AGAINST ALIEF

Penultimate Draft (forthcoming in *Philosophical Studies*)

Abstract

This essay attempts to clarify the nature and structure of aliefs. First I distinguish between a robust notion of aliefs and a deflated one. A *robust* notion of aliefs would introduce aliefs into our psychological ontology as a hitherto undiscovered kind, whereas a *deflated* notion of aliefs would identify aliefs as a set of pre-existing psychological states. I then propose the following dilemma: one the one hand, if aliefs have propositional content, then it is unclear exactly how aliefs differ from psychological states we already countenance, in which case there is no robust notion of aliefs; on the other, if aliefs just contain associative content, then they cannot do the explanatory work set out for them, in which case there is no reason to posit aliefs at all. Thus, it appears that we have little reason to posit the novel category of robust aliefs.

In a series of recent papers Tamar Gendler has put forward the provocative idea that there exist a set of mental states, aliefs, which play an integral role in our mental economy (Gendler 2008a, 2008b). It is unclear whether aliefs have been introduced in a *robust* sense, where they are a hitherto undiscovered psychological entity that would be a new entry into our ontology of the mental, or whether they are introduced in a *deflated* sense, where aliefs are just a name given to a group of psychological states we already countenance. The robust notion of alief is intriguing, far-reaching, and would call for a reconceptualization of many well-known, though perhaps not well-understood, psychological phenomena. Consequently, the project of assessing whether there is a robust notion of alief deserves serious attention and scrutiny.

I am skeptical that there is a robust notion of alief that survives such scrutiny. My discussion proceeds as follows. First I review Gendler's characterization of aliefs and propose that there is a crucial ambiguity in her formulation about their content. Either aliefs have propositional content or they can only contain associative content. I then propose the following dilemma: on the one hand, if aliefs do have propositional content, then they do not appear to differ from psychological entities that we already countenance, in which case there is no robust notion of alief. On the other hand, if aliefs do not have propositional content, then they cannot do the explanatory work Gendler sets out for them and we have no reason for positing aliefs of either the robust or deflationary kind. I conclude the essay by offering some suggestions as to how the functional role of aliefs may serve as a foundation for a fully fledged, robust notion of alief.

1. Aliefs and Associations in Structure and Content

Before we can adequately assess the status of aliefs, we must be clear on what exactly aliefs are supposed to be. This task is not easy, since the notion of alief is still in its infancy. Consequently, I will begin by relying heavily on Gendler's actual words in order to explicate the idea. Gendler writes that an alief is a mental state with "associatively linked content that is representational, affective and behavioral, and that is activated—consciously or nonconsciously—by features of the subject's internal or ambient environment. Aliefs may either be occurent or dispositional" (Gendler 2008a, p. 642).

Since a lot of the issues of the issues regarding aliefs turn on how the 'associative' part of their characterization is understood, it is worth spending a bit of time getting clear on exactly what these associations are. Prima facie, Gendler has associations essentially ranging over the three types of psychological entities that constitute aliefs: the representational, affective, and behavioral (viz., a content, a valence, and a motor response). However, one might well wonder whether the representational component of aliefs also has to be associative; that is, one might wonder whether the representational content part of an alief cannot have propositional structure and thus must consist of either a single representation or a mere association between multiple representations. The differences between these two types of associations are important: if it is the 'association-between-psychological-types' reading that Gendler wants, then, there is no de facto restriction on the sort of representational content that can be a part of an alief. In which case aliefs could, in principle, take the same content as a belief. If this is the intended reading, then we are allowed to expand or constrict the representational content part of an alief as much as we would like. However, if the sense of association at play is the

association-between-content sense, then we have a much stronger, and I will argue less plausible, characterization at hand.

Gendler clearly intends aliefs to at least be an association between psychological types. Yet it appears that she also wants to allow aliefs to be able to have non-associative (viz. propositional) content. For example, when further characterizing the representational aspect of aliefs, Gendler writes, "In paradigmatic cases, an activated alief has…the representation of some object or concept or situation or circumstance, perhaps propositionally, perhaps non-propositionally, perhaps conceptually, perhaps non-conceptually" (2008a, p. 642). Thus, it seems that Gendler is unsure whether aliefs can take a propositional content, though she allows for the possibility.[1] Gendler sees that there is a question about what content aliefs can take, but she puts aside this question, presumably because she does not think it looms particularly large in the exposition and defense of aliefs. However, I will argue that this aspect of aliefs—whether they contain propositional content or not—does important theoretical work. If aliefs do have propositional content (that is, they aren't necessarily associative with regard to not just their structure but also their content), then it is unclear how they differ from other psychological states we already countenance. In order to show this we'll have to go somewhat slowly through the properties of aliefs. In the process I will argue that none of

[1] The reader might be wondering what the difference between propositional and associative content amounts to. To a first approximation, (states that take) propositional contents can have satisfaction conditions, whereas (states that only take) associative contents cannot. Satisfaction conditions are presumably determined by the syntax and meanings of contents, whereas associative contents are (relations among) mental representations that lack any syntactic structure. Accordingly, propositional contents can play a role as a premise in valid inferences, whereas associative contents never play such a role, but rather only enter into associative chains of thought (where such chains cannot be valid or invalid).

the properties of aliefs distinguishes aliefs from other, more mundane psychological states. If aliefs are not distinguishable from our pre-existing psychological states, then there can be no robust notion of aliefs.

