# Believe It or Not: On Multiplying Classes of Belief-like States

J. Robert Thompson

Department of Philosophy and Religion, Mississippi State University

This paper explores whether it is justified to add any new taxa concerning informational states to our psychological taxonomy. Such exploration will not lead to a straightforward decision between remaining steadfast with the taxonomic *status quo* and adding only one new taxon. A careful analysis of when one would be warranted in positing a new taxon for informational states will reveal similarly compelling reasons to posit all sorts of additional taxa. As an antidote to such proliferation, I suggest a reinforcement of traditional taxonomies of the mental by allowing belief and a range of extant taxa to play their requisite explanatory roles, thereby obviating the need for the postulation of any novel taxa.

*Keywords:* belief, alief, folk psychology

## 1. Puzzles and prognoses

Humans often behave in puzzling ways. For instance, people may—with no hint of deception or confusion—act in a way that is discordant with what they profess. A set of such behaviors recently introduced by or associated with the work of Gendler (2008a; 2008b) includes cases ranging from the politically charged (e.g., racist behavior by self-avowed egalitarians) to the mundane (e.g., a trembling hiker who, while standing upon a transparent, though sturdy, outcropping over the Grand Canyon sincerely denies that he is experiencing fear). The presence of these *discordant behavioral patterns* is a pervasive aspect of human activity that has not only provoked ancient and contemporary discussions of *akrasia*, but also become critically important in theorizing about similarly puzzling phenomena such as delusion (Bortolotti 2010), confabulation (Hirstein 2005), blindsight (Weiskrantz 1986), addiction (Poland and Graham 2011), and perhaps change blindness (Simons and

*Corresponding author's address:* J. Robert Thompson, J. Robert Thompson, P.O. Box JS, Mississippi State University, Mississippi State, MS, 39762, USA. Email: jrt260@msstate.edu.

*Studia Philosophica Estonica* (2016) **9.1**, 79–110
Published online: December 2016
Online ISSN: 1736–5899
*www.spe.ut.ee*
*http://dx.doi.org/10.12697/spe.2016.9.1.04*

Rensink 2005). In order to situate these puzzling cases within a comprehensive theory of human action, it appears that something needs to give: either the rationality of the agent should be called into question (the *rationality* gambit), the psychological unity of the agent should be called into question (the *disunity* gambit), or the number and nature of informational states should be called into question (the *bifurcation* gambit). This paper focuses on the last of these options, and in so doing, will bracket the other gambits. The bifurcation gambit explains the discordant behavior by postulating two streams of behavior, each arising from different types of informational state. One stems from *beliefs*, the other from some alternative state, perhaps yet to be explained.

Since taking the bifurcation gambit would result in the addition of at least one new taxon to our psychological taxonomy (PT), I will consider the conditions under which it might be justified to add alternative informational states (AISs) to that PT.[1] I will show that if discordant phenomena require an AIS, a careful analysis of when one would be warranted in positing such a state will reveal similarly compelling reasons to posit all sorts of AISs. For those hesitant to embrace such a consequence, I develop a gambit for avoiding such proliferation—a *belief-saving gambit* that stresses the ways that belief and its close kin might be able to handle both the discordant cases and some puzzling phenomena in psychological development. I will conclude that bifurcationists have heretofore offered insufficient reasons to pursue their gambit, especially given the often overlooked, yet nuanced, taxonomy that is already available for those who choose to persist in belief-centered explanations. Novel arguments that stress the limitations of belief-centered explanations will need to be developed if belief-centered explanatory resources are to be unseated and replaced by resources stressing multiple informational states.

## 2. Three gambits

According to the *Rationality Gambit*, the discordant behaviors in question—affirming $p$ while doing something non-$p$-like—need not cast doubt upon whether human minds are relatively unified or signal the need for additional varieties of informational states. Instead, the behaviors show that we should be willing to judge the agents in these scenarios as believing both $p$ and its negation, and as such, lacking in some aspect of rationality (Gertler 2011;

[1] For the purposes of this paper, a state counts as an AIS if it is an informational mental state that is neither a belief or one of its kin—the superdoxastic, the doxastic, and the subdoxastic, as described in §5—nor an informational state like pretense or imagination whose functional roles enable them to be quarantined in an appropriate way from practical reasoning systems.

Greco 2014; Horowitz 2014). Although most theorists agree that not all of the discordant cases may be dismissed as mere cases of irrationality, considerations about rationality may still figure prominently. Perhaps the most interesting cases involving questions of rationality are those recently raised in discussions of *epistemic akrasia*. In these puzzling cases, it seems reasonable to capture the mind of the agent in question as believing something of the following form: *p*, but I ought not believe that *p* (Greco 2014, 201).

We could handle such apparent discordance by pursuing one of the other gambits: perhaps *S* believes that *p*, but holds that she should not believe that *p* in some other regard; or perhaps one fragment of her mind believes that *p*, but another part holds that she should not. However, this is not what the rationality gambit suggests. Despite the fact that many theorists have recently suggested that it might be *rational* to be epistemically akratic, others have argued (Greco 2014; Horowitz 2014) that appealing to different mental fragments or different types of informational states is less preferable than finding fault with the agent's rationality. The roots of the discordant states can be identified (perhaps there are two sources of evidence, one for each clause), but the best explanation will be to attribute irrationality in such an epistemic agent (or at least withhold judging them as rational).

According to the *Disunity Gambit*, another avenue to pursue in these cases is to suggest that such discordant behaviors have their roots in different fragments of a single mind (Egan 2008; Egan 2011; Lewis 1982; Stalnaker 1984). Such a mind does not consist of a unified, uniformly assessable set of beliefs whose joint consistency or inconsistency leads to attributions of rationality. Instead, the persistence of these discordances suggests that each fragment may be assessed in terms of epistemic virtues, but the fragments as a whole may not be fruitfully assessed in terms of those virtues. Not only do we withhold the condemnation of such agents as irrational, there is no need to appeal to any new additions to our PT's categories of belief, desire, perception, imagination, behavioral tendency, etc. in explaining the discordances. We need to recognize that the entities present in the etiology of behavioral patterns are not a larger unified set of beliefs, but a fragmented, compartmentalized collection of sets of beliefs, each of which are cognitively impenetrable to the other sets. The beliefs themselves are not of a different kind (there is no need to posit an AIS), and the individual mind is not to be counted as irrational just because two of these sets may be in conflict. In general, the discordant cases can be handled by recognizing this fact about the human mind.

According to the *Bifurcation Gambit*, such cases are not to be diagnosed in terms of an issue about rationality or a fragmentation of minds, but rather in terms of two different *types* of mental states (Gendler 2008a; Zimmerman

2007). For our purposes, taking the bifurcation gambit suggests that the other gambits are not worth pursuing. The key move is to amend the PT by adding new entities. What matters is that there are actually two streams of behavior flowing from two different types of mental states.

For example, Gendler suggests that we add a new taxon, *alief*, to our explanatory repertoire in order to explain discordant behavior. Aliefs are more developmentally basic, nonconscious informational states that, like beliefs, function to pair with conative states in the production of behavior. To give a quick example, Gendler offers a tidy analysis of disturbing work from social psychology showing that people who are self-professed egalitarians nevertheless show behaviors incongruous with this egalitarianism on the Implicit Association Test. On this exam, people, to their own dismay, reveal that they implicitly associate black and white faces with negative and positive characteristics, respectively. Gendler insists that the best explanation is that their aliefs are racist,[2] while their beliefs are not. Such a distinction between the distinct types of informational states allows the theorist to explain all sorts of discordant behaviors—including many morally significant behaviors—where the agent's aliefs and beliefs appear to diverge.

In what follows, I will explore the reasons for taking this sort of taxonomic move, and the consequences that follow for our understanding of human agency. First, however, I will explore the extant taxonomy that appears to stand in need of amendment. This exploration is necessary because those theorists who pursue the bifurcation gambit are rarely very forthcoming regarding the most critical aspects of the taxonomy to which they would like to make changes, such as whether the states they aim to elucidate are a species of belief or its rival. Hence, I will need to clarify the taxonomy before the scope and nature of the bifurcationist gambit can be understood.

## 3. The varieties of belief

Lycan's (1986, 61) *locus classicus* on doxastic matters warns that "serious work needs to be done if we are to carve doxastic reality at whatever joint(s) it may actually have." In a recent study, Engel (2012, 18–19) fills in various members of this *doxastic family* of informational states and further differentiates the role belief plays amongst that group. Once one begins probing into the doxastic family there seems to be some consensus that there is (or needs to be) more than one category of states. But, beyond that agreement, the deficiencies of the basic doxastic state of belief pull the theorist in two directions. Belief seems too epistemically sophisticated to explain some phenomena, but

[2] If aliefs are not truth-apt, as non-propositional, then a parallel judgment of their racist nature will need to be provided, a point made by Mandelbaum.

not sophisticated enough to explain others. We need to make some headway in figuring out just when and why we should countenance *any* or *all* of these states.

