



## Philosophical Review

---

Doxastic Deliberation

Author(s): Nishi Shah and J. David Velleman

Source: *The Philosophical Review*, Vol. 114, No. 4 (Oct., 2005), pp. 497-534

Published by: Duke University Press on behalf of Philosophical Review

Stable URL: <http://www.jstor.org/stable/30043693>

Accessed: 15-10-2017 18:08 UTC

---

JSTOR is a not-for-profit service that helps scholars, researchers, and students discover, use, and build upon a wide range of content in a trusted digital archive. We use information technology and tools to increase productivity and facilitate new forms of scholarship. For more information about JSTOR, please contact [support@jstor.org](mailto:support@jstor.org).

Your use of the JSTOR archive indicates your acceptance of the Terms & Conditions of Use, available at <http://about.jstor.org/terms>



*Philosophical Review*, *Duke University Press* are collaborating with JSTOR to digitize, preserve and extend access to *The Philosophical Review*

## Doxastic Deliberation

Nishi Shah and J. David Velleman

Believing that *p*, assuming that *p*, and imagining that *p* involve regarding *p* as true—or, as we shall call it, accepting *p*. What distinguishes belief from the other modes of acceptance? We claim that conceiving of an attitude as a belief, rather than an assumption or an instance of imagining, entails conceiving of it as an acceptance that is regulated for truth, while also applying to it the standard of being correct if and only if it is true. We argue that the second half of this claim, according to which the concept of belief includes a standard of correctness, is required to explain the fact that the deliberative question *whether to believe that p* is transparent to the question *whether p*. This argument raises various questions. Is there such a thing as deliberating whether to believe? Is the transparency of the deliberative question *whether to believe that p* the same as the transparency of the factual question *whether I do believe that p*? We will begin by answering these questions and then turn to a series of possible objections to our argument.

### What Is Distinctive about Belief?

Propositional attitudes can be divided into the cognitive attitudes, which treat their propositional objects as *satisfied* or *true*, and the conative attitudes, which treat their propositional objects as *to be satisfied* or *to be made true*.<sup>1</sup> Philosophers sometimes try to express this characterization of the cognitive attitudes by saying, for example, that to believe something is to believe it true. But placing ‘true’ in the position of predicate adjective doesn’t yield a characterization that differentiates the cognitive from the conative, since to desire something is to desire it true, to wish something is to wish it true, and so on. In order to pick out the cognitive attitudes by using the word ‘true’, we must say that they treat or regard their propositional objects *as true*.<sup>2</sup>

This characterization fits not only belief but other cognitive attitudes as well. Assuming that *p*, supposing that *p*, and even imagining that *p* entail treating *p* as *satisfied* rather than *to be satisfied*, or regarding it as *true* rather than *to be made true*. The question is how to differentiate the concept of belief from the concepts of other attitudes that involve regarding-as-true. The answer cannot be that belief plays a distinctive motivational role, because the motivational role of belief is one that it

shares with other cognitive attitudes. Assuming that  $p$  and supposing that  $p$  resemble believing that  $p$  in that they dispose the subject to behave as if  $p$  were true; and even imagining that  $p$  may resemble belief in this respect.<sup>3</sup> What, then, is the difference between regarding  $p$  as true in the way that constitutes a belief and regarding it as true in the way that constitutes an assumption?

For ease of exposition, we will adopt the term ‘acceptance’ for the generic attitude of regarding-as-true, which is shared by the cognitive attitudes. Our question then becomes how accepting a proposition in the manner of belief differs from other modes of acceptance.

One respect in which belief differs from the other cognitive attitudes is the way in which it is formed, revised, and extinguished—or, as we shall say for short, the way in which it is regulated. One adopts an assumption for the sake of its utility in inquiry or argument, and one retains it just so long as it continues to serve that heuristic or polemical purpose. One’s adoption and retention of an assumption are not responsive to whether it is true. In forming and retaining a belief, however, one responds to evidence and reasoning in ways that are designed to be truth-conducive. Hence belief is regulated for truth, whereas other, non-belief-involving cognitive attitudes are not.<sup>4</sup>

In our view, being regulated for truth is part of the very concept of belief: to conceive of an attitude as a belief is to conceive of it as a cognition regulated for truth, at least in some sense and to some extent. But we think that the concept of belief must include more than the manner in which the attitude is actually regulated. Also part of the concept is a standard of correctness. Classifying an attitude as a belief entails applying to it the standard of being correct if and only if it is true.<sup>5</sup>

In the past, one of the authors has claimed that the standard of correctness for belief is derivable from the constitutive aim of belief.<sup>6</sup> Now, it is clear that if belief literally aimed at truth, in the sense that one would not qualify as believing that  $p$  unless one accepted  $p$  with the aim of doing so only if  $p$ , then the norm of correctness for belief could be explained in descriptive terms. Just as one’s other goals establish criteria of success for the activities that they regulate, so the goal of accepting  $p$  only if  $p$  would establish a criterion of success for one’s attitude of acceptance.<sup>7</sup> And whether one has the goal of accepting  $p$  only if  $p$  is a straightforwardly descriptive matter.

But belief cannot be required to have a literal aim, since only some instances of belief are caused by the goal-directed activity of their sub-

jects; many others are the product of processes such as perception, which don't involve any agential goals or intentions. If the metaphor that belief aims at the truth is not to rule out most cases of belief, it will have to draw on a wider notion of truth-directedness, encompassing non-agential mechanisms that track the truth. Being regulated by any truth-tracking mechanism will then be sufficient for a cognition to count as truth-directed in the sense required for belief.<sup>8</sup>

Once the notion of aiming at the truth is broadened in this way, however, the claim that belief's standard of correctness can be derived from its constitutive aim becomes harder to sustain. That claim would now require the assumption that if an attitude is regulated by a truth-tracking mechanism, then it is correct if and only if true. And this last assumption, though previously defended by one of the authors, has subsequently been abandoned in the face of objections from the other, to the effect that how an attitude *ought* to turn out is not necessarily determined by how it *is* regulated.<sup>9</sup> Our shared view is now that conceiving of an attitude as a belief must entail not only conceiving of it as regulated for truth but also, and independently, applying to it the standard of being correct if and only if true. The concept of belief, in short, is that of a cognition that is governed, both normatively and descriptively, by the standard of truth.<sup>10</sup>

Why do we think that the standard of correctness for belief must be included in the very concept, given that it is not derivable from what is included, in any case, about the manner in which belief is regulated? For an answer to this question, one of the authors has drawn on a familiar feature of belief that also serves to differentiate it from the other cognitive attitudes.<sup>11</sup> We shall call this feature *transparency*, although it is not exactly the same feature as one to which that term has been applied by others, as we shall presently explain. The feature that we call transparency is this: The deliberative question *whether to believe that p* inevitably gives way to the factual question *whether p*, because the answer to the latter question will determine the answer to the former.<sup>12</sup> That is, the only way to answer the question *whether to believe that p* is to answer the question *whether p*.<sup>13</sup> By contrast, the answer to the question *whether p* will not settle either the question *whether to suppose that p* nor the question *whether to imagine that p*, and so those questions do not give way to it—or, as we shall say, are not transparent to it. The best explanation for the transparency of doxastic deliberation to factual inquiry—that is, of the question *whether to believe that p* to the question *whether p*—is that the very concept of belief includes a standard of cor-

rectness, to the effect that a belief is correct if and only if it is true. This argument to the best explanation, presented elsewhere by one of us, raises various questions and objections, which it will be our joint aim to address in this paper.

### Background and Statement of the Argument

Why isn't the transparency of *whether to believe that  $p$*  to *whether  $p$*  explained by the descriptive fact that belief is regulated for truth? After all, we have admitted that belief's being regulated for truth is not merely a contingent fact but a conceptual truth. Hence, anyone who deliberates about whether to believe that  $p$  must recognize that his deliberation can potentially produce the envisioned attitude—that is, a belief that  $p$ —only by regulating that attitude for truth. Why, then, isn't transparency merely a reflection of the subject's recognition of this constraint on his deliberation?

Here we summarize an answer presented elsewhere by one of the authors; we refer the reader to that presentation for a fuller treatment of the question.<sup>14</sup> An adequate account of belief must explain not only the fact that truth occupies the sole focus of attention in doxastic deliberation but also the fact that evidentially insensitive processes, such as wishful thinking, occasionally influence belief. An account of belief rendered solely in terms of truth-regulation may be able to explain one or the other of these facts, depending on how the term 'truth-regulation' is interpreted, but no single interpretation of 'truth-regulation' can explain both facts at once. Hence, an adequate account of belief must include more than the fact of its being regulated for truth.

Being regulated for truth consists partly in responsiveness to evidence. The belief that  $p$  tends to be formed in response to evidence of  $p$ 's truth, to be reinforced by additional evidence of it, and to be extinguished by evidence against it. A crucial question is how strong these dispositions must be in order for the attitudes possessing them to qualify as beliefs. If belief can be influenced by evidentially irrelevant processes such as wishful thinking, then its responsiveness to evidence must be weak enough to leave room for such additional influences. And if this interpretation of truth-regulation is correct (as we believe), then the manner in which belief is regulated for truth can't be cited to explain the role of truth in doxastic deliberation. For when one deliberates *whether to believe that  $p$* , this question not only gives way to the question *whether  $p$*  but does so to the exclusion of any other, competing

question, such as whether  $p$  would be in one's interest. Yet if belief were required only to be weakly regulated for truth, then the potential outcome of deliberation could be envisioned as a belief that  $p$  so long as  $p$ 's truth were treated as relevant to that outcome, without necessarily being treated as having absolute priority over opposing considerations: the question *whether*  $p$  would not have to crowd out competing, non-epistemic questions.<sup>15</sup>

In order to explain transparency, the degree of evidence-responsiveness required by the concept of belief would have to be such as to rule out other influences, so that the outcome of deliberation could be envisioned as a belief that  $p$  only if the truth of  $p$  were treated as exclusively relevant to that outcome. In that case, however, the concept of belief would no longer allow for the possibility of a belief's being influenced by wishful thinking, since any attitude influenced by extra-evidential factors would fail to qualify as a belief. Interpreting the concept of belief as requiring evidence-responsiveness strong enough to account for transparency would therefore entail denying that it leaves room for other influences, whereas acknowledging that belief's responsiveness to evidence leaves room for other influences entails accepting that it is not strong enough to account for transparency. We choose the latter option, because transparency can be explained instead by the hypothesis that the concept of belief includes a standard of correctness.