2. Why Robust Aliefs Cannot Be Propositional

Aliefs are given two types of characterizations. There are the explicit characteristics that Gendler gives in the quote above, and there is the implicit characterization that aliefs receive from the work they are supposed to do in specific cases. Here I will focus on the explicit characterization of aliefs, following Gendler's words closely and in particular allowing for the possibility that aliefs can have propositional content. If there is a robust notion of alief to be had, then some of the properties of aliefs have to differ from psychological states we already countenance. The goal of this section is to show that it does not appear that any of the explicit characteristics of aliefs distinguish aliefs from other quotidian mental states; all of the characteristics apply to at least beliefs and concepts (and perhaps other mental states too, such as imaginings, suppositions, etc.).[2]

As discussed, aliefs must consist of at least an association between a representation, an affect, and a motor response. However, other mental states that we antecedently acknowledge also appear to have the same types of associations. For

[2] NB: With regard to the present argument, it doesn't matter whether the putative properties of aliefs apply to beliefs per se; all that matters is that there is some set of mental states that can have these properties, in which case the explicit characterization of aliefs doesn't provide a unique distinguishing set of properties. That said, since I suspect that beliefs do happen to have all the properties give to aliefs, they will serve as the main example in the text. Of course, this is not meant to show that the properties of alief are sufficient for something to be a belief; it's not even necessary that they be necessary. The argument against the robust notion of alief simply requires that the properties of alief are not uniquely distinctive when compared to more familiar mental states.

example, the belief THAT IS A TIGER IN MY BED tends to co-activate other behavioral and cognitive states (such as a fear response, a belief that one should run away, and a motor response readying the escape behavior)[3]. For a different example, imagine an (idealized) output of a Fodorean visual module (or more accurately, of the last module in the visual system whose output goes to central cognition; see Fodor 1983). Let us say the output is LO, A PANTHER. This output (a) is representational; (b) has an affective component (thinking of nearby panthers most likely causes sweating, fear, etc.); and (c) is associated with a behavioral component, in this case, the readying of the fight-or-flight routine. Yet, LO, A PANTHER is just another run-of-the-mill (perceptual) belief.

It is not just beliefs that have representational, affective, and behavioral components. Take the concept PENGUIN. This concept appears to be closely associated with the following information: PENGUIN is pronounced (pĕng'gwĭn). We have every reason to suppose that when we think of penguins we are quicker to say 'penguin,' spot penguins, and mistake other animals for penguins. Merely tokening the concept PENGUIN readies us for penguin-related behaviors, which is, prima facie, bad news for the alief supporter. Gendler writes that "alief does not involve the execution of these motor routines; it merely involves their activation" (Gendler 2008a, p. 644). However, the same holds for all types of quotidian mental states. Tokening PENGUIN (or believing THERE IS A PENGUIN) does not involve executing motor commands either, but it does ready them; if it did not why would we, for example, be faster at lexical decision tasks involving

[3] Small caps will be used throughout to denote structural descriptions of concepts. The structural descriptions are stipulated, but the stipulations will not affect the arguments in the text. If, for example, it turns out that TIGER is a complex concept, then one can substitute one's preferred simple concept in for TIGER.

'penguin' after we have activated PENGUIN?[4] Lastly, tokening PENGUIN is also generally associated with some affect (thinking about penguins should make you feel warm and fuzzy). Once again, the hallmark properties of alief appear to be identical to the properties of other canonical mental states.

Perhaps you are skeptical about the affective property of PENGUIN; some readers may feel that thinking about penguins still leaves them cold. However, introspective intuitions here, as in so many other places in the philosophy of mind, may mislead. Recent work coming out of Mike Tarr's lab shows that the mind appears to be rife with 'microvalences' (Lebrecht et al, ms.; Lebrecht and Tarr 2010). Microvalences are subtle pieces of affect associated with concepts. They are *micro*valences because the affect is often too subtle to clearly notice from (at least cursory) introspection. The Tarr lab's recent evidence suggests that all concepts appear to have such microvalences, and that the function of such microvalences is to facilitate speedy motor responses without incurring higher cognitive cost (if it helps one can think of these as a ubiquitous form of Damasio's 'somatic markers', ubiquitous because of their attachment to all concepts see, e.g., Damasio 1994).

To get the feel of the use of microvalence, suppose it is early in the morning and you have just brewed some coffee and need to put it in a mug. When you open your cabinet you will see many mugs available for coffee consumption. In such a situation, how do you choose which mug to drink out of? One possibility is that you have a decision-theoretic metric with a preference ordering amongst the mugs. However, such a procedure would be costly in terms of cognitive efficiency. Instead, Tarr and colleagues

[4] Of course, such tokenings need not be conscious (as can be seen by the efficacy of subliminal priming).

posit that each mug has a microvalence, and that these microvalences help to guide motor responses and facilitate quick decision processes.[5] Now assuming what should be untendentious to assume: that occurrent beliefs are, in part, constructed out of concepts, beliefs would inherit the microvalences that are attached to concepts, and in turn inherit the motor routines that are tied to microvalences. If so, then as a matter of fact we would never have 'cold', valence-less beliefs; instead beliefs would always be associated with affect and motor routines, just like aliefs. If this is the case, then all beliefs do, de facto, come with microvalences.[6] In sum, it appears that having a content that is associated with a motor routine and a valence doesn't distinguish aliefs from beliefs.[7]