For example, unlike Gendler, Engel focuses on informational states that occur at the other end of the informational state spectrum. In particular, he emphasizes the features of *acceptance* as an example of a more inflated informational state. According to theorists such as Bratman (1992), Cohen (Cohen 1989; Cohen 1992), and Lehrer (Lehrer 1990; Lehrer 2000), acceptance has the essential characteristics needed in more stringent accounts of knowledge, i.e., when mere belief is not enough. For these theorists, acceptance involves the conscious, voluntary sifting-through of one's justifications for some claim. Although these theorists see a critical role for acceptance in such acts, they differ in how they articulate this state. Lehrer is adamant that acceptance, though not necessarily belief, is evidence-sensitive, and that acceptance is a relatively stable feature of some proposition held over time, whereas Bratman and Cohen see acceptance as featuring in the justification of occurent informational states, as figuring into instances of planning, reasoning, or deciding, rather than in terms of standing or dispositional informational states.

Engel suggests a continuous taxonomy where there is no sharp distinction between accepting and believing. But, he also notes that the majority opinion has been that belief and acceptance are distinct in that they inhabit distinctive functional roles:

> But most views imply that belief and acceptance are different functional states. On the paradigm notion of belief, belief is a dispositional or functional state, involuntary and passive, sensitive to evidence, and inferentially promiscuous, whereas acceptance is most often understood as a conscious mental act, tied to the speech act of assertion, sensitive either to evidence or to pragmatic goals, and expressing a commitment in future doxastic deliberation. (Engel 2012, 20)[3]

Are there varieties of belief itself, and how do they compare to alief, acceptance, and other similar notions? Theorists have been quick to generate these various notions but not thorough in situating them within extant categories.

A reasonable place to start is by distinguishing between beliefs that are held explicitly, and those that are not, i.e., those that are only held tacitly or

[3] Mandelbaum, in correspondence, inquired as to why acceptance's inferential promiscuity is not explicitly compared to that of belief in Engel's discussion. I take it that acceptances, *qua* acts, will be prone to appear in current and future deliberation, which seems like the parallel feature to the inferential promiscuity of functional states, when applied to something like an act.

implicitly by the agent in question.[4] It is not at all clear how this explicit-implicit distinction should be drawn, but it is quite clear that it must be drawn before one decides whether to add more taxa of informational states. A natural way to elucidate and distinguish categories like *explicit* and *tacit* (or *implicit*) beliefs would be by stressing that the former is an *act* of mental judgment while the latter is merely a *disposition* to judge. This elucidation should be resisted, however. Not only is there linguistic evidence against such a demarcation within the class of beliefs—Lycan (1986) stresses the point that 'believe' is a state-verb, not an event-verb (see also Fodor 1981, Vendler 1972)—but some theorists claim that occurent states should not be classified as beliefs at all (Schwitzgebel 2002), or that there is at least a case to be made that standing beliefs should serve as the paradigmatic cases of belief (they also constitute the largest class of beliefs) (Gertler 2011). Hence, in limning the doxastic, we should maintain a firm contrast between acts of judging and beliefs proper (Dennett makes a similar point in 1978, 303), where the latter are taken to be standing states of a creature, often cashed out dispositionally (for a summary, see Schwitzgebel 2015), and we should conclude that there is a need for something like an explicit-implicit distinction to be drawn *within* the class of standing states.

Nevertheless, there is a clear role for verbally avowed judgments to play in a theory of action, where that judgment is either rendered as the result of an active sifting of evidence for a particular situation or as something stemming from the standing beliefs of an agent (as in a transparent report of the standing state). Hence, it seems as though we need some account that includes acts of acceptance, judging, or something like it, as a member of the doxastic family. I will label it as a '*super*doxastic state', the state that is most closely tied to our verbal behavior and is held to the highest epistemic standards, i.e., such acts of acceptance stand as our all things considered judgments.[5]

Regarding beliefs (standing or occurent (if some occurent states are to count)), a distinction can be drawn between the explicit and the implicit va-

[4] If nothing else, there seems to be a distinction between beliefs that are currently accessed consciously and those that are not. Although there have been philosophers who have insisted that all beliefs must be conscious, most theorists will allow that beliefs need not be currently accessed consciously. If they do insist upon this, they are most likely thinking of something like acceptance, rather than belief, but these precise distinctions are what is at issue.

[5] Although it is important to stress that acceptance is an act or event for some theorists, there is also a superdoxastic state that is achieved and can be attributed to the agent who accepts or judges that $p$ at some moment $t$. For ease of exposition, and since the paper is not about acceptance but more basic doxastic elements, I will sometimes refer to the phenomena (the act and its aftermath) involved in acceptance somewhat loosely as a 'state.'

rieties in rough functional terms. One might distinguish the explicit beliefs as ones where the informational state is *actually* generated and/or verbalized, or one might mark their explicitness in that they are more readily *poised* to achieve such a status than their implicit kin. Implicit belief states, then, differ from explicit ones in that they are informational states that serve as sources of action that are not actively generated or assessed, consciously accessed, or verbally reported, but are still generable, accessible, and reportable, should circumstances warrant.

Another phenomenon that an account of non-explicit belief is supposed to handle is the sense in which an unbounded number of beliefs seem attributable to a reasonably rational agent. Notably, these include at least some of the beliefs that are entailed by their other beliefs. The generation of these beliefs would seem to require just that—the *generation* of those states through inference. Hence, in some sense, these beliefs are at an even greater remove in terms of being poised to achieve explicitness than those held without the need for further inference, i.e., the implicit ones. Gertler (2011) distinguishes between these non-explicit states, noting that some are discovered or uncovered pre-existing states whereas others are brought about through some sort of additional inferential procedure. In the end, it seems appropriate to say that the agent *believes* all of these things—those that are present explicitly, to be sure, as well as both those that the agent is well-poised to bring to explicitness and also those that are capable of being inferred as the (somewhat obvious) consequences of the agent's other beliefs (for complications, see Lycan 1986). If we were to establish some terms for these different types of state, we could categorize them as *explicit*, *implicit*, and *tacit*, respectively, and they constitute the doxastic states (narrowly construed).[6] To do so would generate the following taxonomy of the doxastic family (Table 1):

[6] 'Doxastic' is used throughout in both a general sense (as encompassing the overall family) as well as a more specific sense (as encompassing the narrower classes of explicit, implicit, and tacit beliefs). It should be clear from the context which sense is intended, and where it may not be as clear, I have added modifiers to clarify whether it is the narrow or broad class.

Table 1: The Doxastic Family

| Category | Level of explicitness | Level of inferential integration | Level of verbal reportability | Level of accessibility to consciousness | Types of states/events involved | Examples |
|---|---|---|---|---|---|---|
| *Super-doxastic* | Explicit | Inferentially integrated to highest degree | Verbally reportable | Consciously accessible | Acts of judging; Acts of acceptance | The defendant is guilty; I like living in Boston |
| *Doxastic* | Explicit | Inferentially integrated to a high degree | Verbally reportable with immediacy | Consciously accessible with immediacy | Standing beliefs; Occurent beliefs | All races are equal; This cat is black |
| | Implicit | Inferentially integrated to some degree | Somewhat poised for verbal report | Somewhat poised for conscious access | Standing beliefs; Occurent beliefs | All races are equal; This cat is black |
| | Tacit | Typically not yet inferentially generated, so not yet integrated | Verbally reportable, once generated | Consciously accessible, once generated | Propositions that are "fairly obvious" entailments of (more) explicit states | 1567885 is an odd number; Bob Zimmerman is a folk singer |

Many puzzling behaviors may be understood by conceiving of intentional action using these extant classifications, yet one rarely finds them mentioned in contemporary discussions of discordant behavior. Neither does there appear to be any widely cited dismissal of such classifications, so contemporary readers must first decide whether the bifurcation gambit is just an exercise in renaming these classifications in novel jargon (with or without the goal of reinvigorating these decades-old discussions), or whether there is some need for genuinely *alternative* informational states (i.e., alternatives to those laid out in Table 1). If contemporary bifurcationists advocate the latter path, the need for such alternatives should be established by identifying critical cases in which Table 1's taxonomy is shown overtly to be unacceptable. After all, the fact that someone's professed judgment that *p* conflicts with his implicit standing belief that not-*p* is not, on its own, a sufficiently mysterious case to posit a unique category of informational state. The truly alternative state must establish itself with proper contrast to these existing elements of the taxonomy.

Although the distinctions captured in Table 1 manifest themselves in folk or commonsense explanations of behavior (though not always labeled as such), perhaps it is better to relinquish the pretense that the PT is a *folk* psychology. The current task, then, will be to elucidate a more regimented PT that is rooted in these folk notions. Such a taxonomy will include taxa for desire, intention, emotion, and the like. Such a taxonomy will also include taxa for informational states like pretense and imagination. The issue that faces the bifurcationist is whether the belief-like taxa (the doxastic, broadly construed) need more clarification or precisification than what is found in Table 1. Is an additional AIS required?