This explanation goes roughly as follows. When one deliberates whether to have an attitude conceived as a belief that  $p$ , one deliberates about an attitude to which one already applies the standard of being correct if and only if  $p$  is true, and so one is already committed to consider it with an eye exclusively to whether  $p$ . When one deliberates whether to have an attitude conceived as an assumption or fantasy, one does not yet apply any particular standard to it, and so one does not yet have any commitment as to how one will go about considering it.

This explanation of transparency leaves room for the possibility that beliefs can be influenced by non-evidential considerations, because it entails that one is forced to apply the standard of correctness only in situations in which one exercises the concept of belief.<sup>16</sup> Not all belief-forming processes require the subject to deploy the concept, and the norm of truth that controls doxastic deliberation needn't control other processes. Our explanation of transparency thus allows for the fact that passions can influence belief.

Our goal in this paper is to fill in the foregoing sketch of doxastic deliberation. We will begin by addressing a question that arises immediately, which is how doxastic deliberation is even possible, given the apparent impossibility of deciding what to believe. After showing how it is possible to deliberate what to believe, we will contrast the case of deliberative transparency that is our focus from a related but different case of transparency. This contrast will allow us to shed further light on the normative dimension of the concept of belief. We will then be in a position to show in more detail how our proposal explains deliberative transparency. We will finish by raising and answering some objections to our explanation.

### How Doxastic Deliberation Is Possible

Is it even possible to deliberate about what to believe? Deciding what to believe is notoriously impossible. How can one deliberate about something that one cannot decide?

We think that doxastic deliberation is not only possible but commonplace. When someone makes an assertion that is not in itself convincing, the question that naturally comes to mind is whether to believe what he has said. When the president asserts that Iraq is harboring weapons of mass destruction, the natural question to ask is not “*Is Iraq harboring weapons of mass destruction?*” but rather “*Should I believe that?*”—whereupon this question transparently gives way to an inquiry into the truth of the president’s claim. Given how often one wonders whether to believe things that one has heard or read, it should be obvious *that* doxastic deliberation is possible; what is not obvious, of course, is *how* such deliberation is possible, or how it is compatible with the impossibility of deciding to believe.

Our explanation, in outline, is this. Deliberation is reasoning that is aimed at issuing in some result in accordance with norms for results of that kind. Deliberating about whether to  $\phi$  is reasoning aimed at issuing or not issuing in a  $\phi$ -ing, in accordance with norms for  $\phi$ -ing.<sup>17</sup> One can deliberate whether to believe that  $p$  because one can engage in reasoning that is aimed at issuing or not issuing in one’s believing that  $p$  in accordance with the norm for believing that  $p$ . And one can engage in reasoning with that aim precisely by considering the question *whether  $p$* . Considering *whether  $p$*  can amount to reasoning aimed at issuing or not issuing in one’s believing that  $p$  in accordance with the relevant norm because the relevant norm is this: believing that  $p$  is correct if and only

if  $p$  is true, and hence if and only if  $p$ . Furthermore, because the norm is contained in the concept of belief, and doxastic deliberation is framed by the question *whether to believe that  $p$* , anyone engaging in doxastic deliberation knows that this is the relevant norm.

The preceding paragraph is merely the outline of our argument for the possibility of deliberating whether to believe. Let us fill in the details of the argument by examining the nature of doxastic deliberation more closely.

Ordinarily, the reasoning that is meant to issue or not issue in a belief is meant to do so by first issuing or not issuing in a judgment. A judgment is a cognitive mental act of affirming a proposition (although, as we shall explain, not all affirmations are judgments). It is an act because it involves occurrently presenting a proposition, or putting it forward in the mind; and it is cognitive because it involves presenting the proposition *as true*—or, as we have said, affirming it.<sup>18</sup> A belief, by contrast, is a mental state of representing a proposition as true, a cognitive attitude rather than a cognitive act. In our view, the same standard of correctness is implicit in both concepts: a judgment, like a belief, is correct if and only if its content is true. Reasoning aims to issue or not issue in a belief that  $p$  in accordance with the relevant norm by first issuing or not issuing in a judgment that  $p$  in accordance with the corresponding norm. Strictly speaking, then, the question *whether to believe that  $p$*  is transparent, in the first instance, to the question *whether to judge that  $p$* , which in turn is transparent to the question *whether it would be correct to judge that  $p$* , and thence to *whether  $p$  is true* and, finally, to *whether  $p$* .

Now, there can be no problem about the possibility of deliberating whether to perform the mental act of affirming that  $p$ . As an *act*, mental affirmation is clearly eligible to be an object of deliberation. A problem might be thought to arise in the transition from that act of affirmation to an affirmative attitude, but to our minds, there is no problem about that transition, either. Exactly how one accomplishes the transition is of course ineffable, but it is a perfectly familiar accomplishment, in which a proposition is occurrently presented as true in such a way as to stick in the mind, lastingly so represented.<sup>19</sup> Affirming that  $p$  typically induces an affirmative attitude toward  $p$ . Hence, deliberating whether to form an affirmative attitude toward  $p$  is no more problematic than deliberating whether to affirm it. Why, then, should there be a problem about deliberating whether to judge and consequently believe that  $p$ ?



There appears to be a problem only because one cannot judge or believe arbitrarily—or, as they say, at will<sup>20</sup>—and whatever one cannot do at will is often thought not to be within one's control. Since one cannot deliberate about that which one cannot control, the impossibility of judging or believing at will is thought to rule out deliberating about judgment or belief.

The flaw in this reasoning becomes clear upon a further look at the reason why one cannot judge or believe arbitrarily. One can mentally affirm that *p* arbitrarily; and one can arbitrarily affirm that *p* in such a way as to produce a standing representation of *p* as true. For example, one can arbitrarily tell oneself a story to the effect that *p*, and one can thus arbitrarily induce in oneself a standing fantasy that *p*; or one can arbitrarily present the hypothesis that *p* in such a way as to produce a standing assumption. Apparently, then, what prevents one from judging or believing arbitrarily is not that one cannot arbitrarily engage in the affirmative mental act or thereby produce the affirmative mental attitude, but rather that one cannot do so arbitrarily while also doing so in the way that amounts to judgment or belief.

Consider first the case of judgment: why can't one arbitrarily affirm that *p* in such a way as to make a judgment? The reason is that an affirmation that *p* qualifies as a judgment, rather than a mental fiction or hypothesis, only when it is aimed at getting the truth value of *p* right—aimed, that is, at presenting *p* as true only if it really is true.<sup>21</sup> Affirming that *p* without regard to whether *p* is true is story-telling or conjecturing, not judging. Yet to say that an affirmation that *p* amounts to a judgment only if it is aimed at getting the truth value of *p* right is just to say that this affirmation amounts to a judgment only if it is aimed at giving the correct answer to the question *whether p*. That's why reasoning as to whether *p* is the only way of deliberating whether to judge that *p*: it's the only reasoning that can aim at issuing in a judgment that *p*, because it's the only reasoning that can aim at issuing in an affirmation aimed at giving the correct answer to the question *whether p*.

Indeed, when reasoning aims at correctly answering the question *whether p*, it *ipso facto* constitutes deliberation whether to judge that *p*, because it aims at issuing in an affirmation that *p* if *p* is true but not otherwise, and an affirmation that *p* arrived at with that aim will amount to a judgment that *p* made in accordance with the relevant norm. Hence, when reasoning aims at correctly answering the question *whether p*, it thereby aims at issuing or not issuing in a judgment that *p* in accordance with the relevant norm,<sup>22</sup> and so it constitutes deliberation

whether to judge that  $p$ . Note that it constitutes such deliberation whether or not it begins with an articulation of the deliberative question *whether to judge that  $p$* . Whether a process of reasoning is an instance of deliberation does not depend on whether it is initiated by a deliberative question; it depends instead on the aim with which the reasoning proceeds.

The obstacle to arbitrarily judging that  $p$  is that one can judge that  $p$  only by making an affirmation aimed at giving the right answer to the question *whether  $p$* , and an affirmation cannot be made arbitrarily if it is to have that aim.<sup>23</sup> Having a non-arbitrary aim doesn't prevent the judgment that  $p$  from being an object of deliberation: it merely entails that the only way to deliberate about this judgment is to reason toward an affirmation aimed at correctly answering the question *whether  $p$* . In short, there is only one way of deliberating whether to judge that  $p$ , but there being only one way of deliberating about judgment cannot entail that there is no way at all.<sup>24</sup>

The same goes for the possibility of deliberating whether to believe. When reasoning aims at correctly answering the question *whether  $p$* , it aims at issuing or not issuing in a judgment that  $p$  in accordance with the norm for such a judgment—namely, the norm that it will be correct if and only if  $p$  is true. And if the resulting judgment induces a standing attitude of acceptance, that acceptance will have been regulated by the same norm and hence in the manner characteristic of belief. The way to reason with the aim of forming or not forming a belief that  $p$  in accordance with the norm for belief is thus to reason with the aim of correctly answering the question *whether  $p$* , by forming or not forming a judgment that  $p$  in accordance with the norm for judgment. Even if this procedure is the only way of answering the deliberative question *whether to believe that  $p$* , there being only one way can hardly entail that there is none.<sup>25</sup>

Like deliberation whether to judge that  $p$ , deliberation whether to believe that  $p$  need not be initiated by an articulation of the deliberative question: one can start right in with the factual question *whether  $p$*  and yet be recognizably deliberating whether to believe. What makes one's reasoning recognizably deliberative becomes clear upon comparison with non-deliberative reasoning about the same question. One can consider *whether  $p$*  in the spirit of idly wondering, without aiming to make up one's mind—in which case, one isn't deliberating about whether to believe that  $p$ . Idly wondering whether  $p$  is different from trying to make up one's mind whether  $p$ , and only the latter constitutes

deliberating whether to believe. The latter constitutes such deliberation despite skipping the deliberative question *whether to believe*, simply because it is reasoning aimed at issuing or not issuing in a belief in accordance with the norm for believing.<sup>26</sup>

### A Related Case of Transparency

Before proceeding further, we should contrast our case of transparency with a closely related but crucially different case. A question that is sometimes characterized as transparent to the factual question *whether p* is the question *whether I believe that p*,<sup>27</sup> which is also factual and hence significantly different from our deliberative question *whether to believe that p*.