The other properties of aliefs don't appear to isolate aliefs from other, more familiar states either. For instance, another property of aliefs is that they may occur unconsciously. Thus, a person may be in a particular state without knowing that she is in that state. Yet it is uncontested that most mental states can be tokened unconsciously (e.g., in the ubiquitous subliminal priming paradigm one tokens a concept unconsciously). Additionally, beliefs and other propositional attitudes can also be tokened unconsciously (see, e.g., the role of desire in Freudian psychology, or the role of

---

[5] Importantly, microvalences are purported to be applicable to not only singular concepts, but also concepts covering general categories (i.e., so it's not that microvalences just apply to THIS MUG; they can also apply to MUG full-stop). Of course, the evidence in favor of microvalences is still reasonably scant, so any inference about the scope of microvalences is far from apodictic.

[6] Moreover, if the essential properties of beliefs are tied to the actual properties of beliefs in this world, perhaps something could not count as a belief unless it too had an associated valence and motor routine

[7] If you assume that other propositional attitudes (imaginings, supposing, etc.) are also constructed out of concepts, then, if the microvalence hypothesis holds, all of our attitudes come with attached valences and motor responses. The consequence is so much the worse for the robust notion of alief.

belief in cognitive dissonance explanations).[8] Thus the property of being unconscious does not separate aliefs from beliefs and other mental states.

The activation conditions of an alief don't appear to do the necessary individuating work either. On the activation conditions Gendler writes, "It may be activated by features of the subject's internal or ambient environment" (2008a, p. 644). Needless to say, the same holds for beliefs (and concepts). One can token the belief I AM HUNGRY by sensing one's bodily states or through certain cues from the ambient environment (e.g., hearing one's stomach rumble or having someone point out that you happen to be mechanically eating potato chips, which you are known to dislike). To round out the properties, Gendler mentions that aliefs may be either occurrent or dispositional. Of course, it is not particularly tendentious to think that the same is true for beliefs.

It should be clear from this short discussion that aliefs appear to have the same properties as other mental states; in particular, aliefs don't seem to differ from beliefs. Consequently, the robust notion of alief is imperiled. We seem to have no need for a new category of mental states, for the explanatory burden that aliefs are supposed to relieve can be carried out by psychological entities we already countenance. If one wants to establish that there are robust aliefs, then one must show how aliefs essentially differ

[8] In most versions of cognitive dissonance people are attributed core unconscious beliefs (such as: I AM A MORALLY GOOD PERSON); see Thibodeau and Aronson 1992. Additionally, some promising explanations of implicit bias may also need the notion of unconscious belief: one way to attempt explain the workings of implicit racism is to posit that the implicit racist harbors an unconscious belief that, e.g., Caucasians are superior to African Americans. Lastly, vision science often posits unconscious beliefs (generally termed 'assumptions' for though they do appear to underwrite inferences from shading to shape, they aren't globally inferentially promiscuous), such as the belief that there is a single overhead light source (e.g., Ramachandran 1988, Scholl); for arguments that such intramodular propositional states are indeed beliefs see Dwyer and Pietrowski (1996).

from beliefs. As foreshadowed, the associative quality of aliefs might provide a distinguishing characteristic of aliefs. For example, if aliefs were essentially *associative in their content*, then aliefs would be sufficiently distinct from beliefs (and presumably distinct from the rest of the propositional attitudes too). Whatever else one thinks about beliefs, they are surely truth-apt—their that-clauses take propositions and are capable of evaluation. Associations, in contrast, are not truth-apt. If the content of an alief had to be essentially associative, then aliefs would have to differ from beliefs and we would have a basis for justifying the inclusion of a robust notion of alief in our ontology. So, for the time being let's suppose that aliefs are essentially associative not just in their structure but also in their content.

In the next section I will analyze one of Gendler's alief examples to show that aliefs must contain propositional content in order for them to do the explanatory work set out for them. This section will establish an 'upper bound' on the content of aliefs. Once we see that aliefs must be propositional in some cases, we can return to the question if there are any ways of retain the robust notion of aliefs while still allowing aliefs to do some explanatory work.

3. Why Explanatory Aliefs Must Have Propositional Content

I will outline two arguments for why an alief's content cannot be essentially associative. The first argument is based on what I will term 'binding.'[9] The second is based on

[9] NB: My use of 'binding' is not quite the use at play in cognitive neuroscientific discussions of the 'binding problem' in visual perception. Although it risks confusion to employ the term in another way, I find the word 'binding' to be helpful for getting at the underlying idea. Additionally, there is precedent for my usage: Adina Roskies sees the

inferential promiscuity. To see how these arguments work, let us consider Gendler's alief-based explanation of Paul Rozin's poison experiment (Rozin et al. 1986; 1990).[10] In Rozin's experiment, each participant is shown two empty bottles. Then, in plain sight of the participant, the experimenter fills each bottle with sugar (taken from a commercially labeled sugar box). The experimenter then shows the participant two labels, 'Sucrose' and 'Sodium Cyanide,' and asks the participant to affix one to each of the bottles in whatever manner he or she prefers. The contents of each bottle are then emptied into separate glasses and mixed with water. Intriguingly, participants are generally more hesitant to drink from the glass that contains the sugar that was poured from the bottle with the 'Sodium Cyanide' label—the very label that they themselves affixed. Gendler concludes that, though the participants believe that both bottles contain sugar (and water), they *alieve* that one of the bottles contains sodium cyanide.[11]

