# KNOWING WHAT ONE BELIEVES – IN DEFENSE OF A DISPOSITIONAL RELIABILIST EXTROSPECTIVE ACCOUNT

Michael Roche

We seem to enjoy a special kind of access to some of our mental states. First, we seem able to know about such states in ourselves via a special, distinctively first-personal method, that is, one that cannot be used to attain knowledge of others' mental states. Second, our knowledge of such states seems epistemically superior to, and different in kind from, any knowledge that others might attain of our mental states. To use Byrne's (2005) terminology, we seem to enjoy both *peculiar* and *privileged* access to some of our mental states. This alleged access makes possible peculiar and privileged self-knowledge.[i] Notable exceptions notwithstanding, philosophers have tended to accept this intuitive picture.[ii] The challenge has been to explain *how* such access is possible.

Attempts to meet this challenge have mostly produced introspective accounts. The inner sense and acquaintance theories are in this tradition. According to the former, we are each equipped with a perception-like mechanism the function of which is to detect one's own mental states.[iii] According to the latter, we are each able to bear a non-causal, metaphysically direct relationship to (some of) our own mental states.[iv] On these accounts one attains peculiar and privileged self-knowledge by directly attending to one's mind.

In contrast to introspection is extrospection. Defenders of extrospective accounts claim that one attains peculiar and privileged self-knowledge, not by directly attending to one's mind, but rather by thinking about the non-mental world. This seemingly paradoxical approach is inspired by the following remark by Evans (1982):

[I]n making a self-ascription of belief, one's eyes are, so to speak, or occasionally literally, directed outward—upon the world. If someone asks me "Do you think there is going to be a third world war?," I must attend, in answering him, to precisely the same outward phenomena as I would attend to if I were answering the question "Will there be a third world war?" (p. 225)

The rough idea behind extrospective accounts is that one can attain peculiar and privileged self-knowledge by "looking through" one's mental states, attending instead to their worldly contents (e.g., a third world war).[v] Such accounts do not posit any special mechanisms or metaphysical relations. Rather, they seek to explain the phenomenon at issue solely in terms of those cognitive mechanisms and processes needed to account for ordinary first-order cognition. This is an attractive aspect of extrospective accounts.[vi]

I assume in this paper that we enjoy peculiar and privileged access to (some of) our *beliefs*.[vii] My primary aim is to develop and defend a novel extrospective account of this access. Producing an adequate account of this kind is more difficult than is often thought, for there are various under-appreciated obstacles facing such a project. Exposing these obstacles is a secondary aim of this paper.

My account shares certain features with an account recently defended by Byrne (2005, 2011). In sections one and two I describe Byrne's account and argue that it rests on a contentious assumption concerning the relationship between judgment and belief. Providing an adequate extrospective account of peculiar and privileged access to our beliefs without helping oneself to this assumption is one of the obstacles mentioned in the

previous paragraph. In the remaining sections I develop and defend a novel extrospective account that overcomes this obstacle and others.

## 1. BYRNE ON PECULIAR AND PRIVILEGED KNOWLEDGE OF BELIEF

Central to Byrne's account is the notion of an *epistemic rule*. Such a rule is a conditional of the form: if conditions C obtain, believe that p. Byrne stipulates that one *follows* an epistemic rule on a particular occasion if and only if one comes to believe that p because one recognizes that the antecedent conditions C obtain (p. 94).[viii] Byrne attempts to account for our peculiar and privileged access to our beliefs by appealing to the epistemic rule, BEL:

BEL  If p, believe that you believe that p.

Byrne claims that one's recognizing that p is true suffices for one's coming to believe that p (p. 96). Thus, if one follows BEL, the truth of the resulting second-order belief is guaranteed.[ix] Byrne calls epistemic rules of this kind *self-verifying* (p. 96). In fact, Byrne regards BEL to be even stronger than this. He writes that the following situation will be commonplace: "trying to follow BEL, one investigates whether p, *mistakenly* concludes that p, and thereby comes to *know* that one believes that p" (p. 98). Because one can know that one believes that p only if one believes that p, this shows that Byrne regards "mistakenly concluding that p" to suffice for coming to believe that p. Let's stipulate that

one *merely tries to follow* BEL on a particular occasion if and only if one comes to believe that one believes that p because one mistakenly concludes (and so does not recognize) that p is true. If the above points are correct, then merely trying to follow BEL (in the sense just described) guarantees the truth of the resulting higher-order belief. Byrne calls epistemic rules of this kind *strongly self-verifying*.[x]

Byrne's claim that BEL is strongly self-verifying thus seems to assume the following:

[Assumption]  One's (correctly or incorrectly) concluding that p is true suffices for one's coming to believe that p.

