharp form: our habits and associations are to some extent responsive to evidence, whereas our dispositions to judge are often not. As we read Schwitzgebel's exposition of this idea, we feel he is very close to our view of the matter, but ultimately there is a difference.

We certainly agree with this much: the idea that belief is responsive to evidence involves a massive idealisation. This feature is part of an idealised model and ascribing such an attitude to people is often literally false. However, evidence-responsiveness is a highly useful feature of the model because it can helpfully distinguish beliefs from other mental states like hard-wired instinctive reactions, or habits. We are trying to hold together

two distinct ideas here: the first is using evidence-responsiveness as part of what distinguishes beliefs from the other attitudes. The second is acknowledging that often beliefs are not responsive to evidence at all.

Schwitzgebel stresses this second point. He complains that the evidence-responsive view of beliefs 'artificially hives off our rational and thoughtful responses from our habitual, automatic, and associative ones. In asserting that only the first are pertinent to belief, Zimmerman and Gendler attempt to separate what is really an inseparable mix' (2011: 540). We are very sympathetic to Schwitzgebel's idea of the 'inseparable mix' – we think that the underlying reality we are trying to understand is really such a mix. But unlike Schwitzgebel, we think it is worth trying to separate features of the mix through some idealisation. Rather than proposing, as a discovery about belief, that belief *really is* something which is not sensitive to evidence, we can keep the model of belief as evidence-sensitive, and leave open the option that attitudes other than beliefs are responsible for the discordant behaviour.

Here is one potential benefit of looking at things this way. Suppose that we regard belief as a reflectively endorsed, evidence-responsive attitude. If we then encounter a false belief, we know how to try to change it: by presenting the subject with contrary evidence. If people mistakenly believe that immigrants in a country have a high rate of criminal conviction and receive a disproportionate share of social benefits, we can cite clear evidence to the contrary.

In other cases, we find that people have problematic attitudes other than beliefs. For example, they may have implicit aversion towards certain ethnic groups that result in unfair treatment of those people when it comes to grading, or hiring, or recommending. The aversion will often disguise itself as a negative assessment of qualifications or capabilities. But the problematic attitudes in these cases are not best categorised as beliefs, so changing them is not a matter of presenting evidence that corrects a mistake. Rather, something else needs to be done.

Of course, Schwitzgebel can respond that he can endorse the same recommendations: since beliefs are not always responsive to reason, changing them will not always be a matter of rational persuasion. Perhaps he is right – this will depend on how the further details of the model fall into place. For the time being, we see a potential benefit in sticking to the ideal of belief as responsive to evidence in certain cases. Schwitzgebel treats implicit bias scenarios (e.g. implicit racist prejudice) as 'in-between' cases of belief, but he has greater sympathy with assigning belief on the basis of the implicit racist attitude than on the openly endorsed anti-racist commitment:

> If in cases like [an implicit racist prejudice] … we attribute belief primarily on the basis of avowals and explicit judgments, then our belief ascription no longer captures what it would seem the main purpose of belief ascription to capture: whether or not her general cognitive stance is racist or egalitarian. We leave out, it seems, what matters most in ascribing belief. (2011: 543)

Here we arrive at the crux of the matter: what *does* matter most in ascribing beliefs? We agree with Schwitzgebel's general answer that it is capturing someone's general cognitive stance, as this is manifested in her actions and reactions. But the cognitive stance is not sharply separated (for example) from the affective stance: it is an 'inseparable mix', in Schwitzgebel's own phrase. Racism is likely to permeate both. If we understand racist behaviour as the product only of cognitive attitudes, the risk is that we over-intellectualise the phenomenon and miss a crucial element in understanding it.

The answer to the question of whether some subject has some specific belief will emerge only once all the elements of the model are in place. A view of beliefs that emphasises responsiveness to evidence runs into a *prima facie* problem, for example, with the usual definition of delusions: that delusions are false beliefs with a bizarre content which cannot be shaken by contrary evidence. Based precisely on this feature, it has been claimed that delusions are in fact not beliefs. Lisa Bortolotti provides an excellent summary of these arguments (Bortolotti 2009: 56). Delusions don't integrate well with other beliefs, they are formed on the basis of insufficient evidence, they resist revision in the face of counterevidence, they don't reliably guide actions and they are not endorsed on the basis of good reason. In all these ways, delusions violate conditions of rationality, which, some claim, are necessary for belief ascription. Hence delusions are not beliefs.

Bortolotti herself thinks that none of these arguments work, and she devotes her 2009 book to showing that delusions are in fact beliefs. Her main argument is this: (1) ordinary beliefs other than delusions do not in fact satisfy the rationality conditions; so (2) delusions' failing these tests is not an obstacle to counting them as beliefs. Bortolotti mobilises a huge amount of material in support of (1), but here we can just consider a very simple example of dissonance among beliefs: "Stephen sincerely reports that his new colleague is hard-working, competent and fully qualified, but he predicts that she will make a mess of the annual report" (Bortolotti 2009: 78). Bortolotti takes it for granted that Stephen expresses two beliefs here, and indeed, using the most common cue for attributing beliefs, sincere assertions, it is natural to go along with this ascriptions. But

something is clearly amiss here, so other options are possible. If we heard Stephen's report, we might wonder if he really does believe whether his colleague is competent. Or perhaps we realise that he is jealous of her, and he predicts that she will make a mess only out of spite. All three strategies are attempts to make sense of Stephen's complex Worldview concerning his colleague.