## 4. Shaping the bifurcation criteria

Table 1 includes varieties of belief as well as a superdoxastic informational state that stands in some close relation to our professed judgments. There is some need, stressed in contemporary applications of the bifurcation gambit, for a state on the other end of the doxastic spectrum. Although often overlooked in these contemporary discussions, Stich (1978; 1983) has already drawn just this sort of useful "distinction which separates beliefs from a heterogeneous collection of psychological states that play a role in the proximate causal history of beliefs, though they are not beliefs themselves" (Stich 1978, 499).[7] He describes the latter collection as involving *subdoxastic* states and offers us valuable criteria that allow us to judge both whether any psychological state $\psi$ should count as a belief or not, as well as what is entailed in that classification. The two crucial properties that beliefs possess and subdoxastic states seem to lack are the now-classic properties of *access* and *inferential integration*.

According to Stich, adults normally have access to their beliefs and are able to subsequently report these states as their beliefs. But, the presence or absence of such access is more nuanced than it might first appear. For example, there may be cases where we are willing to countenance beliefs that are not *immediately* reportable (e.g., if there is some mechanism actively blocking access, perhaps, so long as that mechanism can, in certain cases, be circumvented). However, Stich (1978, 505) explains that we would "be much more reluctant to countenance a special category of beliefs which are by nature not open to conscious awareness or reporting. It is quite central to our concept of belief that subjects under ordinary circumstances have access to their beliefs." Hence, when there is some delay in access, we will want to be able to distinguish cases where the state is *retrieved* by the agent from cases where being prodded about the issue resulted in the *formation* of a previously absent state. When the prodding provides a state that is accessed by the agent in awareness and then reported upon, this product should count as a belief, but we must distinguish that belief from the subdoxastic state itself (which is not the state directly displayed, but rather one of its causes). The prodding "serves rather to instigate a process of belief formation in which, perhaps, the pre-existing subdoxastic state plays a role" (Stich 1978, 506). In these cases, belief formation, rather than retrieval, seems to be involved, and the class of subdoxastic states are not available for report. Beliefs, however, are normally available for report, even if they are not reported.

[7] He discusses the need for these states by raising examples dealing with depth perception, perceived attractiveness in others, and language processing. Stich's actual aim is to show, *contra* Gilbert Harman, that the informational states under discussion are not beliefs, but something else.

The other feature stressed by Stich is that the subdoxastic states have limited relations to other mental states, when compared to full-fledged beliefs. This point captures something akin to Evans' (1982) *generality constraint*, or what is sometimes described as *inferential promiscuity*. In an especially apt passage, Stich explains how we are to view beliefs as compared to subdoxastic states:

> Now it is my contention that part of the reason we are intuitively inclined to say subdoxastic states are not beliefs is that subdoxastic states, as contrasted with beliefs, are largely inferentially isolated from the large body of inferentially integrated beliefs to which a subject has access. This is not to say that subdoxastic states do not play any role in inference to and from accessible beliefs, but merely that they are inferentially impoverished, with a comparatively limited range of potential inferential patterns via which they can give rise to beliefs, and a comparatively limited range of potential inferential patterns via which beliefs can give rise to them. (Stich 1978, 507)

Importantly for our purposes, the distinction he draws is *not* one between complete isolation and full integration, complete abstinence and complete promiscuity. The difference is a matter of degree, for he finds it implausible that *only* beliefs can enter into inferences. There are, then, inferential connections between subdoxastic states and both the mental states that bring them about and those that are brought about by them. Consider the distinct relations to a (propositional) visual content, *d*:

> The subdoxastic state can lead directly only to a restricted class of beliefs about apparent relative depth (and perhaps some other aspects of the visual field). By contrast, the belief [that *d*], if supplemented by suitable additional beliefs, can lead to just about any belief....There is also a striking contrast in the ways other beliefs can lead to either the subdoxastic state or the belief. A subject might inferentially acquire the belief [that *d*] in numerous diverse ways [through testimony, logical inference, etc.].... On the other hand, it is most likely the case that there are *no beliefs at all* which can lead inferentially to the subdoxastic state that represents the fact that *d*. (Stich 1978, 509–510)[8]

Subdoxastic states will play essential roles in belief formation *and* maintenance, though the relationship seems to be asymmetric.

Evans elucidates this idea of limited inferential integration by exploring how competent speakers might cognize their theory of meaning (in this

[8] Importantly, though beliefs alone may not lead to subdoxastic states, this can occur in limited ways when combined with other subdoxastic states, "Similarly, when a subdoxastic state can result from an inference with beliefs among the premises, the range of beliefs that can serve in this capacity is restricted and specialized" (Stich 1978, 507).

case, usually thought of as a Davidsonian truth theory) (Davies 1989; Evans 1985; Miller 1997). He describes this knowledge of language in terms of its contributory role to other *projects* the agent might pursue, "the information is not even potentially at the service of any other project of the agent, nor can it interact with any other beliefs of the agent to (whether genuine beliefs or other 'tacit' beliefs) to yield further beliefs" (Evans 1985, 339). Integrating the two theorists' terminology, such knowledge of language may be inferentially involved in every intentionally produced and understood linguistic utterance, yet because it is limited to that project only, such states must be subdoxastic.

Any theorist who hopes to supplement the layers of the taxonomy developed in Table 1 must be quite clear about which features distinguish his or her preferred informational state(s) and explain why these features cannot be captured in the other layers. These additions are not to be identified with belief states, implicit, explicit, tacit, or otherwise nor are they states of imagination or pretense. The truly novel state needs to be a different *sort* of entity, and the bifurcation gambit is an attempt to add a new taxon to this taxonomy. Given Stich's insight (bolstered by Evans), we can conclude that there is an explanatory need to posit at least these sorts of subdoxastic states that can be utilized only in a limited number of projects. Because they are isolated to this extent, they differ from beliefs, tacit or otherwise. Assuming that our best models of the mind will include at least some modules,[9] it is likely that such modules will contain subdoxastic states; yet, these states need not be limited to modules. Such an addition leads us to the taxon in Table 2.

Table 2: Addition to the Doxastic Family

| Category | Level of explicitness | Level of inferential integration | Level of verbal reportability | Level of accessibility to consciousness | Types of states/events involved | Examples |
|---|---|---|---|---|---|---|
| *Subdoxastic* | Non-explicit | Inferentially integrated in limited or constrained ways | Not verbally reportable | Not consciously accessible | Modular informational states; meaning axioms or theorems | T-sentences; Many may not be verbalizable |

I will argue that with this single addition the *belief-centered PT* resulting from Tables 1 and 2 (hereafter 'BPT') suffices to explain the sorts of relevant discordant behaviors captured above, as well as some other puzzling

[9] This requires that at least some of the representations in these systems are to some degree informationally encapsulated and opaque to conscious access (we need not take any deeper commitments about modularity than that the representations in these systems possess these two features (Carruthers 2006; Fodor 1983)).

developmental phenomena. Hence, there is no need to posit alternative or additional informational states.

## 5. Is an additional state required?

If we set aside tacit beliefs and acts of judgment, we may focus upon the nature of informational states that are part of the causal production of behavior—states that are present in the agent that join with conative states to produce one or more behavior streams. Our question becomes: In addition to explicit beliefs, implicit beliefs, and subdoxastic states, is there a need for *another* informational state?

In perhaps the most compelling recent bifurcation gambit, Gendler (2008a; 2008b; 2011; 2012) insists that there is a pressing need to posit a novel AIS if we are to get a handle on discordant behavior:

> Without such a notion…either such phenomena remain overlooked or misdescribed, or they seem to mandate such a radical reconceptualization of the relation between cognition and behavior that traditional notions like belief seem quaint and inadequate. In short, I will argue that if you want to take seriously how human minds really work, and you want to save belief, then you need to make conceptual room for the notion of alief. (Gendler 2008a, 642)

Hence, if a bifurcation is required, we should seriously consider her rich and provocative account:

> So what is alief? To have an alief is, to a reasonable approximation, to have an innate or habitual propensity to respond to an apparent stimulus in a particular way. It is to be in a mental state that is (in a sense to be specified) *a*ssociative, *a*utomatic and *a*rational. As a class, aliefs are states that we share with non-human animals; they are developmentally and conceptually *a*ntecedent to other cognitive attitudes that the creature may go on to develop. Typically, they are also *a*ffect-laden and *a*ction-generating. (Gendler 2008b, 557)

Gendler remains avowedly agnostic about her project, insisting that instrumental utility in explanations warrants the identification of states that have, for the most part, gone unnoticed in folk psychological explanations (e.g., Gendler 2008a, 642; 2012, 809). Aliefs seem well-suited to describe the causes of behavior that remain resistant to change, that diverge from avowed beliefs, and that might emerge early in phylogeny and/or ontogeny.