What is it to conclude that p is true? Perhaps it is to *judge* that p is true. A second possibility—assuming that judgment and belief are distinct—is that it is to *come to believe* that p is true. But when treated as distinct, it is quite natural to follow Byrne's claim that "… judging is the act that results in the state of belief …" (p. 102n22). I regard judgment and belief to be distinct states. This is, in part, because I think of a judgment as a typically short-lived mental *event*, whereas I think of a belief as a typically long-standing mental *state*.[xi] Given this conception of judgment and belief, and because I take "concluding that p is true" to be a mental event, it is most natural to understand [Assumption] in terms of judgment:

[Assumption_J]  One's (correctly or incorrectly) judging that p is true suffices for one's coming to believe that p.

In what follows, I take Byrne's claim that BEL is strongly self-verifying to depend upon [Assumption_J]. Its plausibility will be examined in section 2.

Given [Assumption_J], BEL is a hyper-reliable epistemic rule, for it is strongly self-verifying. Because Byrne assumes a reliabilist account of justification, he concludes that BEL is knowledge-conducive: following (or merely trying to follow) BEL will tend to yield knowledge. But can BEL yield peculiar and privileged self-knowledge? Byrne argues that it can (pp. 96-98).

First, BEL cannot be used to attain knowledge of others' beliefs. Concluding that another person believes that p as a result of one's judging that p is true is not a truth-conducive method of reasoning. So BEL seems able to accommodate peculiar access. Second, because BEL is strongly self-verifying, following (or merely trying to follow) BEL will produce higher-order beliefs that are epistemically quite safe.[xii] Plausibly, those methods (whatever their nature) that we use to detect others' beliefs are not strongly self-verifying, in which case BEL seems able to accommodate privileged access.

Finally, notice that BEL is what Byrne calls a *neutral* epistemic rule (p. 94). Although following (or merely trying to follow) BEL results in a belief about a belief, the process leading to this result does not involve the recognition of any of the rule-follower's mental states. Suppose that S comes to believe that he believes that the penny has landed heads because he has recognized that the penny has landed heads. S thus follows BEL. The point about neutrality is that S forms the higher-order belief via the recognition of the penny's landing heads, not via the recognition of any of his mental states. An example of a non-neutral epistemic rule is: if you are angry, believe that your blood pressure is elevated.

The importance of BEL's neutrality cannot be overstated. If it were not neutral, then one could come to believe that one believes that p only by first recognizing in oneself a particular mental state. A worry about circularity would loom. Moreover, to give up on neutrality would be to give up on that which is most distinctive and interesting about extrospective accounts of self-knowledge, namely, their claim that self-knowledge can be attained by attending to the non-mental world. Following Byrne, I regard the account's neutrality as essential to it.

## 1.1. TWO PROJECTS: EPISTEMOLOGICAL AND PSYCHOLOGICAL

Note that there are at least two distinct, but related, projects to keep in mind when considering self-knowledge. One project is to make a compelling case that some particular first-personal method is, in fact, the first-personal method that we use to form beliefs about our mental states. This is primarily a *psychological* project.[xiii] A second project is to argue of some particular first-personal method that it can yield peculiar and privileged self-knowledge. This is primarily an *epistemological* project. Thus, even if BEL can accommodate peculiar and privileged knowledge of one's beliefs, this is not to say that BEL is the first-personal method that humans use to acquire such knowledge.[xiv]

Because I do not have the space in this paper to address both projects, I will proceed on the assumption that a BEL-based account is psychologically correct of us: following (or merely trying to follow) BEL is our first-personal method for forming beliefs about our beliefs. My task is to show that this method can yield peculiar and privileged self-knowledge.

