MANAGING MISMATCH BETWEEN BELIEF AND BEHAVIOR

BY

MAURA TUMULTY

Abstract: Our behavior doesn't always match the beliefs attributed to us, and sometimes the mismatch raises questions about what our beliefs actually are. I compare two approaches to such cases, and argue in favor of the one which allows some belief-attributions to lack a determinate truth-value. That approach avoids an inappropriate assumption about cognitive activity: namely, that whenever we fail in performing one cognitive activity, there is a distinct cognitive activity at which we succeed. The indeterminacy-allowing approach also meshes well with an attractive view of folk-psychology: that ascriptions can help shape the attitudes they report.

Introduction

Sometimes, stretches of our behavior don't match the beliefs that we attribute to ourselves (or that others have attributed to us). Sometimes that's because we've done a good job of lying about what our beliefs really are. A corrupt politician who campaigns on an anti-bribery platform, for example, in fact believes that it is advisable to take all the bribes one can get away with taking. And she also believes one will get away with more bribe-taking if one is falsely believed to be an anti-corruption activist. Her belief in the advisability of bribe-taking, her bribe-taking, *and* her lying about bribe-taking fit perfectly together. The only tension is between all that and the content of the anti-corruption beliefs attributed to her (by herself, and the gullible members of her public). But some cases of belief/behavior mismatch don't appear to stem from deliberate hypocrisy or any form of insincerity. Those cases of mismatch raise questions about the nature of belief and the importance of belief-attributions. This article compares two approaches to such cases, and argues in favor of the one that allows some belief-attributions to lack a determinate truth-value.

That approach avoids an inappropriate assumption about our cognitive activity, and meshes well with an attractive account of the function of folk-psychology. Before outlining the two approaches, it will help to have some examples to consider.

Frances, the forgetful friend: Frances has a close friend, Harry, and she knows that having his birthday remembered is very important to Harry. In the fall and winter, if you had asked her when Harry's birthday was, she would have (correctly) answered, '13th April.' (If you asked her in early April, she'd still have answered that way before exclaiming, 'Good grief! I totally forgot!') But she has been very stressed lately, feeling quite overwhelmed with work, and she fails to plan to send anything to Harry, and fails to call him on his birthday. Speaking at any time after 14th April, can we still say she believes Harry's birthday is on 13th April?

Colin, the committed co-parent: Colin is in a heterosexual partnership and is raising children with his partner in their shared household. Colin and his partner (suppose it's Frances) have agreed that house-work and child-rearing are to be equally shared. Colin participated enthusiastically in the discussions he and Frances had to establish this arrangement, and he talks now about how glad he is that this is how he and Frances chose to do things. But Colin has a hard time moti-vating himself to do house- or child-related tasks when they aren't particularly pleasant. His motivation is especially low when the tasks in question are still strongly culturally associated with the female partner in heterosexual relationships, and have to be done privately. That is, doing them doesn't effectively telegraph to a larger audience that he's a cool, progressive guy. And he's slower than Frances is to recognize when such tasks need to be done. Suppose he would be horrified to have these facts brought to his attention. Colin didn't intend his talk about co-parenting to be mere posturing. But does it express his settled beliefs?

Juliet, the egalitarian professor: Juliet is a white philosophy professor who takes herself to be committed both to a vision of racial equality *and* to a number of subsidiary claims that support that vision. One of those claims is that intelligence is not differentially distributed by race, and that when it appears to be, this is due to the effect of dis-advantages (in, say, nutrition or education) that are distributed by race. Juliet teaches ethics, and works with a diversity training group, and so she talks about these positions, and publicly endorses them, quite a bit. But many of her reactions appear to be affected by racial bias. In the first few class meetings of a new semester, she's more surprised when a student of color asks a theoretically sophisticated question than if a white student does; when a white student's writing

is hard to understand, she is apt to work a bit harder to see if there is a good idea underneath the bad prose, and stick to that task a bit longer than she would if the writer were a student of color; and so on. At least some of these behaviors are marked enough to be noticed by others. Suppose a first-year student who is a member of a racial minority asks an older student (also a member of a racial minority), 'Is she one of those professors who just thinks the white kids are smarter?' How should the older student answer?[1]

An initially appealing description of Frances, Colin, Juliet, and similarly situated subjects, is that they have beliefs on which they don't (or can't or won't) act. We might say: they take the world to be a certain way, and yet they don't entirely succeed in deliberating, acting, and feeling in the ways that would make sense, given that view of the world. The trouble is that in many contexts, what those individuals do (and fail to do) amounts to evidence that they *don't* have the beliefs in question at all. That is, Juliet's grading behavior may not be an oddity, in tension with her egalitarian beliefs: it might be evidence that she doesn't have egalitarian beliefs at all. And perhaps, if Frances *really* cares for Harry that much, she misrepresented her belief when she said his birthday was in April. Perhaps that was a slip of the tongue, and she actually – though falsely – believes it is in August. Folk-psychological standards of evidence appear to allow either relevant attribution in each of these scenarios (for example, that Frances believes that Harry's birthday is on 13th April, and that she doesn't). So which attribution is in fact more appropriate?

Before comparing two tactics for answering that question, we should consider how such questions can be serious. Sometimes, after all, we can feel the pull of such a question for a moment – and then realize that there is an obvious, non-controversial way of answering it. Usually, these answers turn on our acquiring new information about the other attitudes of the subject who puzzled us. Because of the holistic relations between intentional attitudes, and the complicated ways those attitudes drive behavior, the addition of evidence for just one desire or belief to a previously puzzling situation can completely remove the puzzle. For example, suppose Harry grievously offends Frances just a few days before his birthday, and she forms the belief that he doesn't deserve a birthday remembrance. If we are initially ignorant of that belief, we will be puzzled at France's lack of preparation for fêting Harry. But once we learn of that belief, our puzzlement is relieved. We no longer see a lack of fit between her beliefs and any stretch of her behavior. And the addition of the new belief in no way undermines Frances's entitlement to the longer-maintained belief that Harry's birthday was in April. Another possible resolution involves our realizing that the subject has lost a previously held attitude. Suppose we learn that all our evidence for Juliet's egalitarian

beliefs stems from interactions we had with her last year, and that she has since been reading white supremacist literature and enthusiastically recommending it to friends. We would likely think Juliet has changed her mind: she no longer holds the egalitarian beliefs she once did. The behaviors that had puzzled us when viewed in the light of egalitarian belief-attributions fit perfectly with the attributions of racist belief we are now inclined to make to her.[2]

We feel the continued pull of questions about attribution in the light of some stretch of behavior when such resolutions aren't available. Either we don't learn of any attitudes whose presence suddenly reveals the appropriateness (at least in the subject's opinion) of the stretch of behavior that had puzzled us, or any attitudes that look like candidates for reducing our puzzlement raise other concerns. And yet it doesn't seem right to completely abandon the attribution that sets up the puzzle. Suppose, for example, we cast about for beliefs that would explain Juliet's behavior. We wonder: does she have beliefs about the procedures under which students of color are admitted to her institution, such that she believes both that intelligence is not differentially distributed by race, and that students of color at her institution are uniformly poorly prepared for college-level work, making it less likely that they will perform well early on in any class? But we can't find any.[3] So our puzzlement remains. Or suppose we discover that Colin's aversion to doing laundry and washing dishes is due to his asthma: chemicals present even in unscented cleaning products cause him to wheeze horribly. That would look like an uncontroversial puzzlement-remover – until we learn that he has no aversion to washing the car, using a powerful scented detergent marketed for automotive use. In that case, too, we would still feel puzzled, assuming Colin continues (for example) to talk with apparent sincerity about the virtues of co-parenting. (Is his wheezing psychosomatic? If not, why is he more willing to put up with it when doing stereotypically masculine activities?) Developing our examples in these ways illuminates a distinction between merely apparent and persistent puzzlement. That is, some cases that initially appear puzzling do so only because of our lack of access to information that, were we to have it, would easily be counted as puzzlement-resolving. Importantly, both folk-psychological standards and the standards appealed to by parties to the philosophical disputes in which I'm interested would count information of these kinds as puzzlement-resolving. The tricky cases are the ones that don't appear to be solvable in that way. In tricky cases, either the problem isn't merely epistemic, or it isn't such an epistemically simple problem. In the noise and mess of ordinary personal interaction, it isn't always obvious when a case is tricky in the relevant sense. When examples are discussed in this article, whether they are original or are drawn from the literature, they are to be interpreted as falling in the 'tricky' category.[4] Getting a handle on such cases requires not (or not just) adding to the

stock of information attributors have about subjects, but recognizing the particular ways in which the subjects in question have gone awry.

1.   Two kinds of solution to the problem of belief/behavior mismatch

The first kind of response to persistently puzzling cases is to say that there *is* a determinate answer to the question about what such subjects believe. (This family of options can therefore be dubbed the Determinacy View (DV).) Behavior that continues not to make sense in light of that determinate answer is explained by appeal to a subject's inner conflict. That basic similarity covers some dramatic differences. For now, consider just Juliet, and the claim that intelligence is not differentially distributed by race. Does she believe it, or not?[5]

Here are three different answers from three different versions of the DV:

(1):   Conflicting Beliefs:
        Yes, but she also believes that intelligence is differentially distributed by race. That is, she has two contradictory beliefs, formed in different ways, and her behavior is sometimes guided by one, and sometimes by the other.[6]

(2):   Conflict Between the Target Belief and Some Non-Belief State:
        Yes. And she is free of the racist *belief* that intelligence is differentially distributed by race. She is, however, in some other state that has a tight connection to the idea of racial differences in intelligence.