Now, in order to motivate the bifurcation gambit, the distinction between any postulated AIS and those in the BPT needs to be sharply honed. Although Gendler explicitly distinguishes aliefs from states of imagination (much of her other work explores imagination) and from *some* features of beliefs, for whatever reason, the voluminous discussion of alief and "in-between" states offered by Gendler and others has largely failed to engage with earlier attempts to map the doxastic family into sub- and superdoxastic states (Mandelbaum 2013 is an exception). Hence, an initial reaction to her suggestion of adding this new taxon might be a call for her to explain why it's not simply the case that the progressive racist in the Implicit Association Test either (a) *accepts* that the races are equal, and yet *implicitly believes* them to be unequal, or (b) believes them to be equal and yet manifests behavior that is driven by subdoxastic states that are outside of her conscious awareness. It seems preferable, for reasons of parsimony, to try in some detail to account for discordant behaviors by using the taxa that are needed for other reasons, especially since they seem able to account for the phenomena fairly straightforwardly.

Gendler's response to (a) might be to insist that whatever beliefs do under their implicit guise, it will not suffice in the progressive racist case. According to such an interpretation, even implicit beliefs must be access*ible* and report*able*, even if unexpected or unanticipated. Supposing for the purposes of debate that whatever informational states are driving the racist reactions are not accessible or reportable, even upon reflection, these states must be non-beliefs.[10] Gendler's response to (b) is harder to anticipate. Since, as far as I can tell, she never mentions Stich's account of subdoxastic states, we are left unsure whether aliefs are simply a subspecies of Stich's state or if they are an additional sub-state consisting of special properties not shared with subdoxastic states. I suspect it is the latter, but the details are worth considering when making the taxonomic decision confronting us.

It is important to stress that on Stich's account, the differences between doxastic and subdoxastic states is not best understood in terms of the information or content of the two states. They are *both* purposed with carrying information for the organism. And, crucially, that role of information bearer requires both states to be sensitive to evidence, even if the effects of this sensitivity are not consciously available, as such. Stich's subdoxastic states also need not have any affective component. Hence, the presence of an *affective* component or a difference in the *type* of content involved might distinguish aliefs from subdoxastic states.

Gendler has said plenty about what makes aliefs unique: Alief is too *hyperopaque* to be either a belief state or an imaginative state[11] and too infer-

[10] One assumes that the individuals involved are not merely pretending to be egalitarian and are being truthful when insisting that they do not in some way harbor such views or are of "two minds" regarding racial issues.

[11] Gendler (2012, 807–808) now notes that the notion of hyperopacity needs more development before it can play this identifying role.

entially shallow to be a belief; it has a content with an affective component, is tied very closely with a behavior pattern, and is unanalyzable into other FP elements. Although alief is a nascent notion, it has already undergone significant poking and prodding, with seemingly every one of its features challenged by critics (Albahari 2014; Brownstein and Madva 2012; Currie and Ichino 2012; Doggett 2012; Egan 2011; Kung 2012; Kwong 2012; Mandelbaum 2013; Nagel 2012; Schwitzgebel 2010). Egan has already given a fairly thorough outline of how a collection of other positions could account for her conception of aliefs, concluding that "we already have some independently motivated and less radical, theoretical tools available for explaining these sorts of phenomena" (Egan 2011, 67–68).

To attempt to rehabilitate her attempted bifurcation from these critiques —and thereby identify possible core features of the AIS—consider her example of how a subject might respond to being the target of a charging bull. According to Gendler, the responses appear to fall into three basic varieties: purely reflex-based, alief-based, and belief-based. A critical divide is between merely reflexive responses, in which the representational content plays no etiological role, and alief- and belief-based responses, in which a particular sensitivity to the content of the informational state plays a role, i.e., the stimulus does not produce the response in an unmediated way. Setting aside mere reflexes, Gendler explores the nature of this informational sensitivity by posing the following query: "Is there additional information about the bull's charging that could change my desire, and thereby change the action-propensity?" (Gendler 2012, 806). If no such informational update could be made, then the two components—the conative and the informational—are not really separable after all, and hence, constitute a substantial unity. Gendler insists that such tight coupling is only to be found in aliefs, and not with belief-desire pairs.

Here we have all of the critical features of Gendler's AIS on display. Its representational content is not altogether irrelevant, but the presentation and sifting of evidence is incapable of breaking apart the conative and the informational unity and swapping in any new information. Belief's sensitivity (and, likewise, alief's insensitivity) to norms and evidence, is critical:

> One—and only one—of the two behavior-generating attitudes can turn on a dime in this way, even in the face of apparent sensory evidence to the contrary. This gives reason to treat the two as not being on a par.
>
> Indeed, the argument can be made on the following simple grounds: Beliefs change in response to changes in evidence; aliefs change in response to changes in habit. If new evidence won't cause you to change your behavior in response to an apparent stimulus, then your reaction is due to alief rather than belief. (Of course, there are strategies for

> changing aliefs as well—but these run through sub-rational mechanisms.) (Gendler 2008a, 566)

Aliefs are molar, shallow, or not fully combinatoric in that their representational content "is not fully integrated with other representational content that has been simultaneously triggered by features of my internal and external environment" (Gendler 2012, 802–803). Aliefs are insensitive to evidence, I take it, because they are isolated to this extent. Unlike mere reflexes, the entities that make up these fused unities *can* be disassociated and then recombined over time. But, in alief, stimulation leads to the affective and action-initiating components in a unique way that may not be interrupted by new evidence, intentional control, or intervention by the agent, hence giving them their automaticity. They cannot be merely *informed* or *directed away*. Hence, it is not the *number* of inferential connections, but the fact that the informational state is cemented to the conative state in a way that even a defeater for that representational state as dramatic as a belief to the contrary will not serve as a defeater of that alief. The only processes that can defeat an alief, i.e., sever its connection to the conative state that is inevitably action-directing, is through some sort of non-inferential process, perhaps like conditioning.

Gendler often highlights associative features of alief. For example, the (primary) changes that take place in aliefs are not evidential, but associative.[12] They are not brute-associative, as in reflexes or instincts. But this fact about how to go about changing or regulating aliefs is a feature that should not be overlooked because it is critical for us, as both theorists and citizens, to know that *whatever* states are driving some (say, racist) habitual behaviors are not going to be changed by the sifting through of evidence. This tells us both how to fix them and what type of state they are. Aliefs are alleged to be evidence- and norm-insensitive.

But, it is not clear that this imperviousness to change via evidence is a feature of some state itself rather than a feature of the epistemic or psychological mechanisms that operate over the states, so the precise role of association in Gendler's account is in need of clarification. Mandelbaum (2015a) has recently argued that associative features arise in at least three ways in accounting for discordant behaviors like implicit biases, and each of these three associationist features are independent of one another. An associative process could explain *how* the bias is learned; an associative process

[12] I qualify here with 'primary' because some theorists, e.g., Mandelbaum, suggest that inferential or evidential changes might change alief-like states in *some* manner, but would not affect their most central or salient properties. This paragraph and the preceding one have benefitted from reading Mandelbaum's (2013; 2015b; 2015a) insightful work in these areas.

could explain thought *transitions* (the process of thinking); or an associationist *structure* might arise in the content itself. Hence, if someone wants to explain implicit bias as based on implicit associations, the nature of this relationship must be qualified—what state is the cause of the racist behavior, and how is association involved in its etiology?

In any account of discordant behavior, it is critical to deal with each associative feature independently because each feature may manifest itself uniquely. Getting clearer about alief's associative features will be critical when assessing its taxonomic viability. For, it may turn out that upon analysis, if alief has fundamentally associative elements, aliefs will, ironically enough, be *ill*-suited to deal with things like implicit bias. The reason for this odd result is that, as Mandelbaum (2015a) argues, there is good reason to think that associative features *cannot* explain the processing and structure of the thoughts or attitudes underlying discordant behaviors like implicit bias. Hence, not only do phenomena like implicit bias not *require* an alternative informational state like alief, but according to Mandelbaum, discordant behaviors' unique sensitivity to evidence and imperviousness to change are better explained with belief-like structures rather than some substantive associationist alternative. The central question is whether the *apparent* differences in sensitivity to evidence and seemingly automatic deployment are sufficient motivation for positing a novel state, i.e., making a *taxonomic* change.

The problems get more vexing for this sort of bifurcation gambit. Both Mandelbaum (2013) and Doggett (2012) have pressed Gendler to clarify *which* features of content and affect are truly unique to aliefs, and several critics (Brownstein and Madva 2012; Mandelbaum 2013) have warned about cleaving the two informational states primarily in terms of their evidence-sensitivity. These critics concede that the sort of states Gendler discusses might exhibit a certain amount of evidence-insensitivity, but not to the extent stipulated by Gendler: "the possibility that such aliefs can fail in *perfectly familiar* contexts shows that they are not ballistic causal reflexes but legitimately norm-sensitive responses, which are, no matter how well honed, always capable of getting things wrong" (Brownstein and Madva 2012, 425). Schwitzgebel pinpoints this stipulation as a central motivation in rejecting alief and making the way for his alternative account of *in-between* beliefs:

> Gendler's main argument against treating habitual and automatic responses as central to belief is this: Beliefs, by their nature, are meant to track the truth and to change in response to evidence. Aliefs—that is, arational, automatic, or habitual response patterns—do not, she says, change in this way. They change in response to (though maybe she should say they are partly constituted by?) changes in habit (2008b, p. 566). [...] This line of reasoning, it seems to me, considerably

> overdraws the distinction. Our habits, associations, and automatic responses *are*, to a substantial extent, responsive to evidence; and our verbal avowals or dispositions to judge are often *un*-responsive to evidence. (Schwitzgebel 2010, 539)

He explains that such bifurcation "artificially hives off our rational and thoughtful responses from our habitual, automatic, and associative ones.... [Gendler] attempt[s] to separate what is an inseparable mix" (Schwitzgebel 2010, 540). Brownstein and Madva (2012, 411) insist that aliefs can figure into normative patterns—these states may produce certain reactions automatically, but they can still "play an integral normative role in the guidance of action."[13] At bottom, it seems that informational states, no matter what state-type, seem to vary in terms of the features Gendler identifies, albeit to greater and lesser extents, whether compared across or within state-type.