The problem for Gendler is that the putative alief looks to be propositional and we need aliefs to be essentially associative in order to underwrite the robust notion. Let us look a bit closer at the Rozin example Gendler uses. Gendler claims that the content of the alief at work is "CYANIDE, DANGEROUS, AVOID" (Gendler 2008a, p. 648). But what is this alief 'telling us' to avoid? To put the question another way, when I token the alief with content CYANIDE, DANGEROUS, AVOID, what am I thinking? If I am just tokening

---

problem I will describe and the traditional binding problem as structurally similar. See Roskies (1999). Nevertheless, I apologize for any confusion my usage might cause.

[10] Gendler addresses this experiment directly in Gendler (2008a).

[11] The locution 'alieve that' might strike one's ears as odd. If aliefs are associative in content it's difficult to see how they can be *propositional* attitudes (even if aliefs can take propositional content, it's not necessarily clear that they would be propositional attitudes). Nevertheless, it is the locution that Gendler adopts; for example, in describing Rozin's subjects who are looking at a pile of vomit-shaped rubber, Gendler writes, "they alieve that it is vomit" (2008a, p. 653). Even though I adopt this locution I remain agnostic as to whether aliefs could count as propositional attitudes.

these concepts in succession (which is what Gendler's 'associative state' talk implies), then why would I show any behavior whatsoever toward the bottle (and its contents) and not, say, the window, my left foot, or the experimenter's forehead? Since the behavior is bottle/bottle-content specific, the putative alief must somehow bind to the bottle (and its contents), or else participants would not show the avoidance behavior toward it. Merely saying that the alief's content is associated with the bottle doesn't explain why the alief binds to the bottle (and its contents) alone.

To see the problem a bit clearer let's compare the alief case to some paradigmatic examples of associationist cognition, such as the type that John Bargh's lab is forever uncovering. Bargh and colleagues have experimentally demonstrated that people associate warmth with friendliness. For example, they have shown that when one is holding a warm cup of coffee, one is apt to act friendlier than usual (Williams and Bargh 2008). But these subjects aren't more apt to act friendly toward the cup of coffee; rather, they are more apt to act friendlier tout court. Their mood gets enhanced across the board, as if the activation of the concept WARMTH has spread throughout their cognitive store. Compare this type of association with Gendler's use of the cyanide case. The subjects in Rozin's experiment aren't acting afraid in general; instead their fear is bound just to the bottle with the cyanide label (and its contents). However one wants to spell out the associative quality of aliefs, it appears to be a very different creature than the associations at play in Bargh-style priming experiments.

So to return to the Rozin case, how can the alief theorist ensure that the alief binds to the bottle? Perhaps the alief can contain a content more akin to THAT [demonstrative standing in for the bottle] DANGEROUS CYANIDE AVOID. But are the participants just

thinking the concepts THAT DANGEROUS CYANIDE AVOID, one after another, with no syntax as it were? If so, then why would we avoid that particular bottle? Instead, it seems like the participants must be thinking something like THAT IS DANGEROUS CYANIDE, AVOID IT.[12]

The present problem is that the content of the alief must somehow bind to the bottle, and the associative content that Gendler specifies for the alief has no way of attaching to the bottle as opposed to anything else. In order to bind in the right way, the content needs to be structured, and associative content cannot provide the right type of structure.[13] This explanation of the problem is what I term 'the binding argument.'[14]

There is another reason that aliefs need to contain propositional contents. Pure associative chains do not allow for inferences, but the putative aliefs do appear to allow for inferences so they must be propositional, in which case these states are truth

---

[12] Of course, they could just think THAT IS CYANIDE, with CYANIDE being linked to DANGEROUS, which itself would be linked to avoidance behaviors.

[13] NB: This problem cannot be fixed by adding a fourth element to the content, such as an iconic representation of the bottle. Say the alief had the content CYANIDE, DANGEROUS, AVOID, PICTURE (where 'PICTURE' stands in for an iconic representation of the bottle). In such a situation it would still be a mystery why anyone would avoid the bottle, because these would be four separate thoughts, albeit thoughts that sequentially followed one another. If you are having trouble seeing the difference perhaps the following example will prove illuminating. Imagine we have two cognizers, one who tokens the thought SEXY WILDEBEEST (a single thought with an adjective noun structure) and the other who tokens SEXY followed by a tokening of WILDEBEEST (two separate thoughts). These are two very different cognizers; the first one clearly has some odd sexual proclivities, whereas the second one just appears to be someone lost in a stream of consciousness. We can predict a decent amount of the first person's behavior from knowing that the person tokened SEXY WILDEBEEST (for example, you probably would not want to let him pet-sit your wildebeest), but we cannot predict much of anything at all about the second cognizer. If aliefs were essentially associative (meaning no propositional content allowed), then alief contents would parallel our second cognizer. But this cannot be right, because we can predict the behavior of the participants in Rozin's experiments: we know they are apt to avoid the poison.