Some versions of option (2) appeal to states that already play a role in folk-psychology. So the conflict is between the non-racist belief and an imagining, or a desire, or a fear that connects to racist content. And Juliet's behavior while she grades papers (for example) is explained by appealing to the way such a state interferes with the ability of her egalitarian belief to guide her behavior.[7]

Other versions share an intuition that standard folk-psychology isn't well equipped to account for the conflict here: new kinds of states must be introduced. For example, Tamar Gendler suggests that an a-rational, associative and automatically generated kind of state she calls *alief* is responsible for producing the behavior that isn't in keeping with the beliefs we attribute (and still should, on her view). On this view, while Juliet *believes* intelligence doesn't differ by race, she doesn't *alieve* it. When she is talking with a student or colleague who appears to be a member of a racial minority, a state with some content like 'Different from me; not smart; don't take seriously' is automatically generated. That state can play

a role in generating action. And because it isn't a belief, we don't have to accuse Juliet of *basic* irrationality. She doesn't have two contradictory beliefs, but rather one belief and an alief that is discordant with it.[8]

Proponents of option (2) are usually motivated by the following thought: belief isn't *just* a pattern of behavior. Belief is tied to our assessments of truth: we believe what we do because we think it is *true*, and we would (for example) give our beliefs up if we came across what we took to be good counter-evidence. But (they continue) belief isn't the only psychological state that can guide or produce behavior in the way that beliefs often do. That's why someone like Juliet can behave as she does without having any racist beliefs.[9]

> (3):   Absence of the Target Belief, Presence of Its Opposite:
> *No*. She has the racist belief. It is what explains her expectations in conversation and her behavior while grading. Her statements about racial equality do not express beliefs of hers, and we need some other story to explain why she makes those statements.[10] That story will appeal to other psychological states she is determinately in. These might be states like desire or fear, states we already recognize and attribute to one another. Or they might be states we don't currently attribute in folk-psychology, so that we would need an argument to convince us that we should admit ascriptions of them to the standard folk-psychological tool kit. But in either case, on this approach, it is Juliet's racist belief that explains her classroom behavior, and a different psychological state (though not, by hypothesis, an egalitarian *belief*) that explains her tendency to make egalitarian claims.

On each of these three options, the question of what Juliet believes is given an absolutely determinate answer. But *whatever* answer is given, some facts about her behavior are odd, in light of that answer. If she really has the egalitarian belief, why does she spend more time trying to figure out whether the white student's sloppy prose concealed great insight? And if she really has the racist belief, why does she spend so much time talking about racial justice? These further questions are answered, in each case, by appealing to one or more additional states Juliet is determinately in. For this family of approaches, the messiness in cases of persistently tricky belief/behavior mismatch is the messiness of *conflict*.

An alternative approach rejects the idea that there is a determinate answer to the question of what Juliet believes. This approach doesn't claim that questions like 'What does Smith *really* believe?' are generally misguided, only that they can't always be determinately answered. On this view, believing – and likewise desiring, fearing, hoping, etc. – is a matter of having dispositions to behave, think, and feel in certain ways.[11] Thus,

believing it is cool but sunny is (among much else) being disposed to wear a sweater and sunglasses; to leave one's umbrella at home; to feel pleased as one looks around while walking outside; to plan hikes but not swims; and so on. There are lots of reasons why someone might fail to manifest one of those dispositions. Someone might believe it will be sunny and still, for example, carry her umbrella about with her – if she needs it as a prop in an acting exercise that afternoon. Some of those reasons explain why someone fails to *manifest* a disposition without undermining the claim that she really *has* that disposition. The need for an acting prop looks like just such an explanation. (She is disposed to leave her umbrella at home, when considering the umbrella simply as protection from rain. And if she finds out her acting class is cancelled, she will leave it at home.) But other explanations of manifestation-failures make it hard to suppose the subject actually has the disposition in question. One can be generally disposed to go to the circus even if one doesn't get tickets for a particular circus performance – if, for example, one was ill. But suppose the explanation for your non-attendance is that you have a persistent delusion that ringmasters are hired assassins, instructed to kill you. That does explain your non-attendance *today*, but only at the expense of suggesting you aren't really disposed to go to circuses at all.

Developing his dispositionalist view of intentional attitudes, Eric Schwitzgebel suggests that it isn't always clear when a particular explanation excuses you for not manifesting a disposition you possess, and when, rather, it indicates you don't have the disposition at all. There will be grey areas. Schwitzgebel suggests we should embrace that fact. In particular, we shouldn't suppose that the cases which appear to have grey-areas – which he calls 'in-between believing' – are cases of agents being determinately in a particular state, one for which we need a new label (such as 'in-between belief').[12] Rather, how things are with these subjects is (usually) best, or at least most carefully, described when we list all the disparate dispositions they have, and refrain from making a summative claim like 'She believes (or doesn't believe) that the various races are intellectually on a par.'

Schwitzgebel does allow that even in these cases, summative claims may be appropriate in certain contexts. What matters are the expectations of one's audience when one makes such claims. The younger student seeking advice about Juliet doesn't really want to know whether she gives money to civil rights groups. The student wants to know how comfortable Juliet's class is for students of color. Since the older student shares that understanding of the situation, it wouldn't be misleading for him to say that Juliet is a racist. But that belief-attribution is only appropriate because of that local communicative context. A general attribution of either racist or egalitarian belief to Juliet would still be neither true nor false.

Suppose we call the view which countenances such in-between ascriptions the In-Between View (IBV). Notice that the IBV takes it that a single

instance of behavior – such as an expression of a judgment – is rarely sufficient to determine what a subject believes. Consider the version of the IBV that proceeds in concert with a broad-track dispositionalist approach to belief. That account acknowledges that most believers are disposed to avow their beliefs in appropriate circumstances. Surely if you believe something, you'll say so if asked, unless disclosure would be harmful. But that is just one disposition among others, and this version of the IBV doesn't give it special status.[13] On this version, a thought (like 'Racial justice is so important') can occur to you; and you can, at a particular time, consciously judge that the races are intellectually equal – without yet counting as *believing* that they are. That's because this approach doesn't privilege *either* rational, intellectual activity *or* 'in-the-world spontaneous behavior' in working out when it is correct to ascribe beliefs (Schwitzgebel, 2010, p. 18). Schwitzgebel asks rhetorically, 'Shouldn't belief be seen as what animates my limbs *and* my mouth, what shows itself diversely in my action and my reasoning and my emotional responses, not just in some pried off subclass of these things?' (ibid.).[14] In-between cases arise when these diverse elements aren't all present, or aren't all pulling in the same direction. Not every IBV proponent will accept a broad-track dispositionalism about the attitudes. So which diverse elements are supposedly involved in believing will vary across different IBV accounts. The key point for any version of the IBV is that the presence of a single element, especially its presence at one particular time, is rarely enough to warrant a flatly true ascription of belief to a subject.[15] When elements aren't lining up, subjects are in in-between states.[16]

On the DV, subjects are fully in whatever states they are in. Messiness at the level of behavior – of the kind that persists in puzzling us – stems from competition or conflict among those fully determinate, fully realized states. On the IBV, a complicated situation like Juliet's is a matter of a subject not *fully* believing something.[17] Juliet is aspiring to a kind of egalitarian attitude, and falling short. On the IBV, we can stop with that description – we don't need to suppose that she is *also* already possessed of some non-egalitarian belief or attitude. On the DV, that's not enough. The failure to display all the behaviors appropriate to the relevant egalitarian belief has to be explained in terms of her being determinately in an additional state.

2.   *Deciding between these two solutions*

Before laying out two reasons for preferring the In-Between View, it must be acknowledged that many attributions of beliefs or other states are determinately true. (If you attribute to Michelle Obama the belief that President Obama was born in Hawaii, your attribution will be flatly

true.[18]) And in some cases where there is a poor fit between some behavior and a possible belief attribution, that is because an additional conflicting state is gumming up the works. In pursuing the IBV, one needn't deny any of that. Rather, taking what I will call the 'pluralist IBV approach' to puzzling cases allows one to make use of some particular explanations that are not available to those who subscribe to the DV.[19] The IBV says that *some* attributions lack a determinate truth-value; the DV says that *none* do.[20] The pluralist IBV position is preferable. Choosing the DV forces us to misdescribe the particular situations of some forgetful subjects, while making a contentious assumption about our cognitive activity. And the IBV is a nice fit with a view of folk-psychology that is independently attractive. Each point is argued in turn.

A. UNMOTIVATED FORGETTING

We tend to think that there is such a thing as unmotivated forgetfulness. That is, especially if we are talking about forgetting we ourselves have done, we tend to assume that sometimes one can forget a fact (that one's friend is allergic to strawberries) or a commitment (that one was supposed to meet a friend at a café at 4PM today) *without* that forgetting depending on one's having previously judged that fact or event to be *unimportant*.[21] So we don't treat all forgetting as motivated by prior judgments that downgrade the importance of the relevant topic. But this fact is in some tension with the Determinacy View. Here's why. Suppose today is Harry's birthday. Recall that having his birthday remembered is very important to Harry, and that Frances cares a great deal about Harry. The fact that Frances is failing to e.g. call Harry, or arrange for flowers to be sent to him, raises the question of whether she really does believe that 13th April is Harry's special day.[22] On the DV, there is a clear yes-or-no answer to that question. But in giving either answer, we need to respect certain parameters: we need to treat Frances as a tricky case, and be careful not to blur distinctions between her situation and the situation of (for example) new acquaintances of Harry's. The trouble is that there is no way to satisfactorily develop either a 'yes' or a 'no' answer to that question while respecting those parameters.

First, if Frances is going to cause us more than momentary puzzlement, it can't be that the explanation for her failure to make plans to fête Harry is due to her recent acquisition of a belief such as, 'Harry doesn't deserve to be celebrated on his birthday unless he apologizes for what he did to me last week.' Secondly, Frances will have some similarities with a new colleague of Harry's who has not even bothered to wonder about his birthday. For example, neither (unless prompted by Harry's announcing 'Today is my birthday') will make any plans to celebrate Harry or wish him well. But there will also be differences between them. Respecting those

differences constrains how the DV proponent can proceed when taking the option of saying that Frances does not believe that today is Harry's special day.

Consider that 'no' answer first (that's the answer that would be given under DV option (3)). There are two ways of proceeding. It could be that Frances fails to believe that Harry's birthday is in April because she has a contrasting belief on the topic of Harry's birthday – such as 'Harry's birthday is in August.' But this isn't satisfying, because it assimilates forgetting too closely to false belief. On the other hand, it could be that Frances fails to believe that Harry's birthday is in April because she fails to have any beliefs at all on the topic of Harry's birthday. Now there is no danger of assimilating Frances to a false believer. Unfortunately for the DV proponent, this option isn't satisfactory either – because it assimilates Frances's situation too closely to those of people who are (for example) just making Harry's acquaintance.