## 6. Lessons: Bifurcation or bust?

A successful bifurcation must accomplish two goals in order to establish the viability of its novel taxon. First, the taxon must be distinguished adequately from its rivals. Second, the taxon must possess a certain level of explanatory gravitas. In this section, I will argue that neither of these goals has been accomplished in discussions of alief. Although nothing put forth here excludes the very possibility of some other successful bifurcation, since alief is by far the most promising suggestion for an alternative taxon, these challenges leave the bifurcationist with no plausible candidate state and remove a critical barrier that might have stood in the way of pursuing the belief-saving gambit.

Best seen as a conservative response to any specific bifurcation gambit, the belief-saving gambit seeks both to undermine the motivation for adding the alternative taxon as well as pave the way for an account that sets itself apart from the other gambits by stressing the centrality of belief and its doxastic kin in explaining both discordant and nondiscordant intentional behavior. Stating this gambit roughly, belief—and not some drastically distinct AIS—can explain the phenomena under discussion. Undermining the explanatory power of an alternative taxon allows for the re-establishment of the explanatory power of belief states and allows for further development of the BPT that obviates the need for radical shifts or re-envisionings of the roots of intentional behavior. In what follows, I expand the considerations of §5 showing that alief has not been properly distinguished from its rivals,

[13] Brownstein and Madva also discuss the possibility of what Arpaly calls "inverse akrasia"—cases in which the aliefs are actually "*more* attuned to the demands of the situation than beliefs" (Brownstein and Madva 2012, 411).

and hence, is not a coherent proposal for an AIS. I then question the explanatory motivation and payoff in postulating the taxon of alief. At bottom, this section will diminish the call for alternatives or additions to the BPT.

As argued in §5, alief is not a well-defined taxon, and in particular, it is not properly differentiated from either beliefs or subdoxastic states. The overarching problem with such attempts to bifurcate—and a primary reason to prefer a belief-saving gambit to a bifurcating one—is that most of the relevant features that are used to demarcate beliefs and their kin are graded, not absolute. As Stich and Evans explain, inferential shallowness and evidence-sensitivity can be (and often are) attributed as a matter of degree, as can automaticity.[14] Upon analysis, informational states at the levels below the superdoxastic have extremely similar features; they often only differ in the degree to which they have them. Moreover, any attempt to drive a wedge between types of informational states by defining one type in terms of a total absence of these features might leave us with a truly distinct state, but one with no explanatory power. Such a state would be inferentially inert, impervious to all evidence, and fully automatic, etc.; it would be, in effect, a reflex.[15]

A more looming worry, however, is that even if Gendler's criteria can be clarified sufficiently to mark a coherent class, the motivation for such a state is still far from clear. Part of the problem in assessing Gendler's proposed bifurcation is that it is not systematically compared to the fine-grained taxa

[14] Automaticity comes in degrees; a task or reaction can be more or less automatic, and even more problematically, "automaticity may not be a single concept in the sense that manifestations of automaticity (such as nonawareness, nonintentionality, efficiency, and noncontrollability) are not aligned, meaning that there are examples of processes that are automatic in one sense but not in the others" (Keren and Schul 2009, 539).

[15] When faced with these difficulties in articulating the bifurcating criteria, bifurcationists might carry on pursuing the gambit, aiming to establish non-arbitrary thresholds for meeting one of the state-indicating criteria discussed above, or perhaps aiming to identify or sharpen further criteria. After all, Gendler offers so many 'a'-like features that perhaps one will survive the myriad critiques, thereby leading to a fruitful bifurcation brought about by the identification of that bifurcating criterion. The proof of this will be in the pudding.

One option that first appears promising is the idea that some level of conscious awareness might be the best sorting criterion. This might pan out, and Stich agrees that we would at least be *reluctant* to call something a belief that is closed off from access or reporting. But, this too seems inadequate as a proper sorting criterion, since beliefs can be tacit or implicit, and aliefs can be conscious (Gendler 2008a, 644).

Another way to identify a difference might be to limit aliefs to modular systems. But, I do not think that we can view all of the behavior Gendler appeals to as being the result of modular activity (and, more generally, there are reasons to doubt that whatever entity is responsible for behavior could be massively modular, cf. Fodor 2000). In any case, it does not appear that the elements Gendler attributes to aliefs are thought (by her) to be merely those captured by theories about modularity.

captured in the BPT. As we have seen, there are independent motivations for positing the BPT taxa, and when pieced together jointly, they seem to manifest the features Gendler seeks to explain by adding aliefs. Yet because Gendler neglected to articulate the extant taxa's shortcomings or contrast them with alief, the burden of proof must shift back to Gendler (and other bifurcationists) to explain what these taxa lack, and why this deficit warrants the addition of an altogether new, previously ignored, informational state, i.e., to explain that there is sufficient reason for "wheeling in the big gun of a new fundamental taxonomical category for mental states" (Egan 2011, 67). Put bluntly, there is no positive reason to add to Tables 1 and 2 when it appears that belief and its kin will do.

In order to establish an adequate level of explanatory gravitas, a taxon should display its fecundity (Kitcher 1982) both in situations noted by its proponents and in situations not noted by its proponents but that appear readymade for its application. What we should *not* expect to see for the suggested taxon is either that the taxa of the BPT are straightforwardly applicable in such situations, or that the novel taxon's inapplicability in some such situation is only resolved by further bifurcation, i.e., by adding yet another similar yet distinct state in order to handle the case where the taxon fails to apply. These outcomes would suggest, respectively, either that new taxa do not seem to be required at all, or that a *bi*furcation is insufficient and that the taxonomic shortcomings of belief and its kin require the addition of many distinct AISs each possessing a limited domain of application. On the latter scenario, each additional AIS would have to be distinguishable from *its* rivals and display its fecundity across diverse situations. An inability to apply in situations for which it would have seemed readymade will initiate a similar series of evaluations, perhaps yielding the need for yet another similar yet distinct informational state to handle the cases in which the newest taxon could not be applied, and so on.

Alief is touted as a fundamental AIS in that swaths of mysterious intentional behavior are supposed to be explainable via the postulation of aliefs. Yet, its application appears to be oddly or artificially limited in two troubling ways. First, the cases for which Gendler thought the postulation of aliefs was necessary appear instead to be explainable by using belief-like states. Second, the cases in which one would have thought that aliefs would suffice (as the obvious alternative to belief) actually require some state *other than alief*. This leaves the utility of alief unclear, and the fact that beliefs seem to be involved in the cases where aliefs fall short serves to stress the centrality of belief-based explanations and support the reasoning behind the belief-saving gambit. In the next section, I will make this challenge to alief's fecundity more concrete by examining the role alief might have been able to serve in developmental psychology. In so doing, I will show that the notion of alief is not able to handle the sorts of ontogenetically early phenomena it was designed to explain, and that the failure of alief to handle these cases presents us with a choice of either adopting yet another AIS to explain them, or proceeding as the belief-saving gambit suggests.

## 7. Convergence lost

Young children present an interesting population in which to explore the taxonomic issues raised above. Since they seem to possess informational states that join with conative states in bringing about certain behaviors, it makes sense to ask whether these states should be categorized as *beliefs*, *aliefs*, or some other AIS. A vast amount is known about the ability of preschoolers to understand and attribute belief states (rather than some other informational states) to agents, but unfortunately for our purposes, very little has been said directly about how and when children come to possess beliefs. Nevertheless, research into the processes of understanding and attributing mental states provides some insight into the types of informational states that play a role in the mental lives of children.

Until the last decade, it was widely (though not universally[16]) accepted that children under the age of four understood the behavior of other human agents without the benefit of a full-blown concept of belief. This was taken as a given because almost all the evidence accumulated to that point, and the meta-analysis examining it (Wellman et al. 2001), found that it was not until that age that children were able to pass the Standard False Belief Task (Wimmer and Perner 1983) at above-chance levels. In a version of this now famous task, children witness a puppet show in which a character places a piece of chocolate in a (closed) cupboard and leaves the stage. Another puppet enters and switches the location of the chocolate to a (closed) basket. The original puppet reenters the scene and the child is asked where the original puppet will look for the chocolate. According to the picture supported in the meta-analysis, children younger than four (on average) failed to grasp the *representational* nature of belief—that other agents would have beliefs (about the location of the chocolate) that misrepresented reality, i.e., they failed to grasp that the agent's beliefs were false and that these misrepresentations of reality would manifest in the agent's behavior. Younger children answered that the puppet would look for the chocolate in its current location, whereas older children were able to express that the puppet would look for the chocolate where he falsely believed it to be (in the cupboard).