## 2. QUESTIONING [ASSUMPTION$_J$]

I explained above that Byrne's claim that BEL is strongly self-verifying seems to depend upon [Assumption$_J$]. Is this claim about the connection between judgment and belief true? This question has recently received quite a bit of attention from philosophers.[xv] The kind of case at the center of this debate is one where an individual allegedly judges that p, but fails to exhibit many of the dispositions normally associated with believing that p. Peacocke (1999) describes a case of this kind:

> Someone may judge that undergraduate degrees from countries other than her own are of an equal standard to her own, and excellent reasons may be operative in her assertions to that effect. All the same, it may be quite clear, in decisions she makes on hiring, or in making recommendations, that she does not really have this belief at all. (pp. 242–243)

Here, Peacocke clearly has in mind a conception of belief on which believing that p involves having various dispositions.[xvi] One who conceives of belief in this way will likely regard [Assumption$_J$] as implausibly strong.

Suppose that Peacocke's administrator comes to believe that she believes that undergraduate degrees from other countries and her own are of an equal standard on the basis of having judged that they are of an equal standard. On these assumptions, she has followed (or merely tried to follow) BEL. If Peacocke's analysis of this case is correct, her higher-order belief is false: despite judging that undergraduate degrees from other

countries and her own are of an equal standard, she does not believe this. In self-ascribing this belief, she is, in Peacocke's words, "relying on the holding of the normal relations between judgment and belief which are not guaranteed to hold" (p. 243). If Peacocke is correct about the contingency of this connection, then [Assumption$_J$] is false, and BEL is not (strongly) self-verifying.[xvii]

Of course there are philosophers who endorse [Assumption$_J$]. Zimmerman (2006), for example, finds "unassailable" the claim that "if a subject judges that p at $t$ that subject believes that p at $t$" (p. 365). Zimmerman addresses Peacocke's example, offering several interpretations that do not violate the alleged "unassailable" fact. On his preferred interpretation, the administrator, upon (sincerely) judging that undergraduate degrees from other countries and her own are of an equal standard, *gains* the belief that they are of an equal standard. She then *loses* this belief when not attending to her judgment.

I am sympathetic to the idea that one's believing that p is (in part) a matter of one's possessing certain dispositions (behavioral and otherwise). I thus incline towards Peacocke's side of this debate. Justifying this position, though, is not a goal of this paper. My more modest aim is merely to show that there is significant disagreement over whether [Assumption$_J$] is true. That Byrne's account appears to rest on such a contentious assumption is certainly less than ideal. Can a BEL-based account succeed without [Assumption$_J$]? I think that it can and will develop such an account in the remaining sections.

## 3. A NEW BEL-BASED ACCOUNT OF KNOWLEDGE OF BELIEF

I noted in section 1.1 that I will proceed on the assumption that a BEL-based account is psychologically correct of us. My aim in what remains is to show that such an account can accommodate peculiar and privileged access to our beliefs. I will do so without relying on [Assumption$_J$]. As a result, BEL, in my hands, is perhaps not (strongly) self-verifying.[xviii] My account is consistent with there being a much *looser* connection between judgment and belief than is apparently supposed by Byrne.

In the absence of [Assumption$_J$], a different claim (or set of claims) regarding the connection between judgment and belief is needed in order to show that following (or merely trying to follow) BEL can yield peculiar and privileged self-knowledge. Providing this supplementation is the task of this section. My account supplements BEL with three claims. The first concerns a tendency of individuals who believe that p. The second concerns a tendency of individuals who do not believe that p. The third concerns the way in which BEL is used. I conclude that BEL, when supplemented with these claims, can accommodate the peculiar and privileged access that we allegedly have to our beliefs.

### 3.1 [BELIEVE, JUDGE] AND [~BELIEVE, ~JUDGE]

The first two of the three aforementioned claims are as follows:

[Believe, Judge]  Those who believe that p at a time, *t*, and who consider at *t* whether p is true will tend to end their deliberation (at some later time) by judging that p is true, if, when deliberating, they (i) make no use of new information and (ii) make no new use of previously stored information.

[~Believe, ~Judge]  Those who do not believe that p at a time, *t*, and who consider at *t* whether p is true will tend to end their deliberation (at some later time) without judging that p is true, if, when deliberating, they (i) make no use of new information and (ii) make no new use of previously stored information.

Notice that [Believe, Judge] and [~Believe, ~Judge] assert mere tendencies and so are not refuted by a single case. They are thus not directly threatened by the above concerns regarding [Assumption$_J$].[xix]

Put a bit more simply, [Believe, Judge] and [~Believe, ~Judge] express the following claim, which is central to my account:

[Central Claim]  An individual's judgment concerning a proposition, p, will tend to align with—and thus indicate—her belief attitude towards p held at the time that she began to consider whether p, if, when deliberating, the two conditions about the use of information hold.