Recall that if Frances's situation is puzzling, then she has said and done things (perhaps as recently as early March) which look like good evidence for ascribing the April belief to her. That won't be true of Harry's new colleague. Even though Frances is failing to make plans to fête Harry, and hence giving us reason to worry that she doesn't believe his birthday is in April, she is in other respects closer to having that belief than is Harry's new colleague. Now, a DV proponent might press us to make use of Robert Audi's distinction between dispositional beliefs and dispositions to believe. Perhaps, she would argue, while Frances no longer believes (even merely dispositionally) that Harry's birthday is in April, she is *disposed* to believe that.[23] It might seem that this enables the DV proponent to safely take the 'no' option in Frances's case, while respecting the differences between Frances and Harry's newest colleagues.

But this move is only an apparent help. A person who is disposed to believe that p does not yet believe that p. She needs to form that belief. But simply thinking of p – by reading p, or hearing p, or being asked a question embedding p – is usually enough to cause her to form the belief that p. If Frances has forgotten that Harry's birthday is in April, and then hears, 'Harry's birthday is in April,' she likely will so believe afterwards. The usually sufficient method would work for her. The trouble is that if 'Harry's birthday is in April' is the sort of proposition one could be disposed to believe, then people other than Frances with that disposition will typically be caused to form the belief only by encountering something that is also an adequate ground of that belief – such as Harry's testimony that his birthday is in April, or the relevant page from his human resources file. Frances, however, may be caused to think, 'Harry's birthday is in April,' and endorse the thought, not only by events that are also adequate grounds. Passing a display of birthday cakes in a bakery, hearing a child boasting about her birthday presents, or seeing Harry's photograph are

also likely to do the trick. The number and variety of experiences that could potentially trigger the relevant thought for Frances suggest that these experiences – even the experience of hearing Harry's testimony – aren't causing the *formation* of a belief, but something else (recall, or re-engagement). So suggesting that Frances is disposed to believe that Harry's birthday is in April, though she doesn't believe it, still won't manage to respect sufficiently the differences between Frances and other people who have yet to think much about Harry.[24] The disposition-to-believe method for pursuing the 'no' option is not any more successful than the others.

If, on the other hand, the answer is 'yes' (as it would be on options (1) or (2)), the DV is committed to claiming that Frances is *also* in a mental state whose content explains her failure to act on her belief. If that content is anything like 'not Harry's birthday today', we are again assimilating forgetfulness to false belief. The content could, however, be something else: it could be something like 'Harry; not important'. But now it turns out that Frances's forgetting is *motivated*. Her forgetting results from a prior evaluation of Harry, or his birthday, as insignificant.

The IBV doesn't have to rule out the very possibility of unmotivated forgetting. If one thinks that our attention isn't always under our intentional control, that our thinking can be swamped by topics of present concern without our having in any relevant sense *allowed* that swamping to occur, one wants the option of simply saying that Frances isn't fully acting as someone who believed that Harry's birthday is today would act. One can say this (or its shorthand equivalent, 'She forgot') without having to suppose that Frances's forgetting *must* be explained in one particular way: by the presence of a different, distinct mental state concerning Harry. Given how often I am plagued by forgetfulness, I appreciate the extra option for understanding my behavior (and for giving accounts of my behavior to others).

One might ask: why can't the DV proponent just note that the relevant content (about Harry's birthday) never occurred to Frances, and leave it at that?[25] Why not say either that she didn't plan to fête Harry because she forgot it was his birthday, and she forgot because the relevant content didn't occur to her – or say that her forgetting just *is* the failure of the relevant content to occur to her? Here is why the DV proponent can't rest with this sort of approach. We typically take the claim that a subject *forgot* something to mean, in part, that a content that ought to have occurred to her (perhaps so as to guide some action) failed to occur to her. We also typically want to know *why* the content didn't occur. A DV proponent who wants to attribute the relevant belief to Frances, but wants to say simply that her forgetting is constituted, or is explained, by its content not occurring to her, needs a way of showing that typical desire for explanation to be inappropriate (at least in this case). But the DV proponent also

needs (a) to distinguish a forgetful person like Frances from a wholly ignorant person like Harry's new colleague; and (b) to explain away the forgetful person's failure to take relevant actions, so that it does not serve as evidence against attribution of the relevant belief. The DV proponent's appeal to a conflicting state is what enables her to perform these tasks. The conflicting state both explains the forgetting and explains *away* what would otherwise be behavioral evidence against the attribution she wants to make. Because the conflicting state does this double duty, a DV proponent can't simply eschew explanations beyond the non-occurrence of content. The IBV proponent, on the other hand, *can* say that the relevant content doesn't occur to Frances, and leave it at that. That statement is part of her description of Frances: as someone who isn't quite being or acting as a believer (in the relevant content) ought to act. The IBV proponent doesn't need an explanation of Frances's being in that in-between state to do double duty. She doesn't need an explanation of why her non-fêting behavior doesn't defeat our positive attribution of belief. After all, the IBV proponent doesn't want to make that positive attribution of belief.

A proponent of the Determinacy View might, however, make a general complaint: that plain old forgetting isn't the sort of thing to which her model is meant to apply. She might argue that therefore it doesn't tell against her model if her model is a poor fit for it. But Frances's situation is quite like the sample cases DV proponents present in their work. They discuss cases of people who have newly acquired some information – for example, that a road they often use will be closed for construction, or that the knives are now on the *other* side of the stove.[26] They are able to articulate that information in response to direct questions. But they find themselves turning down that road, instead of the other, open one; and they keep reaching for the knives in the wrong spot. They declare that something is true, but can't effectively, let alone habitually, act accordingly. Frances's case is no different – except that she only counts as failing to act on her information once Harry's birthday actually rolls around. But for every day after that, she's like the person taking the wrong route. She's unlucky in that many of her daily activities will fail to suffice for prompting her to recall Harry's birthday. She'll have to run into him, or attend a party for someone else, or visit a bakery, and those events may not happen for some time. If you have to drive home from work every day, then at some point *every* day, you will be confronted with the effects of your failure to act on your information. So you won't forget for more than twenty-four hours at a stretch. But you too required a specific prompt, and would not have altered your behavior without it. The DV requires us to explain *all* such cases by the addition of another state. We aren't, even for the purposes of folk-psychology, allowed just to stick with the claim that subjects in these cases are failing to make effective use of information.

There is a more general lesson here. Consider, again, Juliet's behavior while grading. Recall that the various types of DV may differ on whether or not Juliet really has the egalitarian beliefs she professes. But they agree in taking her to have another determinate state (a belief or something else) with racist content.[27] For them, Juliet's failure to act like someone with egalitarian beliefs is grounded on a *success*: her success in having, or being in, racist states. The proponents of the DV assume that someone who says, 'She doesn't fully believe that the races are intellectually equal' has not explained Juliet's behavior so much as marked a need for it to be explained. That much may be true. But they also assume that the *only* thing that could explain it would be the attribution of a determinate state whose content was directly relevant to racial egalitarianism. That is, a proper explanation will have to include a positive claim like 'She believes the races are not intellectually equal' – or any other variant we might get by talking about aliefs, imaginings, or other states distinct from belief. Perhaps sometimes this is required for a proper explanation, but it need not always be. We offer these sorts of negative ascriptions of one another all the time: we state that someone doesn't really feel or think something. And we don't always mean that therefore they really do feel or think the opposite. The person who doesn't really think you're wonderful need not be thinking you are not-wonderful, because they need not have gotten around to giving you enough thoughtful consideration in the first place. Or suppose you are reporting on a subject in a change-blindness experiment. The subject is speaking to two people who (undetected) swap in and out for each other.[28] We might say, of the subject, 'She doesn't realize there are two people there' without meaning 'She believes there is only one person there'. In some cases, it might be that were she paying enough attention for that second ascription to be flatly true, she would have noticed the switch (and so abandoned the belief).

One reason for our sometimes counting such negative claims as satisfying explanations is that there are different ways in which one can fail to do something, or fail to do it well. Sometimes when one fails, one succeeds in doing something else. If Harry fails at being faithful in a relationship, it is by virtue of having succeeded in having an affair, or at least a one-night stand. And when a state is defined negatively – like the state of not-being-in-pain – one fails to be in that state by succeeding at being in a relevant contrasting state – i.e. by being in pain. But not all activities and states are such that failing to do or be in them is like that. Noting that you aren't quite managing to carry the tune of 'Tea for Two' does not require us to find some other song that you are managing to sing.

There is a distinction between criticizing someone for doing the wrong thing, and criticizing her for doing the appropriate thing badly, awkwardly, or incompletely. When one has certain roles to play – friend, citizen, teacher, parent – one can play those roles badly, or fail to fulfill

one's responsibilities, without playing some other role instead. (Because it is true that some bad-playing could be tantamount to abandoning the role, we have the grey areas: the cases where it isn't determinate whether someone is in the state or not.) Perhaps some of our cognitive activity is like fidelity: we only fail at it by succeeding at something else.[29] But we can't assume all our cognitive activity is like that unless we counter-intuitively deny the possibility of unmotivated forgetting. So at least some of our cognitive activity will be better modeled on role-playing: something one can do more or less well without thereby doing anything else instead. The capacity to accommodate a fuller spectrum of kinds of cognitive activity is the first reason for preferring the IBV response to tricky cases.

B.   THE REGULATIVE VIEW OF FOLK-PSYCHOLOGY

Appreciating the second reason in favor of the pluralist version of the IBV requires some background. Recent work on the nature of folk-psychology has argued that folk-psychology is not exhausted by description, prediction, and explanation.[30] Some of that work has suggested that we've been over-looking the *regulative* role of folk-psychology. Folk-psychological practices help shape us into the kinds of individuals who are properly described in folk-psychological terms. The descriptions we offer of one another's behavior, and the explanations we offer for puzzling or norm-violating behavior, also act as *prescriptions*. They lay down elements in a large background sense of what is rational, moral, appropriate, and personally coherent. We aim, among other things, to ensure that we are explicable in standard folk-psychological terms, and we sometimes do that by changing ourselves to bring ourselves in line with a description of us that's already on the table.