[16] Some theorists offered evidence of some grasp of belief emerging at younger ages, typically in three-year-olds (Carpenter et al. 2002; Clements and Perner 1994).

It seems that children under four understand the behavior of others as resulting from some nonrepresentational informational state, i.e., one other than belief. Gendler introduces alief in order to play the role of precursor to belief, "they are developmentally and conceptually *a*ntecedent to other cognitive attitudes that the creature may go on to develop" (Gendler 2008b, 557). So, if alief is truly a fecund taxon, one would *expect* alief to aid in explaining this phase of development. Indeed, there are at least two potential roles for alief to play. First, although psychologists generally assumed that the children (say three-year-olds) who failed false belief tasks nevertheless possess beliefs themselves, there is at least some reason to doubt, upon reflection, that these are *real* beliefs because the children lack the concept of belief. Taking inspiration from philosophers like Davidson (1982), theorists might insist that a child cannot be properly attributed beliefs until she fully grasps what beliefs are. Second, aliefs might play a role in the children's mindreading practices, as aliefs closely resemble some of the notions psychologists developed to describe the child's take on the nonrepresentational informational states involved in this more basic form of mindreading, such as *preliefs* (Perner et al. 1994) or *registrations* (Butterfill and Apperly 2013). Such children, then, could be alievers but not believers, possessing the precursor to belief and even using attributing aliefs to other agents, but not yet possessing beliefs or understanding that others have beliefs.

The existence of such developmental contexts—in which informational states are involved but where there is genuine doubt as to whether the states are *really* beliefs (as opposed to some distinct informational states)—bolsters the case for introducing a taxon like alief. These children would seem to act according to an alief-desire psychology, and to the extent that they engage in social cognition, would attribute, at best, an alief-desire psychology to others rather than a belief-desire psychology.

A recent wave of empirical results, however, suggests that even very young children (as young as 10 months) are not limited to alief-desire psychologies in these ways. Rather, these children already seem to grasp that other agents have beliefs (including false beliefs). Starting with Onishi and Baillargeon's (2005) groundbreaking results, teams of researchers have radically re-envisioned the standard timeline for development and illuminated the rich mental lives of children in these stages of development (Baillargeon et al. 2010; Caron 2009; Low and Wang 2011; Luo 2011; Southgate et al. 2010; Southgate et al. 2007; Surian et al. 2007). By posing versions of the sort of change-of-location false belief task described above in which eye-gaze or other forms of attention are measured (rather than verbal response), these results have indicated that much younger children display that they expect an agent's behavior to reflect the presence of *representational* informational states, i.e., they expect behavior to reflect an agent's false take on reality rather than the way the world actually is or the way the child represents it. One might at first resist describing what these children master as an understanding of belief. Perhaps what they have mastered is a more nuanced understanding of alief. But, the results from these studies involve states that are more evidence- and norm-sensitive than Gendler allows. Indeed, Gendler makes it clear that aliefs *cannot* be understood representationally:

> Aliefs involve habitual responses to apparent actual stimuli, but things may not be as they seem, the world may change, and one's norms may demand that the way things are is not the way things ought to be. *Aliefs by their nature are insensitive to the possibility that appearances may misrepresent reality, and are unable to keep pace with variation in the world or with norm-world discrepancies.* By contrast, beliefs are (modulo error) responsive to the way things are: not merely to the way things tend to be or to the way things seem to be. Actions generated by beliefs are generated by a mental state that is proportioned to all-things-considered evidence and subject to rational and normative revision; actions generated by aliefs are generated by a mental state that is not...So it should come as no surprise that human animals are rife with (the tendency to manifest) belief-discordant aliefs, and that our non-human counterparts are rife with (the tendency to manifest) teleofunctional-discordant aliefs. (Gendler 2008b, 570–571, emphasis added)

At bottom, then, Gendler's conception of alief precludes its role in explaining these precocious infant abilities.

Alief's inapplicability in explaining these abilities casts serious doubts upon its fecundity. Almost all participants in these debates will agree that even if these children lack a conception of false belief,[17] they have the ability to track reality-incongruent mental states and are able to predict how those lead to certain types of behavior. So, either the notion of alief will have to be altered in order to deal with this aspect of misrepresentation[18] (which would make it much more belief-like), or the children will have to grasp some other state that does and attribute it to others—either belief or yet *another* informational state. This is an unfortunate result for those trying to demonstrate the utility of alief, since the notion was designed by Gendler to serve as *a*, if

[17] Although some skeptics still challenge the significance of these results by noting that these young children might be relying upon rules linking observable situations to certain behaviors (drawing parallels to similar debates in primatology) (Penn et al. 2008; Perner 2010; Povinelli and Vonk 2004), others are instead trying to decide how to characterize the understanding of these infants in some sort of mentalistic terms (Baillargeon et al. 2010; Carruthers 2013).

[18] Albahari (2014) notes that aliefs could (or should) be able to capture misrepresentation in at least some teleofunctional sense.

not *the*, more basic alternative to belief in cases where it seems implausible to expect beliefs to function (as in these early stages in development).[19]

In other work, (Thompson 2012; Thompson 2014; Thompson 2015), I have argued extensively for a belief-based explanation of these results, i.e., that the most empirically and conceptually robust account of this developmental data supports the claim that these children understand false beliefs and deploy this understanding in robust mindreading. Aliefs cannot apply in these cases, and it looks like beliefs actually can. Hence, we are left with another case where some alternative to belief appears at first to be required, and yet upon reflection, the state that best fits that proposed role turns out to be belief itself. For the purposes of the current paper, this interpretation of the results takes a significant step towards realizing the belief-saving gambit.

Now, it could be that the belief-centered interpretation of the infant results is flawed. If alief cannot be applied in such a case, this may be no reason to fall back on belief but rather another reason to add taxa. In addition to *a*lief and *be*lief, perhaps we should add '*ce*lief' to the taxonomy to capture the states these children grasp and attribute to others during these periods in development. Perhaps we should even add another broad category of the *intradoxastic* for similar cases as they arise, as in Table 3:

Table 3: Alternative Informational State Candidates

| Category | Level of explicitness | Level of inferential integration | Level of verbal reportability | Level of accessibility to consciousness | Types of states/events involved | Examples |
|---|---|---|---|---|---|---|
| *Intra-doxastic* | Implicit | Insufficently integrated/not fully combinatoric/not fully molar; Non-representational | Not (usually?) verbally reportable | Not (usually?) consciously accessible | Aliefs? | Not all races are equal;<br>This bridge is unstable |
|  | Implicit | Insufficently integrated/not fully combinatoric/not fully molar; Representational | Not (usually?) verbally reportable | ?? | Celiefs? | The predator sees the food;<br>The chocolate is in the box |

[19] In an interesting twist, three-year-olds end up producing an example of discordant behavior similar to those Gendler discusses. These children verbally indicate the "wrong" location in the false belief task whereas they nonverbally indicate the "right" location, i.e., they look to the location where the protagonist last saw the chocolate but say that the agent will look in its current location. Crucially, alief fails to offer a straightforward analysis of this discordance. First, it's not clear that one can have aliefs about what another alieves, or aliefs about what another believes, so it appears that beliefs would need to be involved in any meta-representational facets of these scenarios—mere alief will not suffice. Second, if there are two streams of behavior, the one that is alief-based (nonrepresentational) is the one that is closely tied to overt judgment and verbal behavior, whereas the one that is belief-based (representational) is the one that is more reactive and embodied. This seems like the opposite of what one would expect, if aliefs operate as Gendler specifies.

Another option would be to regard aliefs and celiefs as distinct types of subdoxastic states that would differ from the subdoxastic states described above, such as modular states.

But, rather than trying to decide the best category in which to place such additional states, at this point, the belief-saving gambit insists that the postulation of some AIS that is grasped instead of belief is unnecessary and unhelpful in explaining either the precocious abilities or the other discordant phenomena. The belief-saving gambit concedes that several taxa *can* be generated, but argues that these additions are superfluous and mask the role that belief is actually playing in these cases. In this instance, aliefs have their explanatory role once again usurped by beliefs—in this case, in a situation for which basic states like aliefs were readymade. If alief were a fecund taxon, we should not expect belief-based explanations to apply fairly straightforwardly in these cases and to the extent that belief-based explanations might appear shaky, we should not expect yet another state to be required for explaining these sorts of phenomena. Unfortunately for Gendler (and the bifurcationist gambit more generally), these are the outcomes that occur.