[14] For a related argument applied to Hume's associationism, see Fodor (2003).

evaluative and appear to work just like beliefs. In other words, aliefs seem to be inferentially promiscuous, but if they are essentially associative then they should be inferentially dormant—after all, one cannot make inferences from associative chains.[15]

To see how the putative aliefs can be inferentially promiscuous, imagine that right after you take part in the Rozin study, you are asked a follow-up question about whether other folks would drink from the bottle with the 'cyanide' label. In this case you would probably infer that others would not want to drink from the bottle. (Perhaps you would go through an unconscious chain of reasoning like THAT BOTTLE CONTAINS POISON, PEOPLE DO NOT LIKE DRINKING POISON, SO PEOPLE WILL NOT LIKE DRINKING FROM THAT BOTTLE.) In short, we should expect people to infer from THAT IS DANGEROUS CYANIDE, SO AVOID IT, to other semantically related (and under the circumstances, reasonable-ish) thoughts, such as that others will want to avoid the bottle labeled 'cyanide,' that the bottle would still be labeled 'cyanide' even if the room were a different color, that the bottle will keep its contents even if it is lifted off the ground, and so forth. There are a seemingly unbounded amount of quotidian inferences we would expect the participants to make, but these inferences can only be made from propositional states. Hence there must be belief-like propositional states in play.

Gendler may respond by allowing that the Rozin situation recruits both aliefs and beliefs.[16] Surely this is a very reasonable response. Regardless of one's take on aliefs, everyone short of eliminativists will want to condone the thought that subjects at least have some high-level occurrent beliefs, such as the fact that they subjects are in an

[15] For the locus classicus on inferential promiscuity, see Stich (1978).
[16] The following objection was raised by a very helpful anonymous reviewer, to whom I am much indebted.

experiment, that they are in a room, that there are bottles in front of them, etc. Indeed, Gendler even explicitly mentions that subjects in the Rozin experiment explicitly believe both bottles have sugar. The question before us is whether those sorts of beliefs can do the inferential work needed for the type of inferential promiscuity needed to satisfy arguments like the one above. It does not seem to me that they can.

Let's start by untendentiously assuming that subjects have (at least) the following beliefs: they believe that there are two bottles in front of them, one labeled 'Sodium Cyanide' and one labeled 'Sugar' and that neither bottle contains any poison. The question is whether these beliefs could serve as the premises in an inference for the multiple conclusions that one might draw. Certainly, such beliefs could serve as premises in some of these inferences; for instance, these beliefs could serve as premises in inferences that the bottle will retain its contents even when superficially moved. However, could these beliefs serve as a premise in an inference that has the conclusion that one's peers would prefer to drink from the bottle labeled sugar?[17] I think it is difficult to argue that this could be so. After all, we are assuming that the subjects believe that neither bottle contains cyanide. Principles of charity allow us to posit that subjects will believe that other subjects will also believe that although both bottles are labeled differently, neither contain cyanide (since this is made clear to all subjects in the experiment); in fact, participants even explicitly acknowledge this to be so. Assuming subjects think that their peers are mostly rational, the subjects should predict that their peers should have no preference for which bottle to grab, and so their beliefs should not

[17] Thinking that subjects would draw such a conclusion is quite reasonable; after all subjects in much more opaque setups tend to be able to predict the (less than fully rational) behavior of how other subjects would act (for a paradigmatic example of such subject based prediction, see Bem 1967).

underwrite an inference for a preference toward the bottle labeled sugar.[18] Thus, it appears that the quotidian beliefs that we can assume that subjects have will not be able to underwrite the inferences that subjects make in alief style cases. To get those inferences we need some state with a content that the bottle does indeed contain poison and this state cannot be a run-of-the-mill rational belief.

Perhaps the reader thinks that some form of simulation theory can take care of such inferences.[19] To sidestep such qualms, consider a different Rozin case of magical thinking, the case of contagion. Contagion cases are very similar to the cyanide case and are rife for alief-style explanations—indeed are the types of cases that motivate alief-style theorizing. In a contagion case, subjects will be (e.g.,) less apt to wear a sweater if Hitler previously owned it, but they will be more apt to don an article of clothing if worn by someone the subject holds in high esteem (Nemeroff and Rozin 1994). Perhaps unsurprisingly, people are also apt to pay more money for a sweater if it was worn by a celebrity they esteem (Newman et al. 2011). Presumably, Gendler would want to explain (e.g.,) the Hitler case by saying though the subject believes that there is no moral

---

[18] It is important to note that in other experiments that have subjects predicting how others would perform, the predictors do not themselves first take part in the study (i.e., they do not first choose a bottle) rather they just have the experiment explained to them and then infer how others would respond. Thus, they are not just projecting from their past behavior to other people's future behavior. This is an important caveat because it blocks one possible response: participants who haven't partaken in the study couldn't reason that since they were in fact hesitant to taste sugar from the bottle labeled sodium-cyanide others would be too. Thanks to Ian Evans for raising this issue.

[19] If so, then the alief theorist would want to posit that even imagination can serve as the stimulus for forming an alief, in which case the alief story would have to be amended to add that aliefs are activatable in imagination. But this seems a bit odd, for aliefs are supposed to preserve a certain amount of ecological validity—they are fitness enhancing precisely because in perception they trade off speed for accuracy in potentially dangerous situations. It is unclear why such states would appear in imagination too where they would lose their fitness enhancing qualities. No doubt, such an explanation is possible, but would be a serious amendment to the alief picture.

contamination of the sweater (because sweaters aren't fit to be the objects of moral contamination) they still alieve that the sweater is morally contaminated (with the triple content containing a representation of the sweater, a negative affect, and an avoidance reaction). So far, so good.