On this view of folk-psychology, it is expected that we will make attributions, to others and ourselves, that aren't, at the time of attribution, flatly true.[31] We will make, for example, *aspirational* attributions. Those attributions can nevertheless count as seriously appropriate if the subjects in question have a good shot at living up to the aspirations – perhaps, in part, because of their awareness of the attribution. And this fits well with the IBV, which allows attitude attributions to fail to be flatly true without being flatly false. The regulative account of folk-psychology is compelling. It copes very well with the fact that we sometimes get a grip on what others are up to by applying a tacit theory to them, and sometimes by simulating them.[32] It highlights and makes sense of links between folk-psychology and ordinary moral assessment.[33] And it appears to fit nicely with some recent work on the evolutionary origins of folk-psychology, and on the way children develop competence with it.[34]

On the DV model for messy situations, we could only embrace the regulative model of folk-psychology if we allow that folk commonly make

false attributions of attitudes to one another. Importantly, these attributions would count as false by folk-psychology's own standards. Often, the standards in question will be ones to which we ourselves have access. Exactly which folk-psychological standards are accessible, and which are tacitly governing our activity without our being able to explicitly reflect upon them, is a complicated and partly empirical question.[35] But it is clear from our common patterns of conversation that many folk-psychological standards are accessible – both in the sense that we can tell when they are violated, and that we can articulate some version of the standard in question. We correct others whom we take to be violating relevant standards. For example, when someone fails to take an obviously sufficient means to an end for which she has just expressed a desire, we typically construe her mistake as one she could correct, and take it that our telling her something about it could play a role in her correction. We perhaps are most explicit about this in our conversation with children, but we talk this way with adult peers too. Consider expressions of exasperation, with a child and a spouse, respectively: 'Stop complaining! If you really want to warm up, put on your sweater – it is right there on your chair!' Or: 'If you really think your uncle shouldn't be allowed to behave that way without apologizing, why are you inviting him to Sunday dinner?' One would presumably risk accusations of condescension if one spelled out the prescriptive standard in question, or its descriptive corollaries. ('When you want x, and y is an appropriate way to achieve x, follow y, other things being equal,' and 'Most people, when they want x, and see that y is an appropriate way to achieve x, will follow y, other things being equal.') But access to such standards is implied by these particular expressions of exasperation, and countless others like them.

So in a messy case, we will often have access to the fact that a given attribution would be problematic, and access to the standard whose violation (by the person we observe) makes the attribution problematic. Consider an aspirational attribution, such as 'You really want to quit smoking.' The evidence that shows you might find my attribution potentially useful as a motivator – such as your pawing through the bag where I know you still keep a few loose cigarettes – is also what makes it potentially false. Your observable behavior is violating a descriptive folk-psychological claim to which we both have access: namely, that people who don't want to do x generally don't take steps whose only point is the preparation to do x. Since we are capable of avoiding these potentially false claims because we have access to the relevant folk-psychological standards, and since, other things being equal, we tend to prefer making true statements to false ones, the DV proponent who wishes to embrace the regulative view of folk-psychology needs to explain why we persist in making so many false statements.[36] She could attempt to do so by appealing to the utility of such statements. Perhaps, for example, a subject who

hears a flatly false ascription about herself could thereby be brought to realize that the attitude in question is one she ought to have, or one she ought to purge. Folk practice could certainly license ascribers to make that kind of false ascription.[37]

While the DV is technically compatible with the regulative view of folk-psychology, it is an awkward fit for a number of reasons. First, on the DV, the self- and other-regulation of attitudes that folk-psychology encourages has to be directed at maintaining or changing attitudes, where neither effort involves intermediate stages of a kind that would threaten to rob an ascription of determinate truth-value. A person who (perhaps as a result of a realization provoked by an ascription) succeeds at such effort moves, at some specific point, from an ascription's being determinately false to its being determinately true. The IBV, by contrast, can allow that not every case of attitude management and change needs to look like that. Given the wide variety of processes that drive attitude change, this pluralism is more sensible. Second, on the DV, an attitude ascription that is not flatly true of an agent at a time is flatly false.[38] This means that all ascriptions must be counted as *descriptions*, even if some of them have the effect of (say) bringing subjects to notice what they should be thinking. Prescriptions, for example, don't become flatly false simply because those to whom they are made fail to follow them. To count as flatly false, barring unusual examples of false normative claims, ascriptions need to be descriptions. As will emerge more clearly in what follows, this is a real hindrance to an effort to combine the DV with the regulative view of folk-psychology. The regulative view emphasizes the diversity of our ascriptive practice, and the variety of ways we exert regulatory control on ourselves and one another. On the IBV, it is possible for ascriptions to be appropriate – not just pragmatically useful – without being flatly true. Finally, admitting ascriptions lacking a determinate truth value allows for greater variety in ascriptions' status. They can be descriptions, but they can also be prescriptions, exhortations, and admonitions.

That an attitude attribution could be admonitory brings us back to Juliet. 'She believes the races are equal' can be an appropriate, if aspirational, attribution if Juliet is aiming to *better* believe that. In some contexts, however, the IBV licensed attributions of racist belief. They were a convenient short hand for the older student to use in telling the younger student what she really wanted to know. But now we could also imagine cases in which someone might appropriately ascribe that belief to Juliet precisely because he wants to bring Juliet up short, to cajole her into taking a good look at herself, and to work at preventing that attribution of racist belief from becoming *flatly* true of her.

So we can distinguish at least three types of ascriptions in the view of ascriptive practice that emerges from combining the IBV with the regulative model of folk-psychology.[39] There are *conversationally acceptable*

ascriptions of determinate states to subjects in in-between states. These ascriptions are made acceptable only by a speaker's (correct) assumption that her audience will not be misled into having inappropriate expectations for the behavior of the subject in question. They are motivated by an interpreter's desire to save her breath.[40] An ascription is easier to produce than is the list of the subject's fractured dispositions. But once we broaden our understanding of what ascription is *for*, we can see that not-flatly-true-ascriptions can nevertheless be significant for reasons other than securing interpreter convenience.

Such *seriously appropriate* ascriptions may be conversationally convenient, but they also serve one or more of folk-psychology's regulatory aims. Recall Juliet's friend, who wants her to notice conflicts between her behavior and her expressed views, and therefore tells her she believes that the races are *not* intellectually on a par. His ascription is useful, as are other such interventions, because it is likely to provoke change. For some seriously appropriate ascriptions, their appropriateness is exhausted by their usefulness in bringing about changes in the interpreted subject's behavior. But other ascriptions, because of their fit with a subject's aim and trajectory, are also likely to become flatly true. Consider Juliet after she's had her behavioral mismatches brought to her attention, and after she's taken some effective steps toward altering the relevant behaviors. She's aiming to believe *better* in the intellectual equality of the races, and staking herself to that aim – by self-ascribing that belief, or allowing others to make it of her without qualification – will help her achieve it. This aspirational ascription is on track to becoming a true descriptive ascription.

Finally, some ascriptions aim to be *flatly true*, to describe a subject as in a determinate state that she in fact is in. Even such an ascription could be deployed as part of a broader regulatory effort. For example, we might use such an ascription to encourage the formation of additional attitudes rationalized, in the light of new information, by the ascribed attitude.

Even when an ascription is serving a regulative purpose, it can still have a declarative form. This is perhaps clearest with self-ascriptions. One can say, 'I don't want another drink,' or 'I don't mind feeding your cat while you are away' even when one is still *en route* to forming the relevant attitude. (The publicity of such a self-ascription helps us along, even if we didn't make the ascription with an eye to taking advantage of that.[41]) But we also use the declarative form when ascribing attitudes to others even when we aren't (or aren't primarily) describing them. We might say, 'You don't really want more cake,' though perhaps we only risk saying so to close friends and family members. And sometimes even when we are addressing an audience other than the subject of our ascriptive activity, we help to make sense of her by indicating something she ought to be doing. (If she hears us, she may realize what she ought to be up to.)

Gareth Evans has an example of this:

> A young student is reading out an ill-prepared essay to his class. It contains the sentence 'A spark is produced electrically inside the carburettor'. 'That's not right,' the teacher says. 'What does he mean, class?' And here someone may say 'He means the cylinder, sir.' In saying this, the second student is *not* committed to the idea that the subject had the thought, or even has the capacity to have the thought, 'I shall say that a spark is produced in the cylinder' (1982, p. 130).

Nevertheless, talking about the cylinder is what the first student, 'given his general plans and his situation . . . should be doing' (p. 130). That is why the teacher can ask for, and the student can respond with, a *prescriptive* ascription. Even though the ill-prepared student was initially in no position to formulate subsidiary plans involving reference to cylinders, he was hoping to talk sense about engines. His comrade can therefore follow the common practice of ascribing to him an intention he ought to have had. Perhaps, unless the first student is completely clueless, hearing this ascription will actually equip him to form the relevant thought. Evans's point is that it was perfectly appropriate for the second student to make the ascription he did, even during the time when the first student wasn't capable of forming the relevant thought.

Evans was explaining how to cope with apparent counter-examples to his account of demonstrative reference rather than arguing for a regulative view of folk-psychology.[42] But he presents a number of examples of ascriptions in declarative form that are prescriptions or exhortations rather than flatly true descriptions, for such attitudes as 'is thinking of', 'wants', and 'intends [to get or to do]'. In these cases, we are making global sense of someone by talking about what it would most make sense for her to be doing – even when we are fully aware that she isn't doing that, or isn't doing it at all well.

The IBV was supposed to fit well with the regulative view of folk-psychology because it can handle an implication of the regulative view: that we will often make ascriptions that are not flatly true descriptions. Yet much of the folk's attitude-ascribing is conducted with apparently declarative statements. We've just seen that there is no real tension between these two points. There are non-descriptive uses of declarative ascriptions in first-, second- and third-person cases.