I am using "alignment" as follows: judging that p is true aligns with a prior belief that p; judging that p is false aligns with a prior disbelief that p; and neither judging that p is true nor judging that p is false aligns with a prior lack of belief and lack of disbelief that p.

Notice that these claims are formulated in terms of alignment between one's judgment concerning p and one's belief attitude towards p held *at the start of the deliberation*. The reason for this formulation is, in part, due to an objection to Byrne (2005) from Gertler (2011). I will say more about this in section 3.5.

In order to explain the need for the two conditions regarding the use of information in [Believe, Judge] and [~Believe, ~Judge], it will be helpful to have at my disposal a simple case. Consider Gareth, a normal adult who believes that Hillary Clinton will be the next president and who lacks the belief that Joe Biden will be the next president.[xx] Let "h" be the proposition *Hillary Clinton will be the next president* and "j" be the proposition *Joe Biden will be the next president*.

Imagine that Gareth considers whether h. Further, imagine that while considering the matter, he hears on the news that Hillary Clinton has been arrested on corruption charges. Surely, this would cause Gareth to either withhold judgment about the proposition, or judge that it is false; he almost certainly would not judge that it is true. This new information would prevent Gareth from judging in alignment with his prior belief attitude towards h. The same kind of case can easily be constructed for Gareth's lack of belief that j. Suppose that while considering whether j, he learns that Hillary Clinton has retired from politics. This new information might cause Gareth to judge that j is true, in which case his judgment would not align with his prior belief attitude towards j. Thus are the

reasons for the restriction concerning the use of new information in [Believe, Judge] and [~Believe, ~Judge].

Imagine next that Gareth believes the following three propositions: President Obama will endorse Joe Biden, if the latter decides to run; Joe Biden will decide to run; and whomever President Obama endorses will be the next president. The contents of these three beliefs entail that j. As stipulated above, however, Gareth believes that h. These other beliefs have (for whatever reason) failed to extinguish that belief. Perhaps Gareth has just recently formed the first of the three beliefs and has not yet had a chance to reflect on its relevance to his belief that h.

The important point is that if Gareth were to consider whether h, the relevance of these other beliefs might become apparent. If so, then they might prevent him from judging that h is true, and so might prevent him from judging in alignment with his prior belief attitude towards h. And the same point can be made with respect to Gareth's lack of belief that j. By making new use of the previously stored information just described, Gareth might judge that j is true. Thus are the reasons for the restriction concerning making new use of previously stored information in [Believe, Judge] and [~Believe, ~Judge].

## 3.2 THE CASE FOR [BELIEVE, JUDGE] AND [~BELIEVE, ~JUDGE]

I will now make the case for [Believe, Judge] and [~Believe, ~Judge]. Consider, again, Gareth's belief that h. Suppose that this belief is an explicit standing belief in something like the following sense: (a) Gareth has in the past endorsed h; (b) Gareth has stored the

endorsed content in memory; and (c) Gareth can readily recall the endorsed content.[xxi] Next, suppose that Gareth begins to consider whether h is true. On the face of it, Gareth's consideration of this proposition would seem to "trigger" his stored endorsement of it, leading him to judge that h is true. Acquiring a new piece of information, or making new use of previously stored information, could interfere with such triggering. But in the absence of these potential interferences, Gareth's explicit standing belief should lead him to judge that h is true. This suggests that [Believe, Judge] is true of explicit standing beliefs of the sort just described.

Is [Believe, Judge] plausible for non-explicit beliefs as well? To see that it is, consider Gareth's implicit belief that Hillary Clinton has ten toes. Suppose that he has this belief by virtue of satisfying the following conditions: (a') he has not previously endorsed the proposition *Hillary Clinton has ten toes*, and so (trivially) has not stored the endorsed content in memory; but (b') if he were to consider the proposition, he would assent to it without acquiring new information/evidence about the matter.[xxii] Similarly, and using these same criteria, most adults implicitly believe that the moon is not made of cheese, that 307 is less than 313, and so on.[xxiii]

The case for [Believe, Judge] is quite strong for implicit beliefs. If we assume that assenting to p is the same as (or involves) judging that p is true, then to implicitly believe that p is, in part, to be such that: if one were to consider whether p is true, one would judge that p is true without acquiring new information about p. Thus, if Gareth implicitly believes that Hillary Clinton has ten toes, then if he were to consider whether she has ten toes, he should judge that she does without acquiring new information about the matter. So, if he makes no use of new information (and makes no new use of previously stored

information) when considering whether Hillary Clinton has ten toes, he should judge that she does. [Believe, Judge] thus seems true of implicit beliefs of the sort just described.