Yet, a very interesting, yet overlooked property of such cases is about to cause trouble for the non-propositional understanding of aliefs. The 'moral' contamination and goodness (in the positive celebrity case) can be eliminated by a small intervention. Take the celebrity case. People will pay lots of money to ascertain an article of clothing worn by a celebrity. Let's say that you really love George Clooney. If you do, you are more apt to pay big bucks for a George Clooney bandana than someone who doesn't care about George Clooney. However, if you are told that the bandana has been laundered after Clooney wore it, then you will be far less likely to pay similarly massive amounts of money for the bandana (Newman et al. 2001).[20] It appears that peoples' knowledge of what happens when something gets washed interacts with their putative aliefs about the object; the magical Clooneynesss of the sweater can get washed off!

The idea behind these cases is that people unconsciously think (to use a neutral word for the moment) that the sweater contains the essence of Clooney, and you can immerse yourself in the essence by wearing the Clooneyed sweater. But somehow people reason that washing the sweater erases the Clooneyness. The point to keep your eye on is that a merely associative account can't explain these types of effects. For example, it's not as if people have strong negative associations with hygiene which could swamp the positive association with Clooneyness. Rather, what's transpiring is that subjects appear

[20] See the 'sterilization condition' in the internal replication (ibid. p. 224).

to have some propositional state that expresses that the article of clothing contains Clooneyish material. This state is then inferentially promiscuous—it can interact with other knowledge stores in inferential ways. In particular, in this case the subjects' knowledge of what washing entails (e.g., it disinfects clothes) interacts with this propositional state to cause the subject to infer that the Clooney essence will be eliminated if the sweater is washed.

Regardless of what we would like to term this state, we can say something definitive about it: it must have propositional structure, for it acts as a premise in (unconscious) inferences. Note that in this Clooney example appealing to people's explicit beliefs won't help, for people don't claim to explicitly believe that the sweater has any real Clooney essence that can be washed off like mud from old jeans. So it appears that the aliefs must have some propositional content in order to allow it to interact in inferential ways with other things that people believe, such as the belief that washing things takes away scents and germs.[21]

***

Thus, the present challenge for the alief theorist is to figure out how to make room for a robust notion of alief. Either the alief theorist has to explain how inferences that appear to interact with aliefs can be explained without positing that aliefs contain propositional structure (in which case aliefs can be a novel, distinctive state in virtue of

[21] The reader might be wondering what type of state can explain such behavior. I suspect it's just old run-of-the-mill beliefs, albeit unconscious and arationally acquired one's that do the explaining. Of course, this would mean that people harbor contradictory beliefs, a conclusion that arises in surprisingly numerous facets of psychology (and one very much goes against the motivation for positing aliefs in the first place). For an expanded theory of arationally acquired and contradictory beliefs of this sort see Mandelbaum (2010); for other evidence that people hold contradictory beliefs, see Ripley (forthcoming); Strickland et al. (2011).

their solely associationistic content), or the alief theorist has to show that aliefs are distinct from other mental states we already countenance on other grounds. Without responding to this dilemma, we are left with a deflated notion of alief, in which case we haven't found a new animal in the kingdom of cognition and much of the revolutionary flavor of aliefs is lost.

Furthermore, this deflated notion of alief wouldn't serve the original motivation for positing aliefs, for it wouldn't rule out that people have certain irrational beliefs, like the belief that one of the Rozin bottles actually contains cyanide. If these beliefs must be implicated in explanations of behavior, then the deflated notion of alief wouldn't actually serve Gendler's original goals because explanations of relatively quotidian behaviors would still often involve attributing irrational beliefs to agents. If we do have to attribute irrational beliefs to agents, then it's unclear what reason we'd have for positing deflated aliefs in the first place.

4. A Suggestion for Finding a Robust Notion of Belief

How could we build up the deflated notion of belief such that it might prove to be a robust notion, a psychological kind unto itself? Everyone already countenances that affect, motor responses, and representations exist, and I for one don't for a second doubt that associations between these three psychological types exist. But as we've seen, merely noting that such associations exist isn't enough to underwrite a robust notion.

If we are to allow a robust notion of alief into our mental ontology, it has to be because aliefs play an explanatory role *as such*; in other words, aliefs should be their own stand-alone psychological kind. Determining the necessary and sufficient conditions for

kindhood is a task far too difficult for the current project. I will settle for something less and argue by examining a different, but no less influential, example of a proposal for a new psychological kind. In *The Modularity of Mind* Jerry Fodor proposed that input systems interpreted as modules may be an overlooked natural kind. Fodor writes,

> I'm about to argue that, if we undertake to build a psychology that acknowledges this functional class [modules] as a neutral kind, we discover that the processes we have grouped together do indeed have many interesting properties in common. (I take it that that is what a natural kind is: a class of phenomena that have many scientifically interesting properties in common over and above whatever properties define the class)" (Fodor, 1983, p 46).