Given that a belief-based explanation works, one would have to find substantial reasons to deny the plausibility of that parsimonious account. It seems that, in particular, one would need to elucidate the explanatory advantages gained by passing over beliefs in favor of celiefs, or any similar state that appears as though it might only be useful in explaining such a narrow period of development.[20] Appeals to multiple distinct alternatives to belief in accounting for a range of behavior, each of which apply in some narrow context, are not explanatorily edifying, especially if belief can apply quite well across these contexts. Pushing for further bi- or tri-furcation in the face of belief's success suggests an insufficiently motivated taxonomic pluralism, i.e., this appears to be a case of insisting on multiple overly fine-grained taxa when a more general taxon that sufficiently covers the phenomenon is available.

Above, I noted that Gendler's notion of alief could contribute in two ways in these discussions of development. Although the psychological literature focuses on the child's developing *conception* of the informational states of agents—and it appears that even infants are not limited to viewing behaviors as alief-driven—perhaps alief could still play a role in the other way,

[20] Theorists might appeal to something like celiefs in more contexts outside this narrow period of development, perhaps in the normal application of low-level mindreading both in children and adults. Such an appeal by two-systems theorists (e.g., Butterfill and Apperly 2013) would require a more robust response including reasons to resist seeing such cases of mindreading as involving something like celiefs rather than beliefs (for such a response, see Thompson 2014).

as accounting for the primary state driving the behavior in young children. However, recall that the default position amongst psychologists was that the younger children already possessed beliefs even though they lacked an understanding of false belief. The central resistance to this position was predicated on the claim that a child cannot be properly attributed beliefs until she fully grasps what beliefs are, i.e., until she understands their representational nature.[21] But, since we now suppose they do not lack that understanding, the central reason for resisting the attribution of beliefs to them is now gone. Once they are granted the conceptual sophistication to mindread and can have beliefs attributed to them, the need to generate some alternative state that can explain their behavior view evaporates. There is no additional role for alief to fill.

## 8. Concluding remarks

Bifurcationists who wish to press the issue have two basic strategies at their disposal. One is a narrow approach in which they produce at least a few cases which are so anomalous that they prove to be deeply *belief-resistant*, i.e., that there is no plausible way to (re)interpret them using the BPT. The strongest examples would be cases like the one attempted (albeit unsuccessfully) in §7, in which belief is absent, either due to the existence of developmental disorders, specific or general cases of trauma, or during a particular range in ontogenetic or phylogenetic development. Such invocations would not merely make intentional explanations of behavior smoother, but would render them available at all. Such a strategy may not lead to the wide-scale revision of the roots of human behavior—to the extent the belief-saving gambit works, the number of cases in which an AIS are required will dwindle—but if the AIS is required to describe the phenomenon properly, it will be a success for the bifurcationist. The traditional belief-centered view of the nature of human agency would remain largely intact in such a scenario, so the significance and frequency of these cases will need to be put into perspective.

The other approach is to embrace a fairly radical pluralism of AISs. As aforementioned, the same reasoning that motivates Gendler's addition of aliefs would motivate the addition of celiefs, and it seems plausible that discussions of delusions, for example, could motivate 'deliefs', and so on for many of the categories listed in §1. By avoiding the belief-saving gambit, one may be forced for consistency's sake into adding new AISs for any or all puzzling or discordant phenomena discussed above. Although consider-

[21] Some theorists have argued that certain *linguistic* features must have been mastered for a creature to be competent with beliefs and have beliefs attributed to it (Bermúdez 2003; Davidson 1982; de Villiers 2007). But, most of the theorists discussed here have considered language mastery to be largely orthogonal to mastery of belief.

ations of parsimony and theoretical taste favor a one-state-explains-all approach, Albahari (2014) suggests that the discordant cases themselves are distinct enough to demand different assessments, sometimes in terms of beliefs, sometimes in aliefs, sometimes as "in-between beliefs", etc., or various combinations of these.

In some ways, this approach is similar to the belief-saving gambit in that both appeal to different combinations of informational states in different contexts to explain the causes of the relevant behavior. If there is significant disagreement, it concerns whether the states included need to diverge from those in the BPT. Although I have argued that alief has not been defined sharply enough to assess its applicability in these situations (or based on how it has been defined, aliefs do not appear to be required), the question of what states will be required is an open question.

But, I want to argue that given the reasoning provided above, the bifurcationist must now take a three-pronged approach in order to establish the case for alief or any other divergence from the BPT. She must thoroughly demarcate between the doxastic states that are already present in these creatures (e.g., implicit beliefs) and the posited alternative, explain why the closest taxon in the BPT is deficient in such a case, and explain what the appeal to alief gains over the added cost of proliferating the taxonomy. Going forward, the case *for* alief or any similar bifurcation needs to begin with this sort of approach, and it will be necessary to look at these additions not just locally (as solutions to a particular range of puzzles) but globally (as complications for our taxonomy as a whole), so that we are forced to come to grips with the theoretical and explanatory costs of invoking such multiple distinct states in our ontology.

I have argued that nothing about the current cohort of puzzling cases indicates a critical taxonomic deficiency that will be remedied by adding (perhaps) several AISs. An ongoing rigorous attempt to assess and regiment our BPT will indicate that such additional taxa will continue to be superfluous. The most appropriate taxonomy (BPT), according to my version of the belief-saving gambit, combines those in Tables 1 and 2 and is captured in Table 4 (setting aside tacit states):

Table 4: A Tentative and Traditional Doxastic Family

| Category | Types of states/ events involved | Level of explicitness | Level of verbal reportability | Level of accessibility to consciousness | Level of inferential integration/ Project participation |
|---|---|---|---|---|---|
| *Super-doxastic* | E.g., opinions, acts of judging, acts of acceptance | Explicit | Verbally reportable | Consciously accessible | Participates in almost all projects, when so directed |
| *Doxastic* | E.g., beliefs—standing or occurent representational mental states | Explicit or Implicit | Often verbally reportable in some contexts, but prone to mistake/confabulation | At least *appears* to be consciously accessible in some contexts | Participates in many/most projects |
| *Sub-doxastic* | E.g., modular representations, meaning axioms | Implicit | Not verbally reportable | Not consciously accessible | Participates in limited projects |

Rather than a novel taxonomy, what might be needed to save belief as the recognizable primary cause of most behavior is a different picture of self-knowledge, one that allows for a mismatch between the causes of our behavior, and what we *take to be* the causes of our behavior. Most dramatically, this mismatch arises in cases of confabulation and other discordant behaviors that are rampant in this literature. Recently, Carruthers (2011) has developed an account of self-knowledge that is built to handle these sorts of discrepancies. Future work along these lines will buttress the belief-saving gambit.

In closing, it should be stressed that pursuing the belief-saving gambit will further reveal the facets of beliefs themselves. Schwitzgebel (2010) argues that adopting something like the bifurcation gambit forces one to place some behavior's etiology squarely in one stream or another—automatic/ reflective, evidence-sensitive/insensitive, inferentially shallow/deep, accessible/inaccessible to consciousness—when in practice it will be exceedingly hard to tease out two distinct elements in the mixture that brings about most behavior. Rather than say that there is definitively either a belief that *p* or a belief that not-*p* in such cases, Schwitzgebel chooses to identify the belief that *p* with a cluster of dispositions, some of which will typically be present and some of which will typically be absent (and some of these dispositions will be more automatic, others not, etc.). This belief-saving gambit does not hope to explain seemingly odd behaviors in terms of informational states other than beliefs, but tries to use these cases to show *what beliefs are actually like* (in this case, dispositional).

Schwitzgebel's aim is to reconstruct our notion of belief so as to widen its applicability, *qua* clusters of dispositions whereas Gendler's aim is to maintain something like the classic conception of belief, but restrict its applicability. According to my version of the belief-saving gambit, discordant behavior and the developmental puzzles help bring out the varieties of belief and their relation to the other members of the doxastic family (broadly construed) that have been sketched above. The puzzles presented in the literature teach us something about the varieties of belief, rather than demonstrate its limited applicability.

## Acknowledgments

The author wishes to thank Eric Mandelbaum, Joshua Shepard, Shannon Spaulding, and the referees for *Studia Philosophica Estonica* for their helpful comments on earlier versions of this paper.

## Bibliography

Albahari, M. (2014). Alief or belief? A contextual approach to belief ascription, *Philosophical Studies* **167**: 701–720.

Baillargeon, R., Scott, R. and He, Z. (2010). False-belief understanding in infants, *Trends in Cognitive Sciences* **14**: 110–118.

Bermúdez, J. (2003). *Thinking without Words*, Oxford University Press, New York.

Bortolotti, L. (2010). *Delusions and Other Irrational Beliefs*, Oxford University Press, New York.

Bratman, M. (1992). Practical reasoning and acceptance in a context, *Mind* **102**: 1–15.

Brownstein, M. and Madva, A. (2012). The normativity of automaticity, *Mind & Language* **27**: 410–434.

Butterfill, S. and Apperly, I. (2013). How to construct a minimal theory of mind, *Mind & Language* **28**: 606–637.

Caron, A. (2009). Comprehension of the representational mind in infancy, *Developmental Review* **29**: 69–95.

Carpenter, M., Call, J. and Tomasello, M. (2002). A new false belief test for 36-month-olds, *British Journal of Developmental Psychology* **20**: 393–420.