Next, consider [~Believe, ~Judge]. Recall that Gareth does not believe that j. Suppose that Gareth, at time *t*, begins to consider whether j is true. Either Gareth has previously endorsed this proposition or not. Either way, [~Believe, ~Judge] is safe. If he has not previously endorsed the proposition, then he will satisfy condition (a') at *t*. If Gareth were to judge that j is true without acquiring new evidence/information, then he would also satisfy condition (b') at *t*. But then, contrary to the initial hypothesis, Gareth would have (implicitly) believed that j at *t*. Consistent with the initial hypothesis, then, he can judge that j is true only if his judgment is based on new evidence/information. If it is not—that is, if the condition about new information holds—then Gareth should not judge that j is true.

Suppose, next, that Gareth has previously endorsed j. Given that he does not (explicitly) believe that j, this prior endorsement has either not been stored in memory, or is not easily accessible. The prior endorsement should then have little or no influence on his consideration of the matter. But then because Gareth does not believe that j, it is difficult to see why he would judge that j is true, if he makes no use of new information and makes no new use of previously stored information when considering whether p.

Thus, on the assumption that Gareth does not believe that j, he should be expected not to judge that p is true, if the two restrictions concerning the use of information hold. I take these considerations to make a strong case for [~Believe, ~Judge].

## 3.3 PEACOCKE'S CASE REVISITED

Before turning to the significance of [Believe, Judge] and [~Believe, ~Judge], I wish to consider an objection to these claims suggested by the discussion from section 2. I suggested there that Peacocke's administrator, despite judging that degrees from other countries and her own are of an equal standard, does not believe that they are. Although not believing that p does not entail *dis*believing that p (i.e., believing that not-p), it is of course possible that she believes that degrees from other countries and her own are *not* of an equal standard. Suppose that this is so.

If she were to consider whether degrees from other countries and her own are *not* of an equal standard, and if when doing so she made no use of new information and no new use of previously stored information, would she judge that they are not? The answer seems to be no, for she routinely judges that they are of an equal standard. But then this case goes against the tendency asserted by [Believe, Judge]. Moreover, supposing that she does not believe that degrees from other countries and her own are of an equal standard, this case also goes against the tendency asserted by [~Believe, ~Judge].

I believe the threat here is only apparent. Even if we grant that Peacocke's administrator disbelieves the proposition that she judges to be true (in addition to not believing that proposition), it of course does not follow that this kind of case is common. It must be common though in order to threaten [Believe, Judge] and [~Believe, ~Judge], for these principles assert mere tendencies. Indeed, Peacocke's remark that the "normal relations between judgment and belief" fail in his case suggests that he views the case to be abnormal. This seems correct to me.

## 3.4 A DISPOSITIONAL RELIABILIST EXTROSPECTIVE ACCOUNT

Notice that to follow (or merely try to follow) BEL on a particular occasion is simply to engage in a certain kind of reasoning. It is to come to believe that one believes that p because one has correctly (or incorrectly) judged that p is true. This is true of any epistemic rule.[xxiv]

Let me stipulate that S *uses* BEL on a particular occasion if and only if S considers whether p is true and follows (or merely tries to follow)[xxv] BEL if and only if she judges that p is true. To illustrate, suppose that S uses BEL on a particular occasion. If, upon considering whether p, S judges that p is true, then S will follow BEL on that occasion. That is, S will come to believe that she believes that p because of her judgment that p is true. If S does not judge that p is true, then S will not follow BEL on that occasion. So S will not come to believe that she believes that p, at least not as a result of using BEL. To use BEL (in the sense just defined) is thus simply to engage in a certain kind of reasoning if and only if one judges that p is true.