#### Asking *Whether I Believe* Tout Court

The question “Do I believe that *p*” can mean either “Do I already believe that *p* (that is, antecedently to considering this question)?” or “Do I now believe that *p* (that is, now that I am answering the question)?”<sup>28</sup> Consider by way of analogy two possible meanings of the question “Do you intend to vote in the next election?” This question can be interpreted as asking either for one’s antecedent state of mind, up to when one was asked, or for one’s state of mind as of when one answers. One can respond to the latter question truthfully by deciding to vote and then saying “Yes”; but one cannot truthfully say “Yes” to the former question by first forming the relevant intention. Similarly, one can answer the question *whether I now believe that p* by forming a conscious belief with respect to *p*, whereupon one’s consciousness of that belief will provide the answer; but one cannot answer the question *whether I already believe that p* in a way that begins with forming the belief.

Now, either of these questions can give way to the question *whether p*. If the question is *whether I already believe that p*, one can assay the relevant state of mind by posing the question *whether p* and seeing what one is spontaneously inclined to answer. In this procedure, the question *whether p* serves as a stimulus applied to oneself for the empirical purpose of eliciting a response. One comes to know what one already thinks by seeing what one says—that is, what one says in response to the question *whether p*.<sup>29</sup> But the procedure requires one to refrain from any reasoning as to whether *p*, since that reasoning might alter the state of mind that one is trying to assay. Hence, asking oneself *whether p* must be a brute stimulus in this case rather than an invitation to reasoning.

By contrast, the question *whether I now believe that  $p$*  is potentially transparent to the question *whether  $p$*  in the capacity of just such an invitation. One reliable way to answer the question *whether I now believe that  $p$*  is to form a conscious judgment with respect to  $p$ , thereby inducing a corresponding acceptance, and the way to form a conscious judgment with respect to  $p$  is consciously to aim at giving a correct answer to the question *whether  $p$* . Hence, the question *whether I now believe that  $p$*  naturally leads to the question *whether to judge that  $p$* , which gives way to the question *whether  $p$* , now posed as an invitation to reasoning.

Note, however, that these two cases of transparency are imperfect. Arriving at the judgment that  $p$  doesn't necessarily settle the question whether one now believes it, since one may find oneself as yet unconvinced by one's own judgment. One may reason one's way to the conclusion that one's plane is not going to crash, for example, and yet find oneself still believing that it will.<sup>30</sup> Similarly, testing one's spontaneous response to the question *whether  $p$*  may yield good evidence as to whether one already believes that  $p$ , but that evidence isn't conclusive: one's first thought upon entertaining a question may be misleading as to one's pre-existing attitude. In both of these cases, a prior question gives way to the alternative question *whether  $p$*  as the most promising route to an answer, but that promise may or may not be fulfilled. One may spontaneously answer the question *whether  $p$*  without thereby learning whether one already believes it, and one may deliberate to an answer without learning whether one now believes it.

Do cases in which asking oneself *whether  $p$*  fails to inform one of what one really believes constitute failures of rationality? Not in the case in which one is attempting to assay an antecedent belief. As we pointed out, one cannot engage in reasoning aimed at answering the question *whether  $p$*  if one wants to find out what one already believes, because such reasoning would contaminate the result by possibly altering the state that one is trying to assay. Since one doesn't reason in delivering up one's spontaneous response as to *whether  $p$* , the fact that one does or does not come to know what one already believes with respect to  $p$  doesn't tell for or against the rationality of the belief. However, in the case in which one is attempting to determine what one now believes, as of when one answers the question, the failure to come to know what one believes by coming to a judgment *whether  $p$*  would constitute a failure of rationality. After all, one is attempting to know *whether one believes that  $p$*  by settling on a belief *whether  $p$* , and one is attempting to settle one's belief *whether  $p$*  by reasoning to a judgment *whether  $p$* ; so if this rea-

soning fails to determine *whether one believes that  $p$* , then one's reasoning has failed to inform one of what one now believes about  $p$  precisely because it has failed to determine one's belief. Take for instance the example from the previous paragraph: one attempts to settle the question *whether one now believes that the plane will crash* by reasoning to a judgment *whether the plane will crash*, but one's negative answer to the latter question doesn't yield a corresponding belief. In this case, an irrational phobia has had a dominant hand in determining what one believes.

A symptom of one's irrationality in this case is that one is in a position to have a thought with the form of Moore's paradox: "The plane will be safe, but I don't believe it." The first half of this thought would embody one's judgment that the plane won't crash, while the second half reported one's failure to form the corresponding belief. What would be paradoxical in an assertion of this form is not that there is no train of thought that it could express; what would be paradoxical is rather that the train of thought expressed would be irrational.

### Asking Whether an Acceptance Is a Belief

In the cases discussed thus far, answering the question *whether  $p$*  fails to settle the question *whether I believe that  $p$*  because it fails, more specifically, to settle the question *whether I regard  $p$  as true*—or, as we put it, *whether I accept that  $p$* . Acceptance is an attitude that is distinct from the act of judgment and not equally easy to detect. And belief requires acceptance and more, with the result that any doubts as to what one accepts must be doubts as to what one believes. But there are instances of *whether I believe that  $p$*  in which one's acceptance of  $p$  is not in question. Watching the president on television, one may find oneself thinking "That's a lie!" and one may be quite sure of regarding this statement as true, without being sure of so regarding it in the manner of a belief. Maybe the president's mendacity is more like a default assumption on one's part—an hypothesis always worth testing—or maybe it is just a paranoid fantasy. In such a case, one's doubts are confined to whether one's acceptance of  $p$  qualifies as a belief.

As before, this question can ask either whether one's acceptance already qualifies as a belief, at the moment when one is asking the question, or whether it now qualifies, as of when one answers. The latter version of the question is also transparent to the question *whether  $p$* . The case largely resembles the one described above, in which one asked *whether I now believe that  $p$*  in a sense that included the question *whether*

*I now accept that  $p$ .* In the present case, one's acceptance of  $p$  has been established, and so the question whether one now believes that  $p$  is easier to settle. For one can proceed to deliberate whether to believe the proposition simply by applying the standard of truth to one's existing acceptance of it, thereby disposing oneself to retain or discard that acceptance in accordance with the outcome of this very reasoning as to the truth of its content. Having applied that standard and adopted that disposition, one will conceive of the acceptance as a working hypothesis, perhaps, but not yet as a belief, since one will not yet regard it as actually regulated for truth. But by considering the question *whether  $p$* , and allowing the answer to determine the fate of one's existing acceptance that  $p$ , one can eventually answer the question *whether I now believe that  $p$* , since the ultimate outcome will be either that the acceptance is extinguished by a conscious judgment to the contrary or that it is reinforced by a conscious judgment that  $p$ , whereupon it will have survived regulation for truth and thus satisfy the remaining descriptive condition for belief. Thus, when one finds oneself thinking that  $p$ , one can ascertain whether one now believes it by considering whether  $p$  in a manner that amounts to deliberating whether to believe that  $p$ .

So much for asking, with respect to a pre-existing acceptance, whether it is now and henceforth a belief. What about the question whether it is already a belief? This question is not transparent to *whether  $p$* , because the answer will be largely determined by the facts about one's pre-existing acceptance—specifically, whether it has been regulated for truth, as required by the concept of belief. The answer does not depend on and cannot be reached by considering the truth or falsity of  $p$ . Yet even if one finds that one's acceptance of  $p$  has been regulated for truth, conceiving of it as a belief will also entail applying the standard of truth to it normatively, as a standard of correctness.

Applying the standard of truth to an attitude as a standard of correctness can be imagined in the first instance as a noncognitive state of accepting a norm in application to that attitude. But it can also be imagined as a way of conceiving of the attitude—conceiving of it, that is, as being correct if and only if its content is true. Conceiving of an attitude as a belief can then be interpreted, in sum, as conceiving of it as a truth-regulated acceptance that is correct if and only if true. And when the concept of belief is interpreted in this way, the question arises whether its components are on a par—specifically, whether conceiving of an attitude as correct if and only if true is just a further descriptive characterization of it, on a par with conceiving of it as a truth-regulated

acceptance. We think not. We opt for an interpretation of the concept of correctness along norm-expressivist lines.<sup>31</sup>

We will defend our choice of expressivism only to the extent of developing a coherent expressivist account of belief-attribution and showing that it explains the relevant phenomena. The phenomena we adduce will include cases in which the question whether a truth-regulated acceptance should be classified as a belief appears to be underdetermined by the facts, just as it would be if classifying such an acceptance as a belief were a matter of bringing it under a noncognitive norm. We shall claim such cases as evidence for our expressivism about that aspect of the concept which characterizes truth-regulated acceptances as correct if and only if true. Yet a full defense of our expressivism would of course require arguments against various forms of realism about this normative aspect of the concept, and such metaethical arguments lie well beyond the scope of this paper.

According to our expressivist interpretation, applying the belief-constituting standard of correctness to one's own acceptance that *p* consists in accepting the norm of truth for that acceptance, where accepting a norm is a conative attitude that, among other things, disposes one to follow the norm and inhibits one from following any alternative. Applying the standard of correctness to someone else's acceptance that *p* disposes one to follow the norm in approving or disapproving of his acceptance, in offering criticism and suggestions, and so on. Whether to apply the truth-norm to an acceptance, in either case, is not a factual question, according to the expressivist view.

How can there be no fact of the matter as to whether an acceptance that *p* must be true in order to be correct? Our answer, to begin with, is that there being no fact of the matter does not entail that it is an open question. In most cases, one will have longstanding normative commitments to hold cognitions of various kinds to the standard of truth, and one may be in no position to reconsider those commitments. Indeed, one will be hard pressed to avoid applying this norm to a cognition that one regards as already having been regulated for truth, for reasons that we will explore in a moment. One will also be hard-pressed to avoid applying it to a cognition that one allows to play an important role in guiding one's behavior.

However, we think that there clearly are situations in which the facts leave open the question whether an acceptance should be brought under the norm of truth. Suppose one recognizes that one has long regarded a fellow philosopher as an adversary, and that this view of him

has survived at least some opportunities for disconfirmation, thus potentially qualifying as having been regulated for truth. Is this view of one's colleague a belief or just a useful fantasy that serves to spur one's intellectual activity? It doesn't seem to us that there is a fact of the matter that could settle this question; rather, one must *decide* whether or not to treat it as a fantasy or a belief. That is, one must decide whether or not to apply the norm of truth to this attitude and thus dispose oneself to regulate it accordingly. Our point is not that one's attitude is determinately not a belief because one has thus far failed to apply the norm of truth to it. The attitude satisfies the only descriptive requirements for belief, since it is an acceptance that has been to some extent regulated for truth. Our point is that whether it should be classified as a belief is not a further factual question, but a question of whether to apply the norm to it.