C.   AN OBJECTION FROM CONTEXT-SENSITIVITY

A final objection to embracing the IBV could come from a DV proponent who decides to acknowledge that our situations with respect to our attitudes are much less stable than we might suppose. She aims to hold on to the DV, however, by arguing that what might look like in-between cases are rather cases where a subject shifts between two or more fully

determinate attitudes depending on context.[43] This move doesn't ultimately succeed, however. It does highlight another way of finding an apparently tricky case of mismatch to be only apparently tricky. But it still leaves a remainder of tricky cases that are better handled on the IBV.

When we are considering the possibility that someone might have contextually circumscribed attitudes, there are two possibilities. There either are, or are not, folk-psychological normative expectations that a subject's attitudes on this topic be cross-contextually stable.[44] If there are no such expectations, then the variability of the subject's attitude is no bar to ascribing the relevant attitudes to her in relevant local contexts. That is, we can say that a particular subject believes that p in context C, and either believes that it is not the case that p, or simply fails to believe that p, in context C^*. Neither ascription is called into question by the fact that there is no answer to the question 'But what does she believe, full stop?' and hence no *general* attitude to ascribe. Furthermore, assuming the subject behaves within each context in a manner consonant with the belief she holds in that context, we don't have any mismatch between belief and behavior. Someone who thought there was a tension between a belief-ascription she wanted to make, and some stretch of behavior she observed, likely failed to respect the distinctness of the contexts here. If there is no expectation of cross-contextual consistency, then refusing to separate the ascriptions by context is wrong-headed. It is just as wrong-headed, in fact, as it would be to complain that someone's past belief clashed with her current behavior, when it was obvious that she had changed her mind. There are ways of limiting *which* stretches of behavior are supposed to match one's attitudes. Appealing to changes of mind is one way, and appealing to contextual shifts, on the assumptions at work here, is another. When the path to contextually circumscribed ascriptions is smoothed by the absence of an expectation of a general attitude, would-be ascribers face no persistent tension. Any apparent tension is revealed as merely apparent once ascribers follow the appropriate ascriptive practice in this case, and parcel out particular ascriptions among the relevant contexts. There is nothing odd or defective about the subjects to which they are attending.

However, folk-psychology does expect cross-contextual consistency for certain attitudes on certain topics. Consider, for example, attitudes of belief and disbelief for propositional contents like those expressed by these sentences: 'Milk sold in stores is usually safe to drink' or 'Few people enjoy being belittled'. There are two sorts of case that can arise when a subject violates this expectation by having only contextually circumscribed attitudes. First, it is possible that her failure to have the relevant general attitude in no way undermines her claim to have the local ones, and hence in no way affects the truth of local ascriptions. It is important to note that, as with subjects' violations of other folk-psychological expectations for attitudes, we may justifiably find a great deal of fault with this subject's

attitudes while being justifiably certain about which attitudes we ought to ascribe to her. Consider, for example, a subject who behaves fully consistently with her belief that p, where we know that p is false. We may correctly judge that the subject shouldn't believe that p – and yet in most of the common cases of this kind, no one would have any reason for finding it awkward or tricky to ascribe that belief to her.

With that analogy in mind, consider a subject who shifts, depending on context, between believing p and believing that it is not the case that p. Where we properly expect a general attitude, we have a few related appropriate criticisms to offer: she should not have such contextually circumscribed attitudes on this topic; she should (depending on the case) either have the belief that p or the belief that it is not the case that p as her general attitude; and (supposing, for now, that the right attitude to have would be p) she should not have the belief that it is not the case that p as an even local attitude in the contexts in which she does. We can make all these criticisms of the subject, and the criticisms can be fully justified, without our facing any uncertainty about what her situation with respect to her attitudes is. Such a case doesn't involve any persistent mismatch. Any apparent mismatch will be revealed as merely apparent once an ascriber realizes she has been viewing behavior observed in one context as if it were relevant to attitude-ascriptions she wants to make to the relevant subject in a different context. Unlike the case above, where there is no expectation of cross-contextual consistency, the ascriber in this situation who seeks to connect her observations and ascriptions across contexts is in line with relevant folk-psychological norms when she expects consistency. But once she recognizes that the subject is violating those norms, she can spell out the subject's situation without feeling uneasy about any of her ascriptions.[45]

On the other hand, it is possible that a subject's failure to have the general attitude folk-psychology expects on this topic does undermine her ability to possess local attitudes on that topic in a full-blooded way. This is the second kind of case. Consider, for example, the belief that milk purchased in stores is usually safe to drink. If someone violates the expectation that this be a general attitude, her violation wouldn't just show her to be doing a poor job with attitudes on this topic. It would instead suggest she doesn't truly have *that* attitude even in the local contexts in which it is somewhat tempting to ascribe it to her. For an analogy, consider someone who, were she to believe that p, would not only have a false belief, but a belief that so flies in the face of evidence she herself acknowledges that we would wonder if she really does believe that p. Some of what initially looks like irrational-believing that p can be *so* irrational that one wonders 'Does she actually understand what p (or one of its components) means?' Sometimes, the answer to the relevant question is 'no' – the person doesn't fully possess the relevant attitude, if she possesses it at all.[46] Now consider someone who, in certain contexts, clearly doesn't believe that milk

purchased in stores is usually safe to drink. Perhaps she even appears to believe that it is not the case that such milk is usually safe to drink. Barring very unusual sets of circumstances, a subject's tendency to have (what we are sometimes tempted to call) her attitude toward this content vary with context raises serious questions about whether she really possesses any attitude, local or otherwise, toward *that* content. We wonder if she understands the work done by 'usually', for example, or the somewhat complex way 'safe' is used.

In such a case, if we allow ourselves the resources of the IBV, we can identify the subject as being in an in-between state. In some contexts it may be conversationally appropriate to ascribe the relevant belief to her. But this is not a situation in which a subject shifts back and forth between two attitudes, each of which is full-blooded and fully possessed. Rather, the subject has a mixed bag of dispositions, some of which bring her a distance towards full belief that *p* (if that is the relevant attitude), and while that's *all* she has in any context, in some contexts, that is enough to render an ascription of belief that *p* conversationally appropriate (though not flatly true).[47]

So when a subject's situation makes it appealing to suppose that she has shifting, contextually-circumscribed attitudes, there are two possibilities. On the one hand, we can ascribe local attitudes to her, but once we do so, we lose any apparent tension with behavior of the kind that could raise questions about our ascriptions. On the other hand, such a tension continues, but it undermines the propriety of ascribing local attitudes. In that case, however, '*shifting* determinate attitudes' can't be the way we describe the tension. How often any of these scenarios arises is obviously not something that can be settled *a priori*. It is quite likely that we have contextually circumscribed attitudes – both ones that do and ones that do not violate folk-psychological expectations for cross-contextual consistency – more often than we might like to think. But when we do fail to manage our attitudes, this doesn't always make it unclear whether we actually possess the attitudes some ascribers are tempted to ascribe to us. When it is unclear which attitudes we possess, the options remain the ones outlined above: the IBV, and the versions of the DV that rely on cognitive conflict to model the problems with the subject that give rise to persistent difficulties for her interpreters. Precisely because of the interest in mutual correction that is effectively highlighted by the regulative view of folk-psychology, it is instructive to look at the class of cases where the kind of criticism, praise, or correction we ought to offer isn't immediately clear, because it isn't immediately clear what a subject's current attitudes are. When such apparent unclarity is merely apparent, appealing to local, contextually shifting attitudes could be a stable interpretive strategy. But when the unclarity persists, appeals to contextual sensitivity won't resolve it.

Conclusion

As a way to handle the often-messy relation between what we say and what we do, the pluralist IBV approach is better overall. DV proponents are wrong to think that every case of persistent messiness must be explained in terms of cognitive conflict. But they are right that determinately true ascriptions (and the states that make them true) are important. There are folk-psychological norms – some of which extend into folk morality – encouraging matching (1) among our deliberations, actions, and reactions; and (2) between our own and others' perspectives on our lives. No one wants to be the kind of person who (sincerely) talks about racial justice and yet sits farther away from colleagues whose race is different from her own. Such mismatching is an impediment to mutual understanding and so to cooperation. It increases everyone's interpretive work if we can't rely on self-assessments being largely accurate, and such an increased workload would make mutual cooperation more difficult (McGeer, 2007b; Zawidzki, 2008). It has negative practical consequences for subjects themselves: they'll be stymied in pursuit of their own ends by the dispositions that aren't lined up with those ends. Finally, mismatch raises moral concerns about authenticity and whole-heartedness.

When matching of both type (1) and (2) obtains, we are in the states that will license flatly true ascriptions. Folk-psychological norms relevant to such matching push toward greater determinacy of individual attitudes. And on the dispositional view of attitudes, greater determinacy helps secure greater harmony among attitudes. (That's because it will be part of a full dispositional profile for (say) the belief that p that one *not* have the dispositions that are central to believing that it is not the case that p; and that one be disposed to form a certain intention if a belief that p and a desire that q are present; and so on.[48]) We want to be integrated, not least in what we present to the world and what we self-consciously present to ourselves. And others want integration from us. But of course we aren't always so integrated. Flatly true description matters as something we want to make possible. We can acknowledge that while recognizing that many ascriptions aren't even aiming to be those kinds of flat descriptions, and would fail to be true if they were.

Schwitzgebel ends his defense of his version of the IBV with a pragmatic argument: adopting it helps ward off inappropriate self-satisfaction with our beliefs. It makes it harder for our verbal statements to suffice for belief-ascription, and so harder for us to count ourselves as believing what we think we should. However, a proponent of the DV could object that the danger of smugness looms on the IBV approach as well. And the danger might seem especially acute for the version of the IBV put forward in this article, which capitalizes on insights from the regulative view of folk-psychology. The objection would run: on this version of the IBV, we are

allowed to use, and frequently will use, ascriptions that are declarative in form but are not flatly true descriptions. But our tendency to view ourselves in the best possible light will make us wrongly interpret (e.g.) aspirational ascriptions concerning us as already flatly true.