The test Fodor puts forth is, no doubt, a heuristic one, but one that seems reasonable enough to serve as a stand in for a necessary condition. The question then is: do aliefs have psychologically interesting properties over and above their constituent, defining parts?

I think that the parade examples Gendler uses do her a disservice. In every example put forward to try to show the explanatory worth of aliefs, it is one of the contents of the triple pair that does the explanatory work and not any emergent properties of aliefs as such. For example, in Gendler's example of being up on a Skyway (2008a, 653) and feeling afraid even though you know you are safe, the fear that one feels is underwritten by affect, the instinct to move away from the railing, by the motor response. So the challenge before the robust alief proponent is to show some behavior that isn't itself explained merely by any of the component parts of the alief (and their elicitation conditions). Robust aliefs would be a category unto itself if there were some behavior (or property) that could be explained in terms of aliefs as such, and not in terms of their parts.

To me it appears that the best avenue for pursuing this question would focus on the functional role of aliefs: that their content does not seem to be affected by incoming evidence. Implicit in Gendler's examples is the idea that aliefs are, in a sense incorrigible; that is, they are unresponsive to evidence. But to show this we would need to be sure that it was not an artifact of the lack of content of the alief; in other words, it can't be because the content is not propositional, for of course non-propositional contents aren't sensitive to evidence—how could they be when they aren't even truth apt! Instead, what would make an alief a natural kind is to (e.g.,) show that the functional role of an alief, it's lack of responsiveness to incoming evidence, is apparent in cases where the aliefs do contain propositional content.

What would it mean for an alief to be unresponsive to evidence? To a first approximation it would mean that the content of an alief is changeable, if at all, by something like 'habit' (where habit can be operationalized as something like the process of extinction in a reinforcement paradigm). The idea is that aliefs would differ form beliefs in the following way. The functional role of the beliefs make them so that their existence is contingent on the status of incoming evidence; disconfirming evidence can drop one's credence in a proposition, or in the limit, destroy a belief, whereas disconfirming evidence (more or less) leaves an alief untouched.[22]

However, the situation is not just that simple for the idea that beliefs change in response to disconfirming evidence has come under some attack recently (see Boudry and Braekman 2011, Huddleston 2010, Mandelbaum 2010). I will not attempt to show that

[22] For a thorough defense of the claim that beliefs are essentially responsive to evidence, a claim that Gendler is sympathetic to, see Adler (2002).

beliefs are completely immune to revision from disconfirming evidence (though it is worth noting that Gendler never attempts to show that aliefs are completely immune to disconfirming evidence). Instead I will just canvass a few venerable findings in the psychology of belief that make it reasonable to cast doubt on intuitive, yet perhaps extreme views about the malleability of beliefs in response to disconfirming evidence. For example, take the confirmation bias, a type of motivated reasoning (in one of its guises)[23] where people form beliefs through a biased informational search strategy with the end result of reaffirming their already held beliefs as opposed to objectively viewing new evidence.[24] A particularly nasty version of the bias can be seen in the 'biased assimilation' paradigm (e.g., Lord et al. 1979). In such experiments subjects show an effect whereby encountering equivocal disconfirming information actually makes people *more* confident in their beliefs. However, it is not just equivocal disconfirming information that affects one's beliefs in a seemingly paradoxical way. The fruitful research program of cognitive dissonance theory has made a living in part showing how the strength of one's beliefs tend to increase as disconfirming information mounts. The program is founded on the principle that disconfirming evidence *hurts* and because it hurts people are motivated to avoid disconfirming information (for evidence for the phenomenological claim see Galinsky et al. 2000). When this policy of avoiding disconfirming information does not work (because, e.g., the information is not avoidable), people will tend to increase their credence in the disconfirmed beliefs. The

[23] Of course, sometimes it is just the name of a positive test search (such as in Klayman and Ha 1987); that use of the phrase is orthogonal to our purposes and should be set aside.

[24] This can happen in different ways: sometimes by discounting the new evidence, other times by merely avoiding it (as in the 'selective exposure' literature).

operative principle appears to be something like this: the more important a belief is to someone (the more someone self-identifies with a belief, see e.g., Thibodeau and Aronson 1992), the more likely disconfirming information will actually increase one's credence in their belief. In the limit, we get cases like cults who believe that doomsday is upon us who somehow increase their credence in their belief system after their doomsday predictions are disconfirmed (see Festinger et al. 1956; for a version of this understood in the selective exposure to information line, see Tumminia 2005). In sum, there is abundant evidence that beliefs do not as a rule get adjusted in a rational manner in response to disconfirming information.

That said, it is intuitively plausible that sometimes we just do adjust our beliefs in response to disconfirming evidence, even if such cases were rarer than supposed. For example, there is no empirical evidence against the idea that people readjust their beliefs based on evidence when they are dealing with a belief that is toward the periphery of their web of belief (and there is a lot of intuitive appeal to the idea that we do so all the time). Thus the plight of the alief theorist is to establish a case where the functional role of an alief does significantly differ from that of beliefs. To do so, would be to find a case where our beliefs seem to change just based on incoming evidence while our aliefs stubbornly persist. Then the skeptics among us would have good reason to embrace robust aliefs as a new creature in the realm of the mental. Until then though, we should be conservative about what we allow into our ontology and view the notion of robust aliefs with some modicum of skepticism.[25]

[25] This paper was written while the author was receiving funding from the American Council of Learned Societies, who is hereby thanked for its generous support.