We are now in a position to appreciate the significance of [Believe, Judge] and [~Believe, ~Judge]. Suppose that S wonders whether he believes that p and consequently uses BEL. If S believes that p, then, as per [Believe, Judge], S should be expected to judge that p is true, if the two restrictions concerning information use hold. Suppose that they do hold. If S judges that p is true, then S will follow BEL and will thereby come to believe that he believes that p. This is the correct result, for S believed that p just prior to using BEL. On the other hand, if S does not believe that p, then, as per [~Believe, ~Judge], S should be expected to not judge that p is true, if the two restrictions

concerning information use hold. Suppose, again, that they do. If S does not judge that p is true, then S will not follow BEL on that occasion and thus will not come to believe that he believes that p, at least not as a result of using BEL. And this is the correct result, for S did not believe that p just prior to using BEL.[xxvi]

This shows that given [Believe, Judge] and [~Believe, ~Judge], BEL is a reliable belief-forming method when used in the *appropriate way*. When one using BEL neither makes use of new information nor makes new use of previously stored information when considering whether p, he should be expected to come to believe that he believes that p if and only if he believed that p just prior to using BEL. The third claim of my BEL-based account asserts that BEL is, for the most part, used in the appropriate way:

> [USE]  Those using BEL will most often make no use of new information and make no new use of previously stored information when considering whether p.

[USE] is consistent with cases where one using BEL either makes use of new information, or makes new use of previously stored information, when considering whether p. It claims only that such cases are infrequent relative to cases where BEL is used and the two restrictions concerning information use hold.

If using BEL is our first-personal method for forming beliefs about our beliefs, then [Believe, Judge], [~Believe, ~Judge], and [USE] show that this method is a reliable one: it most often produces true beliefs. Assuming a reliabilist account of justification, these beliefs are justified and, if true, should qualify as knowledge. My account endorses a reliabilist epistemology.

The level of reliability here would seem to be much greater than the reliability of those methods (whatever their nature) that we use to form beliefs about others' beliefs. First, the inference from *p* to *I believe that p* is exceedingly simply. Second, the higher-order belief's truth is in no way dependent upon the truth of the first-order judgment. BEL will be reliable even if its users routinely reason from false (first-order) premises. My account thus seems able to accommodate privileged access. Moreover, Byrne's point that BEL cannot be used to attain knowledge of others' beliefs still stands. My account can accommodate peculiar access as well.

To sum up, my extrospective account supplements BEL with three claims: [Believe, Judge], [~Believe, ~Judge], and [USE]. The account is *dispositional* in the sense that [Believe, Judge] and [~Believe, ~Judge] assert mere tendencies of those who believe and do not believe that p, respectively. Because it also endorses a *reliabilist* view of justification, it can be called a "dispositional reliabilist extrospective account." I defend [USE], the third claim of my account, in section 4.


### 3.5 AN ATTRACTIVE FEATURE OF THE ACCOUNT


Before turning to that task, however, I wish to quickly highlight what I take to be an attractive feature of my account. Gertler (2011) claims that an adequate account of our special access to our beliefs must accommodate, not only peculiar and privileged access, but also the following alleged fact: "[w]here $t_1$ and $t_2$ are separated only by a moment in which the subject uses some procedure to determine whether she believes that p … [i]f (at $t_1$) I do not believe that p, and I happen to wonder whether I believe that p, I will not

(at $t_2$) self-attribute the belief that p" (p. 128). Gertler criticizes Byrne's account on the grounds that it runs afoul of this alleged fact. On his account, one who does not believe that p when turning to BEL might come to believe that she believes that p as a result of judging that p is true.[xxvii] Moreover, his account gives us no reason to think that such an occurrence should be infrequent. Adequately responding to Gertler's point is an additional obstacle facing extrospective accounts of self-knowledge.

Notice that on my account such cases should be relatively infrequent. Given [~Believe, ~Judge], one who does not believe that p will tend not to judge that p is true, if the two restrictions concerning information use hold. Given [USE], these conditions most often hold for those using BEL. My account, but not Byrne's, can thus accommodate the following slightly weaker version of Gertler's claim: where $t_1$ and $t_2$ are separated only by a moment in which the subject uses some procedure to determine whether she believes that p, if (at $t_1$) I do not believe that p, and I happen to wonder whether I believe that p, I will *most often* not (at $t_2$) self-attribute the belief that p.

Must an adequate account of self-knowledge accommodate the stronger version of this claim? I think that it need not, although I cannot address that issue here.