First, even if this objection could not be met, we are in no greater danger of inappropriate self-satisfaction on this version of the IBV than we are on the DV. And second, the objection can be met. It rests on the assumption that hearing a declarative-form ascription of a belief to us – where so believing would be to our credit – could only cause us to cease working to improve the match between our actual dispositions and the profile for that belief. Making that assumption begs a question against the regulative view of folk-psychology. It rejects a key part of that view: that we are motivated to keep our publicly observable behavior in line with the ascriptions, especially the self-ascriptions, that are relied on by others in their interactions with us.

So in the effort to avoid smugness, the IBV still looks like the better bet. The IBV also helps ward off paralyzing defensiveness. We are sometimes reluctant to take responsibility for attitudes that reflect poorly on us, and refuse to engage their consequences. The IBV lets us treat many of the cases in which improvement is required as cases of in-between belief. We can acknowledge that many of us who need improvement don't fully believe (for example) sexist and racist claims. That may make it easier to convince others and ourselves to take the steps needed to achieve the status of fully believing the egalitarian claims we verbally avow. The regulative view of folk-psychology highlights the fact that choosing how to describe ourselves is bound up with how we want to be. The IBV's stringency about flatly true belief-ascriptions, and its acknowledgment that we often fail fully to believe what we wish we did, will help provoke serious self-reflection and appropriate self-improvement.[49]

Department of Philosophy
Colgate University

NOTES

[1] This character and her name are from Schwitzgebel, 2010. I have altered some features of her situation.

[2] A variation on this scenario would give Juliet different beliefs in different contexts – and behavior consonant with the relevant belief in each respective context. Restricting our ascriptive activity to one context at a time, we find nothing to puzzle us. Contextually shifting attitudes present a number of complications, however, and more complex cases are discussed below.

[3] I am grateful to an anonymous referee for pressing me to think about this kind of case; see Note 27 below for related discussion.

[4] While one can describe a case in such a way that the 'tricky' label becomes compelling, it isn't possible to prove conclusively that a particular case is tricky. That is due to the general

difficulty of proving a negative. To prove conclusively that a particular case is tricky, one must show that there is *no* attitude currently possessed by the subject, and *no* change in attitude recently undergone by the subject, that would dissolve an interpreter's puzzlement were she to learn of it. The more information we have about a subject's attitudes, the more confident we may be that the general, negative claim – to the effect that the subject possesses no hidden puzzlement-dispelling attitudes – is true. For a real-life – and emotionally wrenching – tricky case with more detail than would fit in a journal article, consider the autobiographical narrative Mark Pierpont contributes to *Finally Free: How Love and Acceptance Saved Us from the Ex-Gay Ministries* (Human Rights Campaign Foundation, 2000). Pierpont's work as a preacher of 'conversion therapy' based in evangelical Christianity, and his eventual decision to come out as gay, is also discussed in Joshua Knobe's 2011 *New York Times* Opinionator blog post (Knobe, 2011), which argues that our folk assessments of other's 'real' selves are based in our own value judgments about worthwhile lives. Pierpont, other contributors to *Finally Free*, and Knobe's philosophical discussion are all given nuanced 'folk' discussion at the feminist, focused-on-pop-culture blog *Jezebel* (where I first learned of Pierpont's case and Knobe's treatment of it; see Kinkaid, 2011, and the subsequent comment threads). The key point is that neither Pierpont nor much of the 'folk' discussion of these issues assume a simple picture on which Pierpont had (possibly hidden) beliefs and desires that make his behavior over the painful period of his life non-puzzling. His case is treated as tricky: our puzzlement over him isn't construed as simply due to a lack of information. Answering the question, 'Did he ever *really* believe that homosexual desires and acts were gravely sinful?' isn't treated as an easy, just-need-more-information, task. Even those individuals who feel that Pierpont always had homosexual desires don't take it that he himself was perfectly clear about that in the way that would be required either for accusing him of deliberate hypocrisy, or for taking the question of his *beliefs* about the morality of that behavior to be less than complicated.

[5] Compare the list of positions given in Schwitzgebel, 2010. Because I will be emphasizing the role of conflicting attitudes, my categorization of the available options is different.

[6] See, for example, Sommers, 2009. Tamar Gendler (2008b) critically discusses a variety of such approaches. Eric Mandelbaum (2013), in criticizing Gendler (2008a and also 2008b), proposes appealing to conflicting beliefs in tricky cases. While Andrew Huddleston (2012) emphasizes the relationship between first- and second-order beliefs, he too is working with a conflicting belief model.

[7] See, for example, Currie, 2000. Currie is talking about delusions, and not run-of-the-mill cases like Juliet. But his proposals could be extended to cover non-pathological cases. Wilson, Lindsey and Schooler, 2000, discusses conflicting attitudes, focusing on evaluations. Depending on how the view were developed, it could fit either in this category or in category (1) (if evaluations were analyzed as beliefs about the worth of the target of evaluation).

[8] Gendler, 2008a and 2008b. Andy Egan (2008) and Aaron Zimmerman (2007) each propose alternative views that fit in this category.

[9] An especially explicit announcement of this commitment comes in Gendler, 2008c.

[10] This approach is suggested by David Hunter's handling of what he terms 'alienated belief'. Hunter is providing an account of only one kind of belief/behavior mismatch (Hunter, 2011). So he isn't obviously committed to supposing it will be the correct account for every kind of case.

[11] For an especially explicit version, see Schwitzgebel, 2001, 2010. Victoria McGeer endorses such an approach in her 2007a (see also her 2007b). Akeel Bilgrami (2008) could be interpreted as sympathetic to this approach. Schwitzgebel (2010) suggests that one could accept the possibility of in-between belief without accepting a dispositional account of belief, and briefly explains how a sample representationalist and a sample functionalist could do so.

While Schwitzgebel, 2010, only considers multi- or broad-track dispositionalism, a narrow-track dispositionalism that identified believing with being disposed to judge could also be made to fit with an in-between approach. Such a narrow-track dispositionalist would say that a subject has the belief that p if and only if he is disposed to judge that p. On such a view, a subject believing that p will manifest that disposition to judge, other things being equal. Because it isn't always determinate whether things are equal – and hence not always clear when a disposition is present yet not manifest – it won't always be determinate whether the subject believes that p. The contrast with the determinacy view is easiest to formulate with reference to a version of the in-between view that is formulated in concert with a broad-track dispositionalist approach to belief. So that is the version that will be used in the discussion in this article.

[12] For an example of the kind of move Schwitzgebel is *not* making, see Huddleston, 2012. Huddleston focuses on subjects who have (1) a first-order belief that p; and (2) one or more second-order beliefs like 'p is false' or 'p flies in the face of the good evidence I have' and so on. Huddleston argues that in many of these cases the state with content p really is a belief – a belief we ought to criticize its holder for having. He terms these attitudes 'naughty beliefs' and enumerates the features (such as high levels of contextual variability) that make them the 'sort of creatures of the mind they are.' Elsewhere he refers to them as a 'species of belief'. Huddleston argues we will fail to levy the criticism deserved by subjects with naughty beliefs if we take their errors as proof that they are in non-belief attitude states (such as alief) instead. He doesn't engage with the issue of in-between belief, so there is no way to know what arguments he would put forward in favor of the DV. His paper takes very seriously the fact that we are often less rational in our believing than we suppose ourselves to be, and assumes that this fact can be fully appreciated within a DV framework. As will be argued below, some of our failures as attitude-holding subjects are appropriately modeled within a DV framework. But what sometimes happens isn't what always happens. To see the distinctiveness of the in-between approach that (it will be argued) is sometimes superior, it is crucial to recognize that it is *not* the introduction of a new kind of attitude state for subjects to determinately occupy.

[13] Richard Moran argues that 'the attribution of belief to a person involves evidence of various different kinds which can in principle conflict with each other,' and acknowledges that we can ascribe a belief to another person even when she could not avow it, though we are 'under a certain tension' when we do so (Moran, 2003, p. 408). However, Moran clearly rejects the view that 'the ability to avow a belief by reflection on its object is *just* one possible way among others for belief to manifest itself' (ibid., emphasis original). That is why he can't be counted as fully sympathetic to a version of the IBV that would proceed in concert with broad-track dispositionalism. Since (as Note 11 noted) the IBV is compatible even with narrow-track dispositionalism, and Moran doesn't always handle complicated cases as if they resulted from conflicting, fully realized states, I take it that he might nevertheless be sympathetic to some other version of the IBV. See Schwitzgebel, 2011 for an application of his dispositional version of the IBV to the question of self-knowledge.

[14] Compare McGeer, 2007a, p. 90.

[15] Of course, the presence of a single element that could potentially (given the right scenario) move a subject from a determinate to an in-between state will not necessarily do so. Suppose a dispositional account of the attitudes. The disposition to flinch is likely to play a role in the dispositional profiles for several attitudes, including the profile for believing that one is in danger. However, when one flinches in response to the depiction of an explosion in a movie at the theater, that will not be sufficient to make it indeterminate whether or not one believes that one is in danger. Presumably the adult who regularly attends movies believes, in this instance, that it is *not* the case that she is in danger. Her flinching does manifest a

disposition that plays a role in the belief that one is in danger. But it doesn't play an important enough role for its presence to render a subject's state indeterminate. In-between views will have attitudes constituted by multiple elements, and they will rarely give any one element total preeminence. So just as saying 'I believe that the races are intellectually equal' may not suffice to make an ascription of that belief flatly true, the eruption of a flinch does not suffice to make it indeterminate whether one believes one is in present danger.

[16] That the above general point would hold even for a narrow-track dispositionalism may not be immediately obvious. However, consider the disposition to judge and the disposition to verbally express judgments. Even an account that restricted its analysis of belief to those two dispositions can allow in-between states. That's because (as discussed above) it isn't always determinate whether a disposition is present, but excusably non-manifest, and when instead the underlying disposition itself is absent. I am grateful to an anonymous referee for pressing me to clarify this matter.