Bortolotti's examples of the irrationality of ordinary beliefs provide an excellent demonstration of the starting point of our view: that people's dispositions, actions and reactions often resist any straightforward generalisation. Very often, we go along with the simplest cue for belief ascription: the explicit expression of an opinion in words. This is the main reason to ascribe beliefs also in the case of delusions: people's sincere insistence on the delusional content. If it seems useful to attribute a belief on that basis, then this is perfectly legitimate, as long as we add the qualification that the belief is delusional. We don't see the assessment of our complex Worldviews as *discoveries* made about beliefs, and delusions — for example a discovery that beliefs are not rational, or a discovery that delusions are beliefs. We see the project rather as balancing and negotiating among the different manifestations of our Worldviews, and trying to offer the best model to understand them.

4. Conclusion

The comparison between beliefs and character traits is meant to show two things. First, that certain crucial concepts used in familiar psychological explanations — like the concept of belief — need not be thought of as corresponding to robust states in order for these concepts to serve their purpose effectively. The same is true of character traits. And second, that denying the reality of beliefs and traits in this way is not supposed to be a rejection of the psychological, or some kind of general eliminativism about commonsense psychology. For the psychological reality — what we call the Worldview and the Habitus — is complex and real enough as it is. And once again, the same is true of character traits: the argument of this paper is not supposed to show that character is unreal. The concepts or character traits and beliefs are among the concepts we use to model our unconscious psychological reality. The fact that we use these models no more shows that psychological dispositions are unreal than the Crick and Watson 'double helix' shows that DNA is unreal. And thinking in these terms makes it easier to understand certain problematic phenomena in the philosophy of belief.

References

Andrews, Kristin (2008). "It's in your nature: a pluralistic folk psychology" *Synthese* 165 (1):13 - 29.

Bortolotti, Lisa (2009). *Delusions and Other Irrational Beliefs*. Oxford University Press.

Boyle, Matthew (2011) "'Making up your mind' and the activity of reason" *Philosopher's Imprint*,11(17), 1-24.

Bratman, Michael (1987) *Intention, Plans, and Practical Reason* (Cambridge, Massachusetts: Harvard University Press)

Crane, Tim (2017) "The Unity of Unconsciousness" Proceedings of the Aristotelian Society CXVII *(*2017) 1-22

Crane, Tim and Katalin Farkas (2022a) "The Limits of the Doxastic" in *Oxford Studies in the Philosophy of Mind*

Crane, Tim and Katalin Farkas (2022b) "Mental Fact and Mental Fiction" in Tamás Demeter and Adam Toon (eds.) *Mental Fictionalism* (London: Routledge)

Dennett, Daniel (1978) "A Cure for the Common Code" in *Brainstorms* (Hassocks: Harvester)

Fiebich, Anika & Michael, John (2015) "Mental Actions and Mental Agency" *Review of Philosophy and Psychology* 6 (4):683-693

Fileva, Iskra (ed.) (2016) *Questions of Character* (Oxford: Oxford University Press)

Fodor, Jerry A. (1985) "Fodor's Guide to Mental Representation" *Mind*

Grzankowski, Alex and Michelle Montague (eds) (2018) *Non-Propositional Intentionality* (Oxford: Oxford University Press)

Hieronymi, Pamela (2009) "Believing at Will" *Canadian Journal of Philosophy*, *supplementary volume* 35, 135-187

Holton, Richard (2009) *Willing, Wanting, Waiting* (Oxford: Oxford University Press)

Hunter, David (2018) "The metaphysics of responsible believing" *Manuscrito* 41 (4):255-285

Kriegel, Uriah (2017) "Brentano's Evaluative-Attitudinal Account of Will and Emotion" *Revue Philosophique de la France Et de l'Etranger* 142 (4):529-548

Machery, Edouard (2016) "De-Freuding Implicit Attitudes" in Brownstein, Michael and Saul, Jennifer (eds.) *Implicit Bias and Philosophy, Volume 1: Metaphysics and Epistemology* (Oxford Oxford University Press)

Millikan, Ruth Garrett (1995) "Pushmi-pullyu representations" *Philosophical Perspectives* 9:185-200

Muller, Hans D. and Bana Bashour (2011). "Why Alief is Not a Legitimate Psychological Category." *Journal of Philosophical Research* 36: 371-389

Myers-Schulz, Blake, and Eric Schwitzgebel (2013) "Knowing That P without Believing That P." *Noûs* 47, no. 2: 371–84. http://www.jstor.org/stable/43828832.

Quilty-Dunn, Jake and Eric Mandelbaum (2018) "Against dispositionalism: belief in cognitive science" *Philosophical Studies* 175: 2353–2372

Quilty-Dunn, Jake, Nicolas Porot and Eric Mandelbaum (2022) "The Best Game in Town: The Re-Emergence of the Language of Thought Hypothesis Across the Cognitive Sciences" *Behavioral and Brain Sciences*:1-55

Radford Colin (1966) "Knowledge: By Examples" *Analysis* 27, (1966), pp. 1-11

Schroeder, Tim, "Desire", *The Stanford Encyclopedia of Philosophy* (Summer 2020 Edition), Edward N. Zalta (ed.), URL = <https://plato.stanford.edu/archives/sum2020/entries/desire/>.

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.

Smithies, Declan & Weiss, Jeremy (2019). "Affective Experience, Desire, and Reasons for Action" *Analytic Philosophy* 60 (1): 27-54.

Zalabardo, José L. (2019). "Belief, desire and the prediction of behaviour" *Philosophical Issues* 29 (1):295-310.

Zimmerman, Aaron Z. (2018) *Belief: A Pragmatic Picture (*Oxford: Oxford University Press)

.