This view may seem to imply that there are no such mental states as beliefs. Do we mean to say that psychologists who study processes that they call belief-formation are making a radical error, by looking for something that doesn't exist? Yes and no. We do think that the concept of belief, taken in its entirety, involves a normative component, which we interpret noncognitively. We are therefore committed to saying that believing a proposition is not a property. But remember that we also interpret the concept of belief as having a fairly substantial descriptive content: it characterizes an attitude as a truth-regulated acceptance, to which it then applies the norm of truth. What we're committed to saying, then, is that believing a proposition is a property *and more*. In many contexts, such as explanations of behavior, the normative aspect of belief recedes into the background, and 'belief' is used to pick out states of truth-regulated acceptance. We suspect that psychologists' attributions of 'belief' are made in just such contexts, and hence that their object of study is truth-regulated acceptance, of which there is no reason to be skeptical.

### Suspending Belief and Deciding to Believe

A positive answer to the question whether one already believes that *p* does not necessarily dictate a positive answer to the question whether one now believes it. The reason is that one can often choose to suspend belief. An examination of this possibility will shed further light on the concept.



Conceiving of a cognition as governed both normatively and descriptively by the standard of truth yields a further normative implication. This conception implies that insofar as the cognition is true, it is not just correct but also a success, in that the mechanisms regulating it have succeeded in bringing it into conformity with the standard; whereas insofar as the cognition is false, it is not just incorrect but a failure, because the mechanisms regulating it have failed.

This normative implication becomes especially significant in first-personal attributions of belief. If one regards oneself as still in the process of regulating a cognition for truth, one may not be prepared to submit it for assessment as a success or a failure, and one may therefore be reluctant to classify it as a belief. Suppose that one adds a long column of numbers and arrives at the total 1,234. One may then regard it as true that the sum is 1,234 and yet distrust one's addition and want to recheck it. One will say, "I *think* it's 1,234, I *suspect* that's the sum, but I won't believe it until I've had a chance to check my addition." How does one avoid classifying this acceptance as a belief?

The answer, we suggest, is that one applies a different norm to the acceptance, by classifying it as an hypothesis instead. The norm implicit in the concept of belief is that of being correct if and only if true. Holding a false belief is contrary to this norm: it's not permissible, because incorrect. But holding a false hypothesis is not contrary to the norm relevant to hypotheses. Indeed, holding some hypotheses that are false is unavoidable given the purpose of hypotheses, which is to test whether they are true or false, by reasoning and acting on them experimentally. One would have little chance of learning what is true if one didn't hypothetically accept some propositions that turn out to be false. Holding a false hypothesis becomes incorrect only when its falsity becomes evident, whereupon continuing to hold it would constitute a refusal to follow through on the epistemic purpose for which it was adopted.<sup>32</sup>

In classifying an acceptance as an hypothesis, then, one gives oneself permission to hold it pending the outcome of further tests. If those tests subsequently reveal it to be false, then one would be incorrect to go on holding it, according to the norm for hypotheses, but one will not have been incorrect to hold it initially.

The concept of an hypothesis also has a secondary normative implication. The permission implicit in this concept is the permission to accept a (possibly false) proposition for the purpose of testing whether it is true. If one doesn't have that epistemic purpose—that is, if one

isn't engaged or disposed to engage in testing the proposition—then granting oneself permission to accept it is contrary to the normative presuppositions of the concept. If one doesn't put one's acceptance at risk, by seeking opportunities to find the proposition false, then one isn't entitled to the permission that one would grant oneself in accepting it under the concept of an hypothesis.

Because one can often decide to go on testing an accepted proposition for truth, however, one can justifiably decide to conceive of it as an hypothesis rather than a belief. One's refusal to conceive of it as a belief will be justified insofar as one puts one's acceptance of the proposition at risk, thus adopting the epistemic purpose for which accepting possible falsehoods is permitted by the norm for hypotheses. In this sense, one can often decide *not* to believe—which is, strictly speaking, a matter of deciding not to accord a cognition the status of a belief. Alternatively, one can sometimes decide to stop checking an hypothesis and rest content for now with the checks that it has survived thus far. In suspending tests of the hypothesis, one will forfeit one's entitlement to the permission of which one availed oneself in conceiving of it as an hypothesis, and one will then be obliged to apply the more stringent norm, thereby conceiving of the cognition as a belief. In this sense, at least, one *can* decide to believe—which is, strictly speaking, a matter of deciding to accord a cognition that status. Of course, one cannot just decide to believe that a column of numbers adds up to 3 because one likes the number 3. But having added the column and arrived at a particular result, one can decide whether or not to believe that result, in the sense that one can decide whether or not to accord the status of belief to one's acceptance of it.<sup>33</sup>

Similar latitude for deciding to confer or withhold the status of belief is sometimes available with respect to an acceptance that has already attained the status of belief. Having ascertained that one already believes that *p*, one may yet be in a position to re-open the question *whether p* in a way that genuinely puts one's acceptance of *p* at risk, and one may then be entitled to regard that acceptance as a working hypothesis or suspicion rather than a belief. (Of course, if one can re-open the question *whether p* in such a way as to suspend *acceptance* of *p*, then the suspension of belief follows trivially; what is not trivial, and requires explanation, is the possibility of suspending belief while continuing to accept what one has hitherto believed.) One cannot in this manner alter the fact that one did regard one's acceptance of *p* as a belief until now, but one can decide whether to continue so regarding

it—which is what accounts for one’s sense of being able to decide whether to go on believing it or to reduce its status to that of an hypothesis. One cannot genuinely re-open the question *whether p* with respect to just any proposition *p*. In the case of obvious truths, one can at most pretend to recheck them, or go through the motions of rechecking them, without genuinely putting them at any risk of disconfirmation. Hence, one can at most pretend that they are in the process of being checked for truth, thereby merely pretending to earn the permission implicit in conceiving of them as hypotheses. In other cases, however, one can decide not to believe what one has hitherto believed in the sense that one can regard it instead as an hypothesis under ongoing investigation.

This latitude in whether to regard one’s acceptance as a belief suggests that the applicable standard of correctness is not a matter of fact. Hence, these cases support our expressivist interpretation of the norm implicit in the concept of belief. The question whether to apply the concept appears to hang on pragmatic considerations of the sort that would determine whether to apply a noncognitive norm.

Note that we have not drawn the distinction between belief and more tentative modes of acceptance on the basis of subjective probabilities. The concept of belief, as we understand it, allows for degrees of credence; and insofar as those degrees depend on the subject’s confidence in how an acceptance has been regulated thus far, they are likely to be correlated with his disposition toward further active testing, and hence with his classification of the acceptance as a belief or an hypothesis. But as we shall see, the classification of an acceptance can depend on considerations other than the subject’s degree of confidence.

### Is ‘Belief’ Judgment-Dependent?<sup>34</sup>

Does an agent’s classification of his acceptance as a belief, or instead as a working hypothesis, determine whether it really is a belief or an hypothesis?

First, notice that we are willing to attribute beliefs to creatures that don’t think about their own mental states at all. Hence, self-ascription of belief isn’t necessary for having a belief. Might it nonetheless be a sufficient? We think not, since we think there can be reasons for differing with a subject as to the classification of his cognitions.

If one thinks that a person is no longer actively testing a cognition—that it is no longer at risk in his thinking—then one may conclude that

he is not entitled to the permission implicit in conceiving of it as an hypothesis rather than a belief. For example, suppose someone claims that he doesn't expect bad news from his medical tests but is merely assuming the worst so as to be mentally prepared, just in case. If he shows no interest in getting the results, one may begin to suspect him of really believing, and not just preemptively assuming, that the news will be bad. Conversely, suppose that someone claims to believe that his cough is nothing serious and even acts as if he regards this proposition as true—for example, by taking no steps to seek treatment, and by going out in the cold lightly clothed. If one finds him continually checking his temperature or reading up on the symptoms of lung cancer, one may concede that he is acting on the assumption that it's nothing serious and yet question his claim to believe it. In this case, the subject's evident apprehensions of the cognition's being disconfirmed lead to the conclusion that he ought to hold it only under the more permissive concept of an hypothesis.<sup>35</sup>

The possibility of doubting a subject's classification of his own cognitions can take on a normative aspect in conversation with the subject himself. One sometimes responds to an assertion with, "You don't really *believe* that!"—a remark that is intended less as a statement of fact than as an unsubtle suggestion. One might equally have said, "You may suspect that, or imagine it, but you haven't tested it sufficiently to say that you believe it—and so you won't think of it as a belief any more, now, will you?"

So when we speak of a subject's latitude in classifying his own cognitions, we do not mean to imply that his classification is always decisive for other observers. There is of course a social norm of deferring to the subject as to whether he believes that *p*, is imagining that *p*, or is treating *p* as a working hypothesis. The main reason for deferring to the subject in this matter is that when one classifies his cognition as a belief, one thereby applies the standard of truth to it, and one ought to hold him, in the first instance, to the standards that he applies to himself. This norm, telling us to hold agents to the norms to which they hold themselves, is in a sense, a norm of respect. Normative regulation is the exercise of a person's rational agency, and holding an agent accountable to a norm that he takes himself to be under is an acknowledgement of respect for his rational agency. We think these reasons for deference go some way toward explaining what philosophers have called the first-person authority that agents have with respect to their

beliefs, although, as we have argued, we do not think that this authority is absolute.

Yet implicit in the concept of belief, as we understand it, is a policy of deference that goes beyond expressing respect for the subject's rational agency. According to our analysis of the concept, one may apply the norm of truth to a cognition by classifying it as a belief only if it is in fact regulated for truth to some extent, as if it were being held to the same norm by the subject himself. Hence, the descriptive aspect of the concept forbids one to apply it, with its implicit norm, to any cognition toward which the subject doesn't behave as if he were applying that norm. Of course, if the cognition is being regulated by the subject's cognitive systems operating on their own, then he will not in fact be the agent of their regulation, and there may be no more than a notional or "as if" exercise of his rational agency to which one can be interpreted as deferring when one considers whether to classify his cognition as a belief. Interpreted more literally, one is deferring to the way in which the cognition is treated by the subject's mind.