[17] I am following Schwitzgebel in using the phrases 'fully believing' and 'not fully believing' to indicate (respectively) cases where subjects do, and do not, do or have everything required for the possession of a particular attitude of belief. This use should not be confused with ways of talking about so-called degrees of belief. After all, someone could be in a determinate state of believing that the likelihood of p's being true was 25 per cent. There would be nothing in-between about such a state, and nothing to cause difficulty for ascribers. Consider a different case. Ascribers might find a case initially puzzling, because they ascribe a subject 100 per cent confidence in the truth of p, and yet observe behavior that matches poorly with that attitude. If the mismatch disappeared once we realize that the subject had only 50 per cent confidence in its truth, we no longer have a truly tricky case. To get a persistent puzzle with degrees of belief, one needs e.g. a normally cautious and frugal person who bets her life savings in a scheme whose success depends on the truth of p, but who also reports verbally that she has a low degree of confidence in its truth. Being less than 100 per cent confident in the truth of some claim isn't on its own enough to produce in-between belief.

[18] In the terms of Schwitzgebel, 2001, Michelle Obama has all the dispositions in the folk-psychological stereotype for that belief, and manifests them appropriately.

[19] If we are talking about all puzzling cases, including those cases where initial puzzlement is easily removed by additional information, then a DV proponent can of course be pluralist in many respects. She could, for example, note the presence of an attitude that makes previously puzzling behavior practically rational, in light of all the attributions we want to make to a subject (the subject's attitudinal situation was more complex than we realized). Or she could appeal to evidence that an attribution we've made in the past is no longer true (the subject has changed her mind). She could use these or other resources to show that an apparent puzzle needn't be a persisting puzzle. But consider the cases where puzzlement persists. In those cases, whatever other variety may be introduced by a DV proponent for explaining why we are puzzled, there is one thing she cannot do: appeal to in-between attitude states. That's because the determinacy view is defined by its rejection of the idea that there are any such states. The IBV proponent counts in this respect as more pluralist, because she can appeal (if a case warranted it) to any of the solutions of which a DV proponent could make use, *plus* solutions that appeal to in-between states. In the cases where DV/IBV differences are revealed – cases of persistent puzzlement – the DV proponent always copes by appealing to a conflict between states a subject is determinately in. An apparent exception will be addressed below.

[20] Accepting the dispositionalist model of the IBV will mean that some cases of messiness can't be handled on the conflict model. Schwitzgebel (2010) notes that on his own view, contradictory belief appears impossible. That's because the dispositions required to make an ascription of belief that p flatly true will include dispositions like this one: being disposed to

judge that the claim that it is not the case that *p* is false. But even if contradictory beliefs are ruled out on this version of the IBV, other sorts of conflict – say, between two attitudes of different kinds – are possible. The set of dispositions associated in folk-psychology with a particular attitude doesn't include dispositions to form or refrain from forming every possible related attitude. See Note 48 for further discussion.

[21] For a discussion of (what I am calling) motivated forgetting, see Smith, 2005.

[22] A version of this case could turn instead on the question of whether Frances believes that *today* is 13th April. We could then ask whether she fails fully to believe that, or whether instead one of the DV options applies to her case. See Schwitzgebel, 2002, for the suggestion that many puzzles over uses of co-referring terms in attitude ascriptions should be handled as cases of in-between belief.

[23] See Audi, 1994. I am grateful to an anonymous referee for suggesting this possibility, and for prompting deeper reflection on this issue.

[24] It is actually somewhat unclear whether '*S*'s birthday is on date *d*' is a good candidate for something one is disposed to believe (in Audi's sense), at least with comparison to his usual examples, such as 'This sentence contains more than two words'. It might rather be that we generally have what Audi terms a *capacity* to believe such claims. Someone might then try and suggest that if one hasn't had beliefs about *S*'s birthday before, than one merely has the capacity to believe it is on date *d*; but that if one had such beliefs in the past, one is now disposed to believe it. This would, at least at the level of terminology, provide for a difference between Frances and Harry's new colleague. But it wouldn't answer the question of why so many things – none of which is an adequate ground for the relevant belief – trigger the relevant thought for Frances. There may be cases where we use the term 'forget' appropriately, by folk-psychological standards, and yet the case is well described as a case of a subject who now has (only) a disposition to believe. But that analysis can't do justice to every case of forgetfulness, and it isn't right for Frances.

[25] An anonymous referee asked exactly that.

[26] See Gendler, 2008a; Zimmerman, 2007, and critical discussion of a variety of other views in Schwitzgebel, 2010.

[27] Assuming, again, that Juliet is a tricky case. As an anonymous referee pointed out, it is possible for someone's behavior to have discriminatory consequences even if the beliefs that lead to it are egalitarian (or at least non-racist). But in that case, there is no persistent puzzle about a lack of fit between Juliet's beliefs and any stretch of her observed behavior. She does treat white students differently from students of color, but (on this alternative) it turns out that differential treatment is made practically rational by her egalitarian beliefs in combination with specific beliefs she holds about (for example) admission and enrollment trends at her institution. An observer might wish Juliet would notice the negative effects of her behavior, since those effects have real and undesirable force despite Juliet's freedom from anti-egalitarian beliefs. But such wishing isn't the same as puzzling over what beliefs to ascribe to her. We can wish someone's behavior were different, even when we recognize that it is rendered practically rational by her beliefs.

[28] Such a set-up is among those discussed in Simons *et al.*, 2002. The point of the example is not to suggest that attention, let alone an intention to pay attention, suffices to make one invulnerable to change-blindness. Rather, the point is that if we attribute as flatly true the belief that the relevant change is not happening, we may be making presuppositions about how things are with the change-blind subject that diminish our resources for explaining her blindness to the change.

[29] If one is very expansive about what counts as 'doing', and what counts as an action, then in some sense, whenever one fails to do action Φ, regardless of the nature of Φ, there is *something* at which one is succeeding (displacing air molecules with one's breath, for example).

This doesn't affect the point about two particular kinds of failure at issue in the above discussion, however. In the case where one's failure to Φ is due to (or is equivalent to) one's having succeeded to Φ*, the two actions are more closely related than are (to continue the example) the actions of singing 'Tea for Two' and displacing air molecules. The nature of the relationship will vary, but one way of glossing the relevant closeness is in terms of the second action being a relevant alternative (in that context) to the first. This could mean that an agent attempting the first action might view the second as an alternative she could intend (whether or not she does), or that an observer interested in whether or not the agent manages to succeed at the first action would be equally interested, and for similar reasons, in whether or not she manages to succeed at the second. We can sort cases of failure, then, according to whether or not the failure results from or is equivalent to a successful action that counts (in some sense) as a relevant alternative to the first. Of course relevance may vary to some extent with context.

[30] See Andrews, 2009; McGeer, 1996, 2007b; and Zawidzki, 2008.

[31] McGeer writes that many of our attributions are not in the business of explaining or predicting behavior, but rather of shaping behavior, by means of 'cajoling, encouraging, reprimanding, [and] promising' (2007b, p. 149).

[32] McGeer, 2007b.

[33] McGeer, 2007a and 2007b.

[34] Andrews, 2009; McGeer, 2007b; and Zawidzki, 2008.

[35] There are likely some quite subtle standards governing when we actually make certain kinds of attributions, and when we actually would judge certain attributions false. Consider Joshua Knobe's investigation of how attributions of intention appear to be governed by the perceived moral outcome of attributed action (Knobe, 2003). There are many reasons why Knobe's thesis provoked so much interest, but it seems likely that part of the explanation is that he uncovered an aspect of the folk theory of intention to which we didn't previously have access.

[36] A contrasting case in which such an explanation would not be necessary would run as follows. Suppose future cognitive neuroscience could show that many of the attitude ascriptions we currently make are in fact false. There is no puzzle about why, at the moment, we persist in making them – because the standards showing them to be false aren't available to us. We would have no idea we were making false statements, and our doing so is easily explained by our ignorance.

[37] Here's how this would work in a sample case. Consider the DV proponent who operates with Gendler's version of the DV, and suppose the proponent is considering Juliet, and an attribution to her of racist belief. He will handle her case by attributing to her aliefs with racist content. But he may still attribute to her a *belief* with racist content, while recognizing that that attribution is false. He can see that it is false because he has observed Juliet happily helping to organize diversity training groups, and giving sincere and impassioned speeches about reducing racial bias on campus. By the lights of his theory and descriptive FP standards to which he has access, he needs to label that belief attribution false (because of the extent of the belief-discordant behavior). He can nevertheless make the attribution if it counts as false-but-useful.

[38] It is possible that some versions of the DV might emphasize the pragmatic inappropriateness of some technically accurate ascriptions – for example, 'He doesn't believe that e = mc^2,' said of a six-month-old infant. Someone holding such a view might prefer a less stark statement of the DV position. However, even such a theorist will accept that in most cases, an ascription that isn't flatly true is flatly false, and so I will avoid complicating the discussion by qualifying the starker claim.

[39] Schwitzgebel mentions McGeer, 1996, in his 2002. The idea that avowals express commitments is deployed in his final argument in his 2010. But neither paper makes a wholesale

reassessment of our ascriptive practice, and neither explicitly recognizes that embracing a regulative view of folk-psychology leads to a reduced estimate of how much ascriptive activity is flatly descriptive.

[40] Scenarios in which interpreters can be confident that ascription won't mislead audiences include a number of cases in which an ordinary concern about referential opacity is lifted. For example, we sometimes attribute beliefs to others using referential or descriptive phrases that are understandable only to our audience. Thus, for example, one might say to a friend, 'The judges think your wife is the best diver' even when none of the judges would have been able to think of the relevant individual in that way. Those judges will, in fact, lack some of the dispositions associated with the belief ascribed: they may be disinclined to draw any inferences about the star diver's marital status; and if they run into her husband later, they won't be disposed to congratulate him. But none of those dispositions matter in this context. The ascription used is less clumsy than many alternatives ascribing beliefs whose dispositional profiles the judges would completely match. Because those alternative ascriptions would be flatly true, it isn't the lack of a relevant flatly true ascription that moves one to make a merely conversationally appropriate one. Rather, other conversational norms and broad pragmatic concerns – such as making most of one's remarks from a perspective one shares with one's conversational partner – make the merely-conversationally-appropriate ascription more fitting.