## 4. DEFENDING [USE]

My account's commitment to [USE] might appear to be in tension with its commitment to neutrality. [USE] might appear plausible only if those using BEL monitor *internal matters* in order to ensure that they make no use of new information and make no new

use of previously stored information when considering whether p.[xxviii] I must defend

[USE] in a way that is consistent with neutrality. In section 4.1 I argue that those using

BEL will most often make no use of new information when considering whether p. In

section 4.2 I argue that those using BEL will most often make no new use of previously

stored information when considering whether p.

### 4.1 USING NEW INFORMATION

My argument in this section consists of three stages. Although the third stage is the most

important to my defense, each of the prior stages paves the way for the next.

### 4.1.1 THE ABSENCE OF RELEVANT NEW INFORMATION IN ONE'S IMMEDIATE ENVIRONMENT

Recall Evans' claim that "in making a self-ascription, one's eyes are, so to speak, *or occasionally literally,* directed outward—upon the world" (p. 225, my emphasis). One

must caution against making too much of the emphasized part of this quotation. In many

cases where S might wonder whether she believes that p, there will be no new

information in her immediate environment that she takes to be relevant to whether p.[xxix]

This is significant because when there is no new information available in S's

immediate environment that she takes to be relevant to whether p, she should not be

expected to make use of new information when considering whether p. S will then satisfy

the restriction concerning new information in a way that does not violate neutrality.

## 4.1.2 NON-PSYCHOLOGICAL INTERPRETATIONS OF "DO YOU/I BELIEVE THAT P?"

Next, consider the following case. Suppose that Billy asks Suzy, while both look up at a darkening sky, whether she believes (or thinks) that there is going to be a storm. In this case, there is clearly new information available in Suzy's immediate environment that she takes to be relevant to whether it will storm. She is already looking up at the sky. For this reason, if Suzy were to use BEL, she would almost certainly make use of new information when considering whether it will storm.

Billy seemingly asks Suzy about her mind. He asks her whether she *believes* (or *thinks*) there is going to be a storm. But this is not the only interpretation of the case, nor is it obviously the most natural. A question of the form "do you believe that p?" (posed to S), can be used in at least two ways. First, it might be used to ask whether S has a particular belief at the time the question is posed; call this the *psychological interpretation*. Second, it might be used to ask about p itself; call this the *non-psychological interpretation*.[xxx]

On the non-psychological interpretation, Billy is not asking Suzy about her mind. Rather, he is merely doing what is quite ordinary, namely, talking with another individual about the (non-mental) world. On this interpretation, Billy's question is equivalent to the question (posed to Suzy) "will it storm?" But to answer this kind of question does not require self-knowledge; it does not require using one's method for attaining peculiar and privileged knowledge of one's beliefs. Consider, for example, answering affirmatively to the question "is Germany part of the European Union?" While this answer is surely the

result of mental states and operations, it need not be the result of one's being aware of such states and operations.

It might be objected that even on the allegedly non-psychological interpretation, Billy's question is *partly* about Suzy's mind. Although his primary concern might be the possibility of a storm, he is also interested in Suzy's opinion on the matter. He poses his question to *Suzy*, after all. Notice again though that Billy is doing something quite unremarkable. He is simply talking with another individual about the (non-mental) world. This kind of engagement between people is ubiquitous. Surely, we do not want to claim that metacognition is this pervasive! We should thus deny that Suzy ought to use her method for attaining peculiar and privileged knowledge of her beliefs when answering Billy's question. But so long as this is denied, the case (interpreted non-psychologically) is irrelevant to the issue at hand.

The main point is that care is required when thinking about the first half of [USE], viz., the claim that those using BEL will most often make no use of new information when considering whether p. On its surface, the case of Billy and Suzy appears to be the kind of case that the first half of [USE] alleges is relatively infrequent. But upon closer examination it might be best interpreted non-psychologically, in which case it is irrelevant to the truth of [USE].

## 4.1.3 THE INSTABILITY OF A PSYCHOLOGICAL, YET RELEVANTLY

## INFORMED CONTEXT

Consistent with the previous two stages of my argument are cases of the following kind: a subject wonders whether she believes that p (interpreted psychologically) and yet new information that she takes to be relevant to whether p is available in her immediate environment. The subject should use BEL, but risks making use of new information when considering whether p. [USE] alleges that such cases are relatively infrequent.