References

Adler, J. (2002). *Belief's Own Ethics*. Cambridge: MIT Press.

Bem, D. (1967). "Self-Perception: The Dependent Variable of Human Performance. *Organizational Behavior and Human Performance, 2*: 105-21.

Boudry, M., and Braekman, J. (2011) "How Convenient! The Epistemic Rationale of Self-Validating Belief Systems." *Philosophical Psychology*,

Damasio, A. (1994). *Descartes Error: Emotion, Reason, and the Human Brain*. New York: Grosset/Putnam.

Dwyer, S., and Pietrowski, P. (1996). "Believing in Language." *Philosophy of Science*, 63 (3): 338-73.

Festinger, L, Riecken, H., and Schacter S. 1956. *When Prophecy Fails*. MN: University of Minnesota Press.

Fodor, J. (1983). *The Modularity of Mind*. Cambridge, MA.: MIT Press.

———. (2003). *Hume Variations*. Oxford: Oxford University Press.

Galinsky, A. D., Stone, J., and Cooper, J. (2000). "The Reinstatement of Dissonance and Psychological Discomfort Following Failed Affirmations." *European Journal of Social Psychology* 30: 123-47.

Gendler, T. (2008a). "Alief and Belief." *Journal of Philosophy* 105 (10): 634–63.

———. (2008b). "Alief in Action (and Reaction)." *Mind and Language* 23 (5): 552–85.

Institutional support was also provided through fellowships held at Oxford University and Yale University during the writing of the essay. The essay has been much improved by conversations with Tamar Gendler, Ian Evans, Adam Lerner, Alex Madva, Fred Dretske, and the St. Catherine's Philosophy of Mind Reading Group. Special thanks to Joshua Knobe and Susanna Siegel for their extended discussions on this topic.

Huddleston, A. (2011). “Naughty Beliefs.” *Philosophical Studies*. DOI 10.1007/s11098-011-9714-5.

Klayman, J., and Ha, Y. (1987). “Confirmation, Disconfirmation, and Information in Hypothesis Testing.” *Psychological Review* 94 (2): 211–28.

Lebrecht, S., and Tarr, M. (2010). “Defining an Object’s Microvalence through Implicit Measures.” *Journal of Vision*, 10 (7): 966.

Lebrecht, S. Bar, M., Feldman-Barrett, L., and Tarr, M. (forthcoming). “Micro-Valences: Affective valence in “neutral” everyday objects.” *Frontiers in Perception Science*.

Lord, C., Ross, L., and Lepper, M. (1979). “Biased Assimilation and Attitude Polarization: The Effect of Prior Theories on Subsequently Considered Evidence.” *Journal of Personality and Social Psychology*, 37 (11): 2098-2109.

Mandelbaum, E. (2010). “The Architecture of Belief: An Essay on the Unbearable Automaticity of Believing.” PhD diss., University of North Carolina, Chapel Hill.

Nemeroff, C., and Rozin, P. (1994). “The Contagion Concept in Adult Thinking in the United States: Transmission of Germs and Interpersonal Influence.” *Ethos* 22 (2): 158-86.

Newman, G., Diesendruck, G., and Bloom, P. (2011). “Celebrity Contagion and the Value of Objects”. The Journal of Consumer Research, 38: 215-28.

Ramachandran, V. (1988). “Perception of Shape from Shading.” *Nature* 331: 163-66.

Ripley, D. (Forthcoming). “Contradictions at the Borders.” In Rick Nouwen, Robert van Rooij, Hans-Christian Schmitz, and Uli Sauerland, eds., *Vagueness in Communication*. Heidelberg: Springer LNCS.

Roskies, A. (1999). “The Binding Problem.” *Neuron* 24 (1): 7–9.

Rozin, P., Millman, L., and Nemeroff, C. (1986). "Operation of The Laws of Sympathetic Magic in Disgust and Other Domains." *Journal of Personality and Social Psychology,* 50, 703-712.

Rozin, P., M. Markith, and B. Ross. (1990). "The Sympathetic Magical Law of Similarity, Nominal Realism, and Neglect of Negatives in Response to Negative Labels." *Psychological Science* 1 (6): 383-84.

Scholl, B. (2006). "Innateness and (Bayesian) Visual Perception," in *The Innate Mind: Structure and Contents*, eds. P. Carruthers, S. Laurence, and S. Stich. New York: Oxford University Press.

Stich, S. (1978). "Beliefs and Subdoxastic States." *Philosophy of Science*, 45, 499-518.

Strickland, B., Fisher, M., Peyroux, E., and Keil, F. (2011). "Syntactic Biases in Intentionality Judgments." *XXXIII Proceedings of the Cognitive Science Society*

Thibodeau, R. and Aronson, E. (1992). "Taking a Closer Look: Reasserting the Role of the Self-Concept in Dissonance Theory." *Personality and Social Psychology Bulletin*, 18 (5): 591-602.

Tumminia, D. (2005) *When Prophecy Never Fails: Myth and Reality in a Flying Saucer Cult*. NY: Oxford University Press.

Williams, L. and Bargh, J. (2008). "Experiencing Physical Warmth Promotes Interpersonal Warmth." *Science*, 322: 606-7.