But surely the subject's mind is also that to which one is required to extend charity by Donald Davison's principle of that name. Davidson's principle of charity requires one to interpret a subject as largely believing what's true and desiring what's good. But the credit for the subject's believing what's true and desiring what's good is often due, not to the subject himself as the witting agent of those accomplishments, but to his cognitive and conative systems. And the deference that is built into the concept of belief, according to our analysis, is just a variant of Davidsonian charity. The concept of belief, as we understand it, doesn't require one to assume that the subject is generally correct; it merely forbids one from holding him to a standard of correctness that he has no disposition to meet.

Note that this minimal degree of charity has a normative element that is missing from Davidson's principle. All that Davidson's principle of charity says is that in order to interpret a subject as a believer, one must interpret him as having beliefs that are largely true. Of course, if one already accepts that the concept of belief involves truth as its standard of correctness, then acceptance of Davidson's principle of charity will compel one to interpret a subject as having beliefs that are not only largely true but also largely correct. But this normative conclusion will then have been arrived at from Davidson's principle of charity plus an independent premise that true beliefs are correct. Davidson's princi-

ple of charity by itself doesn't express anything normative about the concept of belief.<sup>36</sup>

### Pragmatic Reasons for Believing

The foregoing considerations about the classification of cognitions imply that there can be something like pragmatic reasons for or against believing. There cannot literally be pragmatic reasons for believing *tout court*, since reasons for believing *tout court* must recommend an attitude conceived as being correct only if true, and an attitude so conceived cannot be recommended by pragmatic considerations.<sup>37</sup> (We turn to a defense of this claim below.) But if one already does accept something and that acceptance is responsive to truth-tracking mechanisms, then there can be pragmatic reasons for or against according that acceptance the status of a belief.

The pragmatic reasons relevant to the classification of a cognition have to do primarily with what is at stake. Sometimes the costs of continuing active testing of a cognition would be high if it were true, whereas the costs of postponing further tests would be low if it were false. One then has pragmatic reason to accord the cognition the status of belief. In other cases, the costs of failing to find counterevidence, if it existed, would be high, or those of continuing to look for it, if it didn't exist, would be low. One then has pragmatic reason for hypothesizing rather than believing. Thus, one can have pragmatic reasons for believing rather than suspecting that there is a predator in the shadows, and one can have pragmatic reasons for suspecting rather than believing that one has managed to prove Goldbach's conjecture.<sup>38</sup> (As before, what we loosely describe as a choice between believing and suspecting is, strictly speaking, a choice between ways of classifying one's own cognition and, ultimately, a choice between norms implicit in the relevant classifications.)

In these cases, the factual question *whether I believe that p* is not at all transparent to the question *whether p*, because it leads one to classify an existing acceptance of *p* in a context where the truth or falsity of the proposition accepted is not the primary grounds for classification. Of course, these are cases in which the deliberative question *whether to believe that p* is not transparent to *whether p*, either. Indeed, the former, factual question is not transparent to *whether p* in these cases precisely because it *is* transparent to the latter, deliberative question, which is not in turn transparent to *whether p*. The question whether one believes

that *p* is here a question of whether to accord one's acceptance of *p* the status of belief, and hence a question of whether *to* believe it rather than merely suspect or hypothesize it. And when the question is whether to believe that *p* rather than merely suspect or hypothesize it—given that one somehow accepts it—the answer may depend on the relative costs of belief versus suspicion rather than on the truth value of *p*.<sup>39</sup>

This concludes our discussion of the question *whether I believe that p* and its transparency to the question *whether p*. We have found that when consideration of the latter supplants consideration of the former, it does so either by way of the deliberative question *whether to believe that p* or at least in the form of deliberation that tacitly pursues that question, insofar as it aims at issuing or not issuing in a belief that *p* in accordance with the relevant norms. Thus, what is fundamentally transparent to the factual question *whether p* is the deliberative question *whether to believe that p*; and this primary transparency will be the exclusive focus of our attention from now on.

### A Constraint on Doxastic Deliberation

We have just argued that the deliberative question *whether to believe that p* can be answered on pragmatic grounds when asked in reference to an existing cognition classifiable either as a belief or an hypothesis. We contrasted this application of the deliberative question with that which asks whether to believe *tout court* that *p*—that is, whether to regard *p* as true in a manner governed both normatively and descriptively by the standard of being correct only if *p* is actually true. We argued earlier that whether to believe *tout court* that *p* can be the topic of deliberation that proceeds via the factual question *whether p*. Our next task is to explain why answering that factual question is the *only* way of deliberating whether to believe *tout court*. This will complete our explanation of why arbitrarily believing *tout court* that *p* is impossible.

We think that the necessity of considering whether *p* as the only means of deliberating whether to believe *tout court* that *p* becomes obvious once this version of the deliberative question is distinguished from that which merely seeks classification of a pre-existing cognition. All apparent counterexamples fall into one or another of the categories of classificatory deliberation discussed in the previous section. That is, they are examples in which the question is not just whether to believe that *p* but whether to believe it rather than hold some other cognitive attitude toward it, given that one already accepts it.

Here is the reason why considering whether  $p$  is the only means of deliberating whether to believe *tout court* that  $p$ . To deliberate whether to believe that  $p$  is to engage in reasoning aimed at issuing or not issuing in a belief that  $p$  in accordance with norms for such a belief. And to conceive of the attitude at issue in one's reasoning as a belief that  $p$  is to conceive of it as an acceptance that will be correct if and only if  $p$  is true. Thus, deliberating whether to believe must consist in reasoning aimed at issuing or not issuing in an acceptance that would be correct if and only if true, in accordance with norms for such an attitude. Since the norm of truth is already applied to the envisioned attitude as part of the concept of belief, it must be the norm in accordance with which deliberation aims to issue in the attitude. Aiming to arrive at an attitude in accordance with a competing norm would be incompatible with subsuming it under the norm of truth in conceiving of it as a belief.<sup>40</sup> Deliberation whether to believe that  $p$  must therefore consist in reasoning aimed at issuing in an acceptance of  $p$  if and only if this acceptance would be correct in virtue of  $p$ 's being true. And the only way of reasoning with that aim is to consider *whether*  $p$ , so as to arrive, if  $p$  is true, at a judgment that would induce an acceptance of  $p$ .<sup>41</sup>

### Replies to Objections

Before we consider objections to this account of transparency, we should note what it implies about the norms governing belief. The standard of correctness implicit in the concept of belief, according to our view, is a biconditional norm: a belief is correct if and only if it is true. But correctness is itself a permissive rather than injunctive notion. A norm of correctness forbids the holding of beliefs that would be incorrect, but it merely permits the holding of correct beliefs. One is not required to hold every belief that would be correct. In deliberating whether to believe that  $p$ , however, one is committed to forming the belief if it would be correct, and this commitment tends to supply the injunctive half of a biconditional norm, mandating a belief in  $p$  if and only if  $p$  is true.

#### First Objection: A Practical Syllogism for Belief?

This explanation immediately raises the following objection. Reasoning cannot aim at issuing in an acceptance of  $p$  if and only if that acceptance would be correct in virtue of  $p$ 's being true, because pursuit of that aim would entail first ascertaining whether  $p$  is true; and ascertain-



ing whether  $p$  is true would entail arriving at a belief with respect to  $p$ , as an intermediate step in deliberating whether to believe it. And believing that  $p$  cannot be an intermediate step in deliberating whether to believe that  $p$ .

This objection claims, in effect, that doxastic deliberation, as we conceive it, would have to conclude with a practical syllogism whose premises were:

I will believe that  $p$  if and only if  $p$  is true.  
 $p$  is true.

Strictly speaking, the minor premise of this syllogism would be embodied in a judgment, not a belief. But that judgment would be equivalent to the one with which we imagine doxastic deliberation as concluding. The objection is that doxastic deliberation cannot require its own conclusion as a premise.

As we have explained, however, the way to deliberate whether to believe that  $p$ —that is, the way to reason with the aim of arriving or not arriving at a belief that  $p$  in accordance with the norms for such a belief—is to consider, not the question *whether  $p$  is true*, but the disquoted version of that question, *whether  $p$* . If consideration of *whether  $p$*  leads to the judgment that  $p$ , it will induce an acceptance of  $p$  in a way that, to the best of one's ability, accords with the standard of correctness implicit in the concept of belief; and so it will, to the best of one's ability, accomplish the aim characteristic of doxastic deliberation. Since considering the question *whether  $p$*  is the way to accomplish that aim through reasoning, it is the way to reason if one has that aim, and it is therefore the way to deliberate whether to believe that  $p$ .

Of course, in order to adopt consideration of *whether  $p$*  as one's means of deliberating whether to believe that  $p$ , one must regard it as conducive to the characteristic deliberative aim, and so one must regard it as likely to issue in a judgment that  $p$  if and only if  $p$  is true. One must therefore have methods for considering whether  $p$  that one regards as truth-conducive. The present objection proves this much, at least: that doxastic deliberation cannot aim at truth directly. In order to aim at accepting the truth with respect to  $p$ , one cannot aim in the first instance at accepting  $p$  if and only if it is true; one must aim at following some truth-conducive methods that will lead to its acceptance.<sup>42</sup> As we have argued, those methods are methods for correctly answering the question *whether  $p$* .

A corollary of this result is that doxastic deliberation requires that one already have some beliefs—namely, beliefs about the truth-conduciveness of one's methods for answering *whether p*. Obviously, then, one's first beliefs cannot be formed deliberatively, in answer to the question whether to believe. Indeed, one cannot have the concept of belief, as a state distinct from the other cognitive attitudes, unless one can conceive of an acceptance as having been regulated for truth, a conception that requires in turn at least the belief that there are truth-conducive methods of regulation. One must therefore have a fairly rich repertoire of beliefs before one can have the concept of belief.

Note that this corollary does not impose any preconditions on *having* beliefs but only on conceiving of them as such, either prospectively, in posing the question whether to believe, or retrospectively, in classifying existing cognitions as beliefs. Toddlers and lower animals can have attitudes classifiable as beliefs without having the concept of truth-conducive methodology; they just can't conceive of what they have as beliefs. They are believers unwittingly.<sup>43</sup>

### Second Objection: Brains in Vats

Do envatted brains have beliefs by our lights? Their acceptances aren't regulated by genuinely truth-conducive methods. Indeed, they may not even have attitudes of genuine acceptance, if a behavioral disposition is an essential component of regarding-as-true. They may not even have the concept of belief, given their lack of rapport with any genuine subjects of propositional attitudes.