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

Carruthers, P. (2011). *The Opacity of Mind: An Integrative Theory of Self-Knowledge*, Oxford University Press, New York.

Carruthers, P. (2013). Mindreading in infancy, *Mind & Language* **28**: 141–172.

Clements, W. and Perner, J. (1994). Implicit understanding of belief, *Cognitive Development* **9**: 377–395.

Cohen, L. (1989). Belief and acceptance, *Mind* **98**: 367–389.

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

Currie, G. and Ichino, A. (2012). Aliefs don't exist, though some of their relatives do, *Analysis* **72**: 788–798.

Davidson, D. (1982). Rational animals, *Dialectica* **36**: 317–328.

Davies, M. (1989). Tacit knowledge and subdoxastic states, *in* A. George (ed.), *Reflections on Chomsky*, Basil Blackwell, Oxford, pp. 131–152.

de Villiers, J. (2007). The interface of language and theory of mind, *Lingua,* **117**: 1858–1878.

Dennett, D. (1978). How to change your mind, *in* D. Dennett, *Brainstorms*, MIT Press, Cambridge, MA, pp. 300–309.

Doggett, T. (2012). Some questions for Tamar Szabo Gendler, *Analysis* **72**: 764–774.

Egan, A. (2008). Seeing and believing: Perception, belief formation and the divided mind, *Philosophical Studies* **140**: 47–63.

Egan, A. (2011). Comments on Gendler's "the epistemic costs of implicit bias", *Philosophical Studies* **156**: 65–79.

Engel, P. (2012). Trust and the doxastic family, *Philosophical Studies* **161**: 17–26.

Evans, G. (1982). *The Varieties of Reference*, Clarendon Press, Oxford. Edited by J. McDowell.

Evans, G. (1985). Semantic theory and tacit knowledge, *in* G. Evans, *Collected Papers*, Clarendon Press, Oxford, pp. 322–342.

Fodor, J. (1981). Propositional attitudes, *Representations*, MIT Press, Cambridge, MA, pp. 177–204.

Fodor, J. (1983). *The Modularity of Mind: An Essay on Faculty Psychology*, MIT Press, Cambridge, MA.

Fodor, J. (2000). *The Mind Doesn't Work That Way: The Scope and Limits of Computational Psychology*, MIT Press, Cambridge, MA.

Gendler, T. (2008a). Alief and belief, *Journal of Philosophy* **105**: 634–663.

Gendler, T. (2008b). Alief in action (and reaction), *Mind & Language* **23**: 552–585.

Gendler, T. (2011). On the epistemic costs of implicit bias, *Philosophical Studies* **156**: 33–36.

Gendler, T. (2012). Between reason and reflex: Response to commentators, *Analysis* **72**: 799–811.

Gertler, B. (2011). Self-knowledge and the transparency of belief, *in* A. Hatzimoysis (ed.), *Self-Knowledge*, Oxford University Press, New York, pp. 125–145.

Greco, D. (2014). A puzzle about epistemic akrasia, *Philosophical Studies* **167**: 210–219.

Hirstein, W. (2005). *Brain Fiction: Self-Deception and the Riddle of Confabulation*, MIT Press, Cambridge, MA.

Horowitz, S. (2014). Epistemic akrasia, *Noûs* **48**: 718–744.

Keren, G. and Schul, Y. (2009). Two is not always better than one, *Perspectives on Psychological Science* **4**: 533–550.

Kitcher, P. (1982). *Abusing Science: The Case against Creationism*, MIT Press, Cambridge, MA.

Kung, P. (2012). Review of *Intuition, Imagination, and Philosophical Methodology*, *Australasian Journal of Philosophy* **90**: 806–809.

Kwong, J. (2012). Resisting aliefs: Gendler on belief-discordant behaviors, *Philosophical Psychology* **25**: 77–91.

Lehrer, K. (1990). *Metamind*, Oxford University Press, New York.

Lehrer, K. (2000). Belief and acceptance revisited, *in* P. Engel (ed.), *Believing and Accepting*, Kluwer, Dordrecht, pp. 209–220.

Lewis, D. (1982). Logic for equivocators, *Noûs* **16**: 431–441.

Low, J. and Wang, B. (2011). On the long road to mentalism in children's spontaneous false-belief understanding: Are we there yet?, *Review of Philosophy and Psychology* **2**: 411–428.

Luo, Y. (2011). Do 10-month-old infants understand others' false beliefs?, *Cognition* **121**: 289–298.

Lycan, W. (1986). Tacit belief, *in* R. Bogdan (ed.), *Belief: Form, Content, and Function*, Oxford University Press, New York, pp. 61–82.

Mandelbaum, E. (2013). Against alief, *Philosophical Studies* **165**: 197–211.

Mandelbaum, E. (2015a). Attitude, inference, association: On the propositional structure of implicit bias, *Noûs* **Early View**: 1–30.
**URL:** *http://onlinelibrary.wiley.com/doi/10.1111/nous.12089*

Mandelbaum, E. (2015b). The automatic and the ballistic: Modularity beyond perceptual processes, *Philosophical Psychology* **28**: 1147–1156.

Miller, A. (1997). Tacit knowledge, *in* B. Hale and C. Wright (eds), *A Companion to the Philosophy of Language*, Blackwell Publishers, Oxford, pp. 146–174.

Nagel, J. (2012). Gendler on alief, *Analysis* **72**: 774–788.

Onishi, K. and Baillargeon, R. (2005). Do 15-month-old infants understand false beliefs?, *Science* **308**: 255–258.

Penn, D., Holyoak, K. and Povenelli, D. (2008). Darwin's mistake: Explaining the discontinuity between human and nonhuman minds, *Behavioral and Brain Sciences* **31**: 109–178.

Perner, J. (2010). Who took the *cog* out of cognitive science? Mentalism in an era of anti-cognitivism, *in* P. Frensch and R. Schwarzer (eds), *Cognition and Neuropsychology: International Perspectives on Psychological Science*, Psychology Press, Hove, pp. 241–261.

Perner, J., Baker, S. and Hutton, D. (1994). Prelief: The conceptual origins of belief and pretence, *in* C. Lewis and P. Mitchell (eds), *Children's Early Understanding of Mind*, Laurence Earlbaum, Hillsdale, pp. 261–287.

Poland, J. and Graham, G. (eds) (2011). *Addiction and Responsibility*, MIT Press, Cambridge, MA.

Povinelli, D. and Vonk, J. (2004). We don't need a microscope to explore the chimpanzee's mind, *Mind & Language* **19**: 1–28.

Schwitzgebel, E. (2002). A phenomenal, dispositional account of belief, *Noûs* **36**: 249–275.

Schwitzgebel, E. (2010). Acting contrary to our professed beliefs, or the gulf between occurrent judgment and dispositional belief, *Pacific Philosophical Quarterly* **91**: 531–553.

Schwitzgebel, E. (2015). Belief, *in* E. N. Zalta (ed.), *The Stanford Encyclopedia of Philosophy*. Winter 2011 Edition.
**URL:** *http://plato.stanford.edu/archives/win2011/entries/belief/*

Simons, D. and Rensink, R. (2005). Change blindness: Past, present, and future, *Trends in Cognitive Sciences* **9**: 16–20.

Southgate, V., Chevallier, C. and Csibra, G. (2010). Seventeen-month-olds appeal to false beliefs to interpret others' referential communication, *Developmental Science* **13**: 907–912.

Southgate, V., Senju, A. and Csibra, G. (2007). Action anticipation through attribution of false belief by 2-year-olds, *Psychological Science* **18**: 587–592.

Stalnaker, R. (1984). *Inquiry*, MIT Press, Cambridge, MA.

Stich, S. (1978). Beliefs and subdoxastic states, *Philosophy of Science* **45**: 499–518.

Stich, S. (1983). *From Folk Psychology to Cognitive Science: The Case against Belief*, MIT Press, Cambridge, MA.

Surian, L., Caldi, S. and Sperber, D. (2007). Attribution of beliefs by 13-month-old infants, *Psychological Science* **18**: 580–586.

Thompson, J. R. (2012). Implicit mindreading and embodied cognition, *Phenomenology and the Cognitive Sciences* **11**: 449–466.

Thompson, J. R. (2014). Signature limits in mindreading systems, *Cognitive Science* **38**: 1432–1455.

Thompson, J. R. (2015). Ruling out behavior rules: When theoretical virtues and empirical evidence collide, *Review of General Psychology* **19**: 14–29.

Vendler, Z. (1972). *Res Cogitans*, Cornell University Press, Ithaca, NY.

Weiskrantz, L. (1986). *Blindsight: A Case Study and Implications*, Clarendon Press, Oxford.

Wellman, H., Cross, D. and Watson, J. (2001). Meta-analysis of theory-of-mind development: The truth about false belief, *Child Development* **72**: 655–684.

Wimmer, H. and Perner, J. (1983). Beliefs about beliefs: Representation and constraining function of wrong beliefs in young children's understanding of deception, *Cognition* **13**: 103–128.

Zimmerman, A. (2007). The nature of belief, *Journal of Consciousness Studies* **14**: 61–82.