There are many interesting questions to be pursued about how exactly conversational norms function in such cases. However, the fact that the field of pragmatics could yield insights here is not an objection to the claim that the relevant ascriptions lack determinate truth-values (when the subjects in question don't fully fit the relevant profiles). A related point is connected to the claim that we sometimes make negative ascriptions without intending to imply an opposed positive ascription. In some of the cases in which we say, 'She doesn't realize that p,' it may be that we can find some related content (even if not p) which we are pragmatically permitted to ascribe. Only the DV, however, would insist that it must always be possible to find such a content to use in a flatly true ascription.

I'm grateful to Kristin Andrews for pressing me to think about this, and for suggesting the useful example of attributions to judges.

[41] Both McGeer, 2007b, and Moran, 2001, discuss this aspect of self-ascriptions.

[42] Evans was explaining why there is no argument from the acceptability of ascriptions like these, to the claim that subjects can really think of and refer to objects without meeting the conditions for such thinking that he has laid out. The ascriptions are acceptable, in these cases, even when the subjects aren't (at the time of the ascription) thinking and referring in the relevant ways.

[43] Rowbottom, 2007, suggests this option in criticizing Schwitzgebel, 2001. (I am grateful to an anonymous referee for pressing me to think about it.) Rowbottom notes that Price, 1969, is a precedent for in-between views. Interestingly, when developing the example of what he terms 'the most important case of all, religious half-belief,' Price discusses this half-believer in terms that actually appear to match a shifting view of his case. That is, Price's religious half-believer appears to shift between full theistic belief on Sundays and determinate non-belief in theism the rest of the week (pp. 305–6). This helps support a weaker version of Rowbottom's claim: *some* cases of what someone with Price's or Schwitzgebel's commitments might initially classify as a case of in-between belief are perhaps better treated as cases of contextually-driven shifts among determinate beliefs. The pluralist version of the IBV defended in this article can certainly accept that weaker claim.

[44] Critically discussing the shifting view, Schwitzgebel appears to be assuming that such expectations apply almost across the board. One of his main reasons for rejecting the shifting view is that it isn't equipped to answer questions about what he terms a subject's 'general

attitude' – the attitude she has toward some content, full stop, rather than the attitude she has toward that content when doing this or that activity (Schwitzgebel, 2010, p. 543). His discussion is not developed enough to permit profitable speculation as to whether or not he would agree with the line I take here.

[45] Examples are likely to be controversial, precisely because normative expectations of consistency are so strong and so common. But here is a possibility. Arguably, folk-psychology expects that a belief such as 'Spring is my favorite season' should be general, and not contextually dependent. And yet many people sincerely utter '*This* is my favorite season' at the beginning of each season of the year, apparently in violation of that expectation. In many cases, folk ascriptive practice appears to cede them the belief, in the early spring, that early spring is the best, in the crisp fall, that the crisp fall is the best, and so on. They have four attitudes instead of the one they ought to have, but they really have all four (successively, in the appropriate contexts).

[46] There are very complicated questions here about how rational one's believing has to be in order for it to be believing at all (or, more generally, how far up to scratch on relevant normative standards one's intentional attitudes need to be in order for them to count as those particular intentional attitudes). The argument doesn't depend on endorsing any particular way of answering such questions. The argument does assume, however, that there is *some* point at which failures of rationality raise concerns about attitude identity and possession – some point at which one properly shifts from observations like 'What odd things she does with her belief that *p*' to 'It isn't clear she really does believe that *p*.' With respect to shifting attitudes, the argument doesn't depend on any particular way of sorting out when context-sensitivity properly raises questions about attitude-possession. It does, however, assume that there are some kinds of context-sensitivity that do raise that question. That is, it assumes there are cases where we properly shift from noting 'How funny that whether she believes that *p* depends on contextual feature *f*' to wondering 'If *f* can make a difference to her intentional attitudes in any way, it isn't clear she believes *p* in any context at all.' For further discussion, see Tumulty, 2012.

[47] Again, I am working with the dispositional model of belief for ease of presentation. Any other version of the in-between view could be substituted.

[48] In Schwitzgebel's judgment, a main part of the appeal of the version of the IBV that works in tandem with broad-track dispositionalism stems from its making links to action internal to belief. Some care is required in developing this view, however. Consider that subjects can share beliefs – that is, the same ascription could be flatly true of each of them – without necessarily behaving in identical ways. That's because they may have different environments, desires, and levels of access to additional information relevant to the topic of their belief. It would not be plausible to suppose that folk-psychological stereotypes for attitudes include hypothetical statements covering all the dispositions a subject might have in any of the infinitely many situations in which a subject with an attitude could be placed. So dispositional profiles for attitudes will likely prescribe (and proscribe) dispositions for only some of the many possible attitude/context combinations. There could, of course, be other folk-psychological norms prescribing combinations of attitudes – they just wouldn't be internal to the attitude itself. So the link between determinacy and harmony is not so strong as to make unworkable the pluralist version of IBV advanced in this article. We can accept that view while allowing that it is possible for there to be some kinds of conflicts between determinately held attitudes.

[49] I am grateful to very helpful comments from two anonymous referees. An ancestor of this paper was presented at a Colgate University Humanities Colloquium in April 2010, and I benefitted from the discussion that followed. An earlier version of this paper was accepted for the Central Division meeting of the American Philosophical Association in March 2011.

REFERENCES

Andrews, K. (2009). 'Understanding Norms Without a Theory of Mind,' *Inquiry* 52, pp. 433–48.

Audi, R. (1994). 'Dispositional Beliefs and Dispositions to Believe,' *Noûs* 28, pp. 419–34.

Bilgrami, A. (2008). *Self-Knowledge and Resentment*. Cambridge, MA: Harvard University Press.

Currie, G. (2000). 'Imagination, Delusion, and Hallucinations,' *Mind and Language* 15, pp. 168–83.

Egan, A. (2008). 'Imagination, Delusion, and Self-Deception,' in T. Bayne and J. Fernandez (eds) *Delusions and Self-Deception: Affective and Motivational Influences on Belief Formation*. London: Psychology Press, pp. 263–280.

Evans, G. (1982). *The Varieties of Reference*, John McDowell ed. Oxford: Clarendon Press.

Gendler, T. S. (2008a). 'Alief and Belief,' *Journal of Philosophy* 105, pp. 634–63.

Gendler, T. S. (2008b). 'Alief in Action (and Reaction),' *Mind & Language* 23, pp. 552–85.

Gendler, T. S. (2008c). 'Self-Deception as Pretense,' *Philosophical Perspectives: Mind* 21, pp. 231–58.

Huddleston, A. (2012). 'Naughty Beliefs,' *Philosophical Studies* 160, pp. 209–22.

Human Rights Campaign Foundation (2000). *Finally Free: How Love and Acceptance Saved Us From the Ex-Gay Ministries*. Available online at www.hrc.org/resources/category/religion-faith (accessed September 2012).

Hunter, D. (2011). 'Alienated Belief,' *dialectica* 65, pp. 221–40.

Kinkaid, R. (2011). 'Philosopher Unsure if the True Self is Gay,' *Jezebel*. Available online at http://jezebel.com/5809577/philosopher-unsure-if-the-true-self-is-gay (accessed September 2013).

Knobe, J. (2003). 'Intentional Action and Side Effects in Ordinary Language,' *Analysis* 63, pp. 190–93.

Knobe, J. (2011). 'In Search of the True Self,' *New York Times* Opinionator. Available online at http://opinionator.blogs.nytimes.com/2011/06/05/in-search-of-the-true-self/ (accessed September 2012).

McGeer, V. (1996). 'Is Self-Knowledge an Empirical Problem? Renegotiating the Space of Philosophical Explanation,' *Journal of Philosophy* 93, pp. 483–515.

McGeer, V. (2007a). 'The Moral Development of First-Person Authority,' *European Journal of Philosophy* 16, pp. 81–108.

McGeer, V. (2007b). 'The Regulative Dimension of Folk-Psychology,' in D. Hutto and M. Ratcliffe (eds) *Folk-Psychology Reassessed*. Dordrecht: Springer, 2007, pp. 137–56.

Mandelbaum, E. (2013). 'Against Alief,' *Philosophical Studies* 165, pp. 197–211.

Moran, R. (2001). *Authority and Estrangement: An Essay on Self-Knowledge*. Princeton, NJ: Princeton University Press.

Moran, R. (2003). 'Responses to O'Brien and Shoemaker,' *European Journal of Philosophy* 11, pp. 402–19.

Price, H. H. (1969). *Belief*. New York: Humanities Press.

Rowbottom, D. P. (2007). 'In Between Believing and Degrees of Belief,' *Teorema* 26, pp. 131–37.

Schwitzgebel, E. (2001). 'In-Between Believing,' *Philosophical Quarterly* 51, pp. 76–82.

Schwitzgebel, E. (2002). 'A Phenomenal, Dispositional Account of Belief,' *Noûs* 36, pp. 249–75.

Schwitzgebel, E. (2010). 'Acting Contrary to Our Professed Beliefs, or: The Gulf Between Occurrent Judgment and Dispositional Belief,' *Pacific Philosophical Quarterly* 91, pp. 531–553.

Schwitzgebel, E. (2011). 'Knowing Your Own Beliefs,' *Canadian Journal of Philosophy* 35, supp. vol., pp. 41–62.

Simons, D. J., Chabris, C. F., Schnur, T. and Levin, D. T. (2002). 'Evidence for Preserved Representations in Change Blindness,' *Consciousness and Cognition* 11, pp. 78–97.

Smith, A. M. (2005). 'Responsibility for Attitudes: Activity and Passivity in Mental Life,' *Ethics* 115, pp. 236–71.

Sommers, F. (2009). 'Dissonant Beliefs,' *Analysis* 69, pp. 267–74.

Tumulty, M. (2012). 'Delusions and Not-Quite-Beliefs,' *Neuroethics* 5, pp. 29–37.

Wilson, T. D., Lindsey, S. and Schooler, T. Y. (2000). 'A Model of Dual Attitudes,' *Psychological Review* 107, pp. 101–26.

Zimmerman, A. (2007). 'The Nature of Belief,' *Journal of Consciousness Studies* 14, pp. 61–82.

Zawidzki, T. W. (2008). 'The Function of Folk-Psychology: Mind-Reading or Mind-Shaping?,' *Philosophical Explorations* 11, pp. 193–210.