The fact remains, however, that envatted brains have a vat-world proxy for the concept of belief, which they can apply to vat-world proxies for attitudes of acceptance that have been vat-ish-ly regulated for the vat-world proxy of truth. Or as Putnam might put it, they have acceptances-in-the-image that have been regulated-in-the-image for truth-in-the-image, and to these they can apply the concept-in-the-image of belief. By their own lights, then, they have the analog of what would count as belief by our lights. And surely the mental life of envatted brains should be no more perspicuous than this.

### Third Objection: Regulation for Something Other than Truth

We have argued that the concept of belief distinguishes it from other cognitive attitudes by applying to it the standard of being correct if and only if true, and by characterizing it as having been regulated, at least

to some extent, for conformity to that standard. Even if you agree that belief is distinguished from the other cognitive attitudes by being subject to a standard of correctness, you might think that truth is not the relevant standard. Perhaps belief is distinguished by being subject to a standard of epistemic rationality, as defined by the very truth-conducive methods to which we alluded above.

You might think that this view offers several advantages over ours. First, it doesn't make the concept of belief dependent on the philosophically problematic concept of truth. Instead of holding belief to the standard of truth via methods regarded as truth-conducive, it cuts straight to the chase, by holding belief to a standard embodied in the methods themselves, thus dispensing with the concept of truth entirely. Second, the view does not make attributions of belief dependent on the presupposition that there are genuinely truth-conducive methods by which cognitions can be regulated. It therefore skirts the embarrassing fact that the truth-conduciveness of most available methods for regulating cognitions is open to question, given the unsolved problems of epistemological skepticism. By contrast, our view invokes the problematic concept of truth and interprets belief-attributors as presupposing the existence of truth-conducive methods.

Unfortunately, the proposed alternative to our view would not yield a unified concept of belief, because epistemic methods vary in accordance with the content of the cognitions that they are meant to regulate. The methods of mathematical proof are different from those of simple induction, which are different again from those of statistical inference, and so on. Hence one's acceptance of the Pythagorean theorem has been regulated rather differently from one's acceptance of the sun's daily reappearance, and differently yet again from one's acceptance of a diagnosis based on medical tests. If the concept of belief applied the standard of being regulated by the relevant methods, then different standards would be applied to different cognitions, and the concept itself would be fragmented. Merely conceiving of the methods as "relevant" would raise the question 'relevant to what?' and the proposed view would have denied itself recourse to the obvious answer—namely, "relevant to the cognition's truth."

For similar reasons, this problem cannot be circumvented by appeal to warranted assertability as the standard of correctness for belief. To begin with, warranted assertability cannot apply directly to beliefs; it can apply only to their verbal expression. And as we shall point out below, the verbal expression of a belief must in this context be con-

ceived as the *correct* expression, whose correctness would initially seem to consist in an identity of truth conditions. More importantly, unless the concept of warranted assertability is given some substantive content in terms of the truth-conduciveness of the speaker's justification (precisely the sort of content that the current objection aims to avoid), it amounts to no more than a name for a miscellany of justificatory methods, which are not rendered any more unified by being corralled under a single name. Without any unifying conception of what makes for warranted assertability in the cases of mathematical, inductive, and statistical statements, for example, the concept of cognitions governed normatively and descriptively by standards of warranted assertability would remain too fragmented to constitute the concept of belief.

#### Fourth Objection: Deflationism

Perhaps an alternative answer is available from the theory known as deflationism about truth. Methods relevant to regulating one's acceptance of the Pythagorean theorem, according to this theory, are simply those which are relevant to ascertaining whether the square of a right triangle's hypotenuse is necessarily equal to the sum of the squares of the other two sides; methods relevant to regulating one's acceptance of a medical diagnosis are those which are relevant to the question whether one has strep throat, or whatever.

Come to think of it, though, deflationism about truth can be applied directly to the problem of distinguishing beliefs from the other cognitive attitudes. For maybe the standard to which beliefs are characteristically subject is not the standard of being correct if and only if true, but a standard that might be described as the disquotational standard—the standard of being correct if and only if *p*, where “belief that *p*” is the expression used to attribute the belief in question. Thus, even if you concede to us, as before, that belief is distinguished from other cognitive attitudes by being subject to a particular standard of correctness, you might insist that the availability of the disquotational standard obviates any need for a standard of truth.

What we have called the disquotational standard is in fact not a standard at all but rather a standard-schema. The schema specifies that the standard of correctness for any belief can be generated by substitution of the belief's characterization in indirect discourse for '*p*' in the sentence 'The belief that *p* is correct if and only if *p*'. In a sense, then, the disquotational standard fails to establish a single standard of correct-

ness for beliefs, and it may therefore appear inadequate to explain how a single, unified concept of belief can be differentiated from the concepts of other cognitive attitudes.

Yet there is a second-order norm that could serve this purpose—a norm that requires evaluation and regulation of any particular belief in accordance with that standard of correctness which would be generated for it by the schema “The belief that  $p$  is correct if and only if  $p$ .” The single, unified concept of belief could then be that of an acceptance governed both normatively and descriptively by the norm of being evaluated and regulated in accordance with a disquotationally generated standard of correctness.

Our claim that the norm of truth is embedded within the concept of belief may thus be compatible with a deflationist interpretation of that norm. In the case of belief, however, deflationism faces an obstacle that is absent from the case of language. Disquotation is a syntactic operation. That is, in the disquotational schema “The statement ‘ $p$ ’ is true if and only if  $p$ ,” the quoted and unquoted occurrences of the variable must be interpreted as standing for tokens of the same syntactic type. They cannot be interpreting as standing, say, for any two expressions that have the same meaning, since such an interpretation risks smuggling in the concept of truth conditions, which the disquotational schema is supposed to help eliminate. Yet in the case of belief, there is initially no syntactic item to extract from quotation marks, or from a that-clause, by the syntactic operation of disquotation. The disquotational schema for generating a standard of correctness must therefore involve the concept of a quoted expression or a that-clause with which a belief would be characterized or attributed and on which the operation of disquotation can be carried out. And the concept of belief cannot be that of an acceptance governed by a standard of correctness generated disquotationally from just any characterization or attribution; it must be, more specifically, that of an acceptance governed by a standard generated disquotationally from a *correct* characterization or attribution. And what is a correct characterization or attribution of an acceptance? Surely, a correct characterization or attribution is one involving a quoted expression or that-clause with the same content as the acceptance being characterized or attributed. Even if there is a syntactic operation by which to generate a standard of correctness for an attitude from the linguistic expression used to characterize or attribute it, *there is no syntactic operation by which to generate the appropriate linguistic expression from the attitude itself*. The concept of an acceptance governed

by a disquotationally generated standard of correctness would therefore have to presuppose a semantic relation between the acceptance and the required target of disquotation.

Whether the deflationist can account for this semantic relation without invoking the concept of truth is a question that we must leave for another occasion. One might suspect, however, that the very concept of belief requires a robust conception of truth, if not in its implicit standard of correctness for belief, then in an implicit standard of correctness for belief-attribution.

#### Fifth Objection: Circularity<sup>44</sup>

According to our account, there are two conceptual truths about belief, one descriptive and one normative. The descriptive truth is that a belief is an acceptance that is regulated for truth, at least to some extent. The normative truth is that a belief is correct if and only if it is true. This latter conceptual truth, employing the normative term 'correct', requires a metaethical interpretation. Instead of attempting a straight definition of correctness, as if it were a property, we favor the noncognitivist strategy of explicating what it is to *apply* a standard of correctness. And we then incorporate this indirect explication of correctness into our account of belief by extending the strategy of indirectness to the latter as well, transposing it into an explication of what it is to attribute a belief. Our account of the concept of belief thus belongs to the expressivist tradition in metaethics.

The final objection that we will consider focuses on this expressivist part of the account. According to our expressivist account of belief-attribution, X judges that Y believes that *p* if and only if X applies the norm of truth to Y's acceptance that *p*. Yet it would seem that for X to have any noncognitive attitude whatsoever toward a mental state of Y's, X must *believe* that Y has that mental state. Our account of belief-attribution thus presupposes that the belief-attributer has a belief about the target of his attribution. And this presupposition would seem to introduce a circularity into our account of the concept of belief.

In fact, two circularities may be charged against us. First, we may be charged with the conceptual circularity of using the concept of belief in the course of explicating that very concept. Second, we may be charged with the pragmatic circularity of making belief-attributions in the course of explicating belief-attribution. To the first charge we

plead innocent, to the second, guilty—but with mitigating circumstances that let us off the hook. We'll answer the charges in order.

To begin with, X can apply the truth-norm to an acceptance of Y's without believing in that acceptance. For there are contexts in which belief-attribution doesn't require that the attributer believe his own attribution. For example, in the course of acting out a role in a game, one can attribute the belief that *p* to another without actually believing that the other either accepts or believes that *p*. One might tease a superstitious friend by pretending that he believes in ghosts, for example, without actually believing that he does. One would still be making a belief-attribution, but it would be an attribution of a fanciful kind.

What's more, our description of this belief-attribution as fanciful just consists in our declining to apply the truth-norm to it. When we say (sincerely), "X attributes to Y the belief that *p*," we are, first, expressing our acceptance of the proposition that X accepts the proposition that Y accepts the proposition *p*; second, we are expressing our acceptance of the proposition that X applies the truth-norm to Y's acceptance of *p*. We may or may not be applying the truth-norm to X's acceptance of the proposition that Y accepts that *p*; this contingency will determine whether we are ascribing to X a serious or a fanciful belief-ascription. Finally, we may or may not be applying the truth-norm, or inviting our hearers to apply it, to any of the acceptances that we ourselves are expressing; this contingency will determine whether we take ourselves, or invite our hearers to take us, as expressing beliefs about X.

We have now formulated our account of belief without using the concept of belief, and so any conceptual circularity has vanished. The concept of belief has been replaced in our account by the concept of acceptance, which we have explicated in turn as the concept of regarding-as-true—an explication that does not use the concept of acceptance itself.<sup>45</sup> What initially raised suspicions of conceptual circularity was our norm-expressivist strategy of explicating belief indirectly, by explicating belief-attribution, since belief-attribution involves a cognitive attitude that we were suspected of having to characterize, circularly, as a belief. Having characterized the attitude involved in belief-attribution as an acceptance, we needn't fear renewed suspicions of circularity, because our explication of acceptance is not expressivist and hence doesn't proceed via an explication of acceptance-attribution.