The reason for thinking that this is true has to do with the norms governing belief. For the most part, we aim for true beliefs.[xxxi] When new information that one takes to be relevant to whether p is true is available in one's immediate environment, one is seemingly in a strong position to evaluate whether p is true, and thus to ensure that one has the appropriate belief attitude towards p. So when such information is immediately available, any interest that one might have as to whether one believes that p should be expected to *shift* to whether p is true. Because this latter interest concerns the (non-mental) world, it would be inappropriate to use BEL or any other first-personal method for attaining self-knowledge.

Suzy has immediately available to her a wealth of information that she (correctly) takes to be relevant to whether there will be a storm. She is thus in a strong position to determine whether there will be a storm. Assuming that Suzy aims to have true beliefs, if and when she considers the psychological question of whether she believes that there is going to be a storm, she should be expected to take advantage of her strong position. She should evaluate the available information, integrating it with other information that she

might also have in mind. Failure to do so would demonstrate an indifference towards her (possible) belief's truth. If Suzy lacks this sort of indifference, then any interest in whether she believes that there is going to be a storm should quickly give way to an interest in the possibility of the storm itself.

The immediate availability of new information that one takes to be relevant to whether p is true thus makes for an *unstable environment* with respect to interest in whether one believes that p. If S becomes interested in whether she believes that p, this (psychological) interest should be expected to shift to the non-psychological question of whether p, if new information that S takes to be relevant to whether p is immediately available. Once shifted, though, the use of BEL is inappropriate.

There is thus reason to think that when one does, in fact, use BEL, there will most often be no new information in one's immediate environment that one takes to be relevant to whether p. Accordingly, those using BEL will most often make no use of new information when considering whether p.


## 4.2 MAKING NEW USE OF PREVIOUSLY STORED INFORMATION


In section 3.1, I described the case of Gareth. He believes that h, but also has three beliefs that together entail that not-h. If Gareth were to consider whether h is true, these beliefs might prevent him from judging that it is. In other words, he might make new use of previously stored information. [USE] claims that those using BEL will most often not make new use of previously stored information.

The problematic kind of case involves a subject, S, who has previously stored information that is at odds with S's current belief attitude towards p. This feature of the previously stored information is important, for it shows that the information has (for whatever reason) failed to alter S's belief attitude towards p. Three possibilities for why this might occur come to mind.

First, the previously stored information might not be easily accessible to S. Perhaps S would easily see the information's relevance to p, if only he could access it. Given this fact about access, S would most likely not make new use of this information when using BEL. Accordingly, this first possibility is not problematic for [USE].

Second, the previously stored information's relevance to p might not be readily apparent to S. Perhaps S could see its relevance to p only if he were to carefully consider the matter. The case of Gareth (from 3.1) is not like this, for the relevance of the three beliefs to h is fairly obvious. Imagine, then, a case where the relevance is less obvious. How likely is it that in such a case the subject would make new use of that information when using BEL?

While this is difficult to say in the abstract, it seems clear that the less obvious (to S) the information's relevance to p is, the less likely that S will make new use of that information when using BEL. Because inquiries concerning one's mind are typically answered (and are expected to be answered) fairly quickly, those using BEL will most often make no new use of previously stored information, if that information's relevance to p is not readily apparent to them.

Third, S might possess information that is easily accessible (to S) and obviously relevant (from S's perspective) to whether p, but which is at odds with S's belief attitude

towards p. For whatever reason, the information failed to impact S's attitude when the information was was first acquired, or the attitude was first formed. If S were to use BEL, he would immediately see the information's relevance to whether p and would make new use of it.

While I admit that this kind of case is possible, I suspect that it is infrequent. I suspect that information that is easily accessible (to S) and obviously relevant (from S's perspective) to whether p will most often impact S's belief attitude towards p at, or around, the time that the information was first acquired, or the attitude was first formed. If this is correct, then the second half of [USE] is not threatened by this third possibility.

CONCLUSION

I have defended a novel account of how we attain peculiar and privileged knowledge of our beliefs. My account, like Byrne's (2005), is based on the epistemic rule, BEL. Unlike Byrne's account, mine does not rest on [Assumption$_J$], the claim that one's judging that p suffices for one's coming to believe that p. This means that BEL, in my hands, need not be (strongly) self-verifying. I have argued, however, that my account rests on surer footing as a result of abandoning this contention assumption.

I have replaced [Assumption$_J$] with [Believe, Judge], [~Believe, ~Judge], and [USE]. Together, these claims show that using BEL is a highly reliable method for forming beliefs about one