Had our account of acceptance been expressivist, then it would indeed have devolved into a vicious circle. We would then have been obliged to explicate acceptance indirectly, by explicating acceptance-

attribution, which clearly involves some cognitive attitude on the part of the attributer; and any cognitive attitude involves acceptance. Hence, our account of acceptance would have relied on that very concept. A circularity of this general form is what was initially suspected in our account of belief but has now been avoided, thanks to our descriptivist account of acceptance.

Nevertheless, our account of belief still contains a pragmatic circularity—though not exactly the one that we adumbrated at first. Initially we suggested that the pragmatic circularity consisted in our having to explicate belief-attribution by making a belief-attribution; but we have now seen that our explication of belief-attribution can make do with an attribution of acceptance instead. Even so, our explication of belief-attribution also uses the concept of acceptance-attribution, since it says that X attributes a belief to Y by attributing an acceptance to Y and applying the truth-norm to it. And if we tried to explicate acceptance-attribution, a pragmatic circularity would now ensue. In explicating what it is for X to attribute an acceptance of *p* to Y, we would have to attribute an acceptance to X—that is, X's acceptance of the proposition that Y accepts that *p*. Hence, we could not explicate the phenomenon of attributing acceptances without ourselves exemplifying that phenomenon.

On the one hand, this pragmatic circularity does place a limit on the understanding that our explication can convey—or, more precisely, the lack of understanding that it can remedy. Someone who does not understand acts of attributing acceptances will not learn to understand such acts by hearing our explication of them, because our explication is carried out in acts of the very kind that he cannot understand. On the other hand, this degree of pragmatic circularity is unavoidable in the philosophy of mind and language. Someone who does not understand assertions cannot learn to understand them by hearing a philosophical explication of assertion, since that explication will be carried out in assertions.

For this very reason, however, a philosophical account of assertion cannot be intended to teach anyone how to understand assertions—that is, how to interpret them—since it can reasonably be addressed only to those who already understand them in this sense. It is addressed to an audience that understands assertions but doesn't know what assertions are. Similarly, our account of belief-attribution is not intended to teach anyone how to understand belief-attributions: it is



addressed to an audience that understands belief-attributions but doesn't know what they are.

*Amherst College* (Shah)

*University of Michigan, Ann Arbor and New York University* (Velleman)

### References

- Bar-on, Dorit. 2004. *Speaking My Mind: Expression and Self-Knowledge*. Oxford: Oxford University Press, Clarendon Press.
- Boghossian, Paul A. 2003. The Normativity of Content. *Philosophical Issues* 13:31–45.
- Bratman, Michael. 1992. Practical Reasoning and Acceptance in a Context. *Mind* 101:1–15.
- Edgley, Roy. 1969. *Reason in Theory and Practice*. London: Hutchinson.
- Korsgaard, Christine. 1997. The Normativity of Instrumental Reason. In Garrett Cullity and Berys Gaut (eds.), *Ethics and Practical Reason*, ed. Garrett Cullity and Berys Gaut, 215–54. Oxford: Oxford University Press.
- . n.d. Self-Constitution: Action, Identity, and Integrity.
- Moran, Richard. 2001. *Authority and Estrangement*. Princeton: Princeton University Press.
- Owens, David John. 2000. *Reason Without Freedom*. London: Routledge.
- . 2003. Does Belief Have an Aim? *Philosophical Studies* 115:283–305.
- Scanlon, Thomas. 1998. *What We Owe to Each Other*. Cambridge: Harvard University Press.
- Schroeder, Timothy. 2003. Donald Davidson's Theory of Mind is Non-Normative. *Philosophers' Imprint* 3.1:1–14.
- Shah, Nishi. 2001. Thinking Through Belief, Ph.D. diss. Department of Philosophy, University of Michigan, Ann Arbor.
- . 2002. Clearing Space for Doxastic Voluntarism. *The Monist* 85:436–45.
- . 2003. How Truth Governs Belief. *Philosophical Review* 112:447–82.
- . Forthcoming. A New Argument for Evidentialism. *Philosophical Quarterly*.
- Velleman, J. David. 1989. Epistemic Freedom. *Pacific Philosophical Quarterly* 70:73–97; reprinted in *The Possibility of Practical Reason* (Oxford: Oxford University Press, 2000).
- . 2000a. Introduction to *The Possibility of Practical Reason*.
- . 2000b. On the Aim of Belief. In *The Possibility of Practical Reason*, 244–81.
- Wallace, R. Jay. 2001. Normativity, Commitment, and Instrumental Reason. *Philosophers' Imprint* 1.3:1–26.

- Wedgwood, Ralph. 2002. The Aim of Belief. *Philosophical Perspectives* 16:267–97.
- . Forthcoming. The Normativity of the Intentional. In *The Oxford Handbook of the Philosophy of Mind*, ed. Brian P. McLaughlin and Ansgar Beckermann. Oxford: Oxford University Press, Clarendon Press.
- . n.d. Contextualism and Moral Beliefs.
- Williams, Bernard. 1973. Deciding to Believe. In *Problems of the Self*, 135–51. Cambridge: Cambridge University Press.

## Notes

For comments on earlier drafts of this paper, we are grateful to David Enoch, Pamela Hieronymi, Don Hubin, Daniel Jacobson, David Owens, Karl Schafer, David Sobel, and Ralph Wedgwood. Shah's research was supported by an Amherst College Trustee Faculty Fellowship and a grant from Amherst College's Faculty Research Awards Program. Versions of the paper were presented to the Philosophy Department of the University of California, Riverside; the Pre-Conference on Truth and Realism at the University of St. Andrews; the Ohio Reading Group in Ethics; and Roger White's epistemology seminar at New York University.

<sup>1</sup> We do not mean to suggest that this classification covers all the mental phenomena for which philosophers have used the term 'propositional attitude'. In order to cover all propositional attitudes, we would have to replace the phrase "to be made true," because conative attitudes include not just desires but also aversions, whose propositional content is regarded as to be made false, or to be prevented from coming true. And it's unclear whether even this description would fit attitudes such as regret, whose content is assumed to be beyond falsifying. We won't pursue these issues here, since we are primarily concerned with the cognitive attitudes. Finally, we do not claim that the classification of attitudes is uncontroversial even in the case of attitudes described as beliefs. After all, descriptivists and expressivists about normative concepts differ as to whether normative beliefs should be interpreted as cognitive or conative states.

<sup>2</sup> Of course, our phrase 'regarding *p* as true' must not mean "believing that *p* is true," lest it generate an infinite regress of beliefs. Some other interpretation of this phrase is therefore required. The interpretation we favor is that an attitude involves regarding *p* as true insofar as it plays the motivational role discussed in the next paragraph—that is, the role of disposing the subject to behave in ways that would satisfy his desires in worlds where *p* was true. (See Velleman 2000b, 255–77.) Our arguments in the present paper do not depend on this interpretation of 'regarding as true'; they merely presuppose that some unproblematic interpretation is available. Even if it turned out that belief had a distinctive motivational role not shared with any other cognitive attitude, this would not be something given to us just by our knowledge of the concept of belief. Our knowledge of the difference between the concept of belief and the concepts of the other cognitive attitudes thus cannot consist in knowledge of

belief's distinctive motivational role.

<sup>3</sup>For arguments in support of the minimal claim that there is at least one type of cognitive attitude with the same motivational role as belief, see Bratman 1992. For arguments in support of the stronger claim that the motivational role picked out by Bratman is shared by all the cognitive attitudes, including imagining that *p*, see Velleman 2000b, 255–77.

<sup>4</sup>The qualifier “non-belief-involving” is meant to allow for cognitive states such as knowledge, which are indeed regulated for truth but, in our view, only because they involve belief.

<sup>5</sup>This thesis is defended by Boghossian 2003.

<sup>6</sup>Velleman 2000a, 16–17.

<sup>7</sup>Owens (2003) argues that an account like ours cannot distinguish between believing and guessing. This argument raises many questions that we cannot discuss in the present paper. (These questions include whether guessing is a mental act or a mental state, and whether it entails belief.)

<sup>8</sup>Velleman 2000b, 254.

<sup>9</sup>Shah 2003. Below we summarize why belief's regulation for truth can't account for a phenomenon that is itself a reflection of the fact that truth is belief's standard of correctness.

<sup>10</sup>See Wedgwood 2002 for a separate defense of the conceptual claim that truth is belief's standard of correctness. Wedgwood interprets the metaphor that belief aims at the truth solely in terms of this normative claim, whereas, as we have indicated, we think that there is both a descriptive and a normative component to belief's truth-directedness. (Wedgwood's forthcoming closes some of the distance between us on this point.) Furthermore, while it seems that Wedgwood understands this normative claim as articulating a *factual* judgment about the essence of belief (270–71), as we shall explain later, we prefer a norm-expressivist interpretation of the concept of correctness according to which applying a standard of correctness to a cognition doesn't involve ascribing a property to it.

<sup>11</sup>See Shah 2003.

<sup>12</sup>For the purposes of this paper we set aside deliberative contexts in which *p* is nonfactual.

<sup>13</sup>But don't I sometimes engage with prudential reasons in deliberating whether to believe that *p*, as when I rehearse Pascal's Wager in deliberating whether to believe that God exists? No. In the sense we have in mind, deliberating whether to believe that *p* entails intending to arrive at a belief as to whether *p*. If my answering a question is going to count as deliberating whether to believe that *p*, then I must intend to arrive at a belief as to whether *p* just by answering that question. I can arrive at such a belief just by answering the question whether *p*; however, I can't arrive at such a belief just by answering the question whether it is in my interest to hold it. Once I've answered the question whether it is in my interest to believe that *p*, I must still take steps to induce the relevant belief. As even Pascal recognized, accepting that it would be in my interest to believe that God exists may lead to the decision to take steps to bring about that belief, but it won't bring me to believe that God exists. See Shah forthcoming for further discussion of Pascal's Wager.

<sup>14</sup> See the section titled “The Teleologist’s Dilemma” in Shah 2003.

<sup>15</sup> As we explain later, in the deliberative context in which one is asking the contrastive question whether to believe rather than hypothesize that *p*, pragmatic reasons are relevant. However, so long as one’s question is whether to believe *tout court* that *p*, one can be guided only by considerations that are congruent with one’s alethic reasons.

<sup>16</sup> Our claim here is not that deliberation about what to believe cannot be influenced by non-evidential considerations; it is that such deliberation cannot explicitly treat such considerations as relevant to the question what to believe. Any influence that such considerations exert must be unacknowledged.

<sup>17</sup> In cases in which there is a constitutive norm for  $\phi$ -ing, this is the relevant norm, and in cases in which there is no constitutive norm, the relevant norms are supplied by the context.

<sup>18</sup> Like the phrase ‘regarding as true’, the phrase ‘presenting as true’ will require explication, which is not on the agenda for the present paper. See note 2, above.

<sup>19</sup> Of course, in order for a representation to take hold in one’s mind it needn’t continue to be the object of one’s conscious awareness.

<sup>20</sup> See Williams 1973, 148, for a much-discussed argument that believing at will is impossible. Williams himself does not go on to assert the further claim that doxastic deliberation is therefore impossible. See Shah 2002 for discussion of Williams’s argument.

<sup>21</sup> Because judgment is an *act*, it differs from belief in that it necessarily has a literal aim. Hence, there is no problem accounting for judgment’s standard of correctness: its standard of correctness is just the criterion of success associated with the intention with which it is made. An agent’s judgment that *p* is correct if and only if it successfully fulfills his attempt to affirm *p* only if *p*—that is, if and only if it is true.

<sup>22</sup> Not necessarily under that guise, however, since the *de dicto* content of the aim needn’t include the concept of a norm.

<sup>23</sup> There are a few exceptions to this rule. Self-fulfilling affirmations can be made arbitrarily without sacrificing this aim; so can affirmations such as the Cartesian *cogito*, which must be true in order to be made. See Velleman 1989.

<sup>24</sup> In this respect judgment is no different from other kinds of aim-constituted activities. For example, you cannot attempt to build a house without aiming to provide a structure capable of providing shelter, because providing shelter is a constitutive aim of house-building. The reason why you cannot attempt to build a house without concern for its capacity to provide shelter is not that you can’t play around with bricks and mortar without such a concern; it’s that your activity won’t count as an attempt to build a house. But isn’t a shoddy builder exactly someone who builds a house without such a concern? Korsgaard (n.d., 24) answers this question nicely:

The shoddy builder doesn’t follow a different set of standards or norms. He may be doing one of two things. He may be following the norms, but carelessly, inattentively, choosing second-rate materials in a random way, sealing the corners imperfectly, adding insufficient insulation, and so on. But he may also, if he is dishonest, be doing this sort of thing quite consciously, say in order to save money. In that case, surely we can’t say he is trying to build a good house? No, but now I think we should say that he is not trying to build a house at all, but

rather a sort of plausible imitation of a house, one he can pass off as the real thing. What guides him is not the aim of producing a house, but the aim of producing something that will fetch the price of a house, sufficiently like a real house that he can't be sued afterward.

Hence, you cannot intend to build a house arbitrarily—that is, without regard to whether the structure you are building can provide shelter. But of course this obstacle to arbitrarily building a house does not entail that you cannot deliberate about whether or how to build a house.

<sup>25</sup> We have not yet established that this is the only way of deliberating whether to believe, and thus we have not yet explained why arbitrary believing is impossible. What we have shown, by describing one procedure for answering the deliberative question *whether to believe that  $p$* , is how doxastic deliberation is possible.

<sup>26</sup> Does this mean that one can deliberate whether to believe without having the concept of belief? Well, it's true that one needn't employ the concept of belief in the foreground of one's thoughts in order to engage in doxastic deliberation. But insofar as one is aiming to form an acceptance that will be true, one must conceive of the forthcoming acceptance as being hereby regulated for truth and, if true, then a success in that respect. If one applies no countervailing standard of correctness to the forthcoming acceptance, then conceiving of it as a success if true may be sufficient for conceiving of it, in effect if not in name, as a belief.

<sup>27</sup> See Moran 2001, 60–65, for a discussion of this type of transparency. Moran, 60–61, cites Edgley 1969, 90, as the source for this use of the term 'transparency'.

<sup>28</sup> Note that the word 'now' in this question is proleptic: it refers to the time of the anticipated answer to the question, not the time at which the question is being asked.

<sup>29</sup> This explanation is proposed by Bar-on (2004, 318): "If asked whether you believe  $p$ , you will directly consider whether  $p$  is to be believed. We can think of this as a way of putting yourself in a position to give direct voice to your (first-order) belief."

<sup>30</sup> See Scanlon 1998, 35, for another example of "akratic" believing, in which one's belief that Jones can be relied on, formed in response to Jones's appearance of warmth and friendship, fails to give way to one's judgment that he is an artful deceiver.

<sup>31</sup> Our norm-expressivist interpretation was initially proposed and defended in Shah 2001.

<sup>32</sup> Here we are using the term 'hypothesis' in a narrow sense that does not include, for example, the wittingly false assumptions that are made for the purpose of conducting what is often called hypothetical reasoning from counterfactual antecedents.

<sup>33</sup> Note that denying the status of belief to a cognition need not entail any difference in one's behavior. Even as one rechecks one's addition, one still suspects or hypothesizes or thinks that the sum is 1,234, and if called upon to act upon the sum of the numbers, one will act accordingly. One is disposed to behave in ways that would be appropriate if the sum were 1,234, because one accepts that result; one merely accepts it tentatively, pending further confir-

mation, and consequently without according that acceptance the status of belief.

<sup>34</sup> This section draws on Shah 2001.

<sup>35</sup> Haven't we contradicted our earlier claim that the concept of belief cannot be distinguished from the concepts of the other cognitive attitudes on the basis of belief's motivational role? After all, whether or not someone is actively testing a cognition that *p*—and thus whether he should be regarded as hypothesizing that *p* or believing that *p*, according to our current claim—is a question about the motivational role played by his cognition. Our current claim, though, is that the less restrictive norm associated with the concept of an hypothesis entails a permission that is not entailed in the concept of belief—specifically, the permission to have been wrong, provided that one corrects the mistake once it is discovered—and the subject has to earn the right to the more permissive norm, by proving that he is still testing the cognition. This last statement, about what it takes for someone to “earn this right,” is a substantive normative claim, not something that can be settled just by examining the descriptive conditions that are built in to the concept of an hypothesis.

<sup>36</sup> See Schroeder 2003 for discussion of whether there is any substantial sense in which Davidson's theory of mind is normative.

<sup>37</sup> There is a minor exception to this statement in the case of self-fulfilling beliefs. See Velleman 1989.

<sup>38</sup> See Wedgwood n.d.

<sup>39</sup> Owens 2000, chap. 2, argues that while non-evidential reasons play a non-deliberative role in belief-formation, they cannot move an agent by way of his reflective recognition of them as reasons for belief because non-evidential considerations are not relevant to answering the question whether *p*, which is the question that frames reflection about whether to believe that *p*. As we shall presently explain, we agree that non-evidential considerations cannot be reasons in deliberating whether to believe *tout court* that *p*, because this question is transparent to the question whether *p*. But the question whether to believe that *p* rather than hypothesize that *p* is the question whether to accord one's acceptance that *p* the status of belief rather than working hypothesis. This latter question, we have just argued, is not transparent to the question whether *p*, since it asks whether to apply a norm that permits one, for the purpose of testing whether *p* is true, to accept it even if it is, in fact, false. And pragmatic considerations of the sort we have just described can certainly be relevant to answering this question.

<sup>40</sup> This explanation relies on a very weak form of internalism about normative thought. The relevant form of internalism does not require a positive disposition to obey any norm that one applies; what it requires is the lack of a disposition to obey a different norm instead. One cannot genuinely apply the norm of truth to an attitude while simultaneously trying only to make it conform to some other, unrelated norm. This form of internalism does not rule out obedience to additional norms compatible with the one applied. One can aim to arrive as quickly as possible at a true cognition with respect to *p*—in which case, one will deliberate in accordance with a norm of speed as well as the norm of truth. One can aim to arrive at a true cognition in a manner that,

if it leads to error, is more likely to err in stopping short of true cognition than in arriving at a false one; one will then deliberate in accordance with norms of caution as well as truth. What one cannot do, according to our weak form of internalism, is to apply a norm in thought while hewing to a competing norm in practice. We leave it as an open question whether in cases of evidential equipoise—where the evidence equally supports  $p$  and not  $p$ —applying a non-epistemic norm to break the tie is compatible with adherence to the norm of truth.

<sup>41</sup> In this respect doxastic deliberation is no different from deliberation of other kinds. One cannot deliberate about something by taking up questions that are obviously irrelevant to it. For example, one cannot deliberate about how to build a house by considering whether snow is white, or whether the Cubs will ever make it to the World Series. Considering those questions just wouldn't amount to deliberating about how to build a house, because the questions are obviously irrelevant to house-building. Similarly, one cannot deliberate about what to believe by considering questions irrelevant to believing. And because only one question is relevant to the topic of what to believe, there being only one standard of correctness for belief, it is the only question consideration of which amounts to deliberating on that topic. But precisely because that question is relevant, considering it does amount to deliberating about what to believe.

<sup>42</sup> Compare with Boghossian 2003, 39, and Wedgwood 2002, 282.

<sup>43</sup> Some philosophers (Wallace (2001, 10), Korsgaard (1997, 248), Moran (2001, 52)) have tried to express the normative relation between belief and truth by saying that believing that  $p$  involves a *commitment* to the truth of  $p$ . We find this locution less than perspicuous. If committing oneself to the truth of  $p$  means *doing* something—performing a mental act—then it sounds to us like a judgment that  $p$ , not a belief that  $p$ . If committing oneself to the truth of  $p$  means accepting the norm of truth in application to one's attitude toward  $p$ , then it would rule out toddlers and lower animals from having beliefs, as they don't have the requisite normative concepts. The way to avoid this consequence is to locate the commitment to truth, interpreted as acceptance of a norm, in the *concept* of belief: unless one accepts the norm of truth for one's acceptance that  $p$ , it will fail to be a belief that  $p$  *in one's own eyes*; but failing to be a belief in one's own eyes is not the same as failing to be a belief in the eyes of the universe. Making the commitment to truth a condition on *conceiving* an attitude as a belief rather than a condition on an attitude's *being* a belief wouldn't exclude toddlers and lower animals from having beliefs; it would just rule out their being able to classify their beliefs as such.

<sup>44</sup> This section draws on Shah 2001.

<sup>45</sup> Velleman 2000b further explicates the attitude of regarding-as-true in terms of its distinctive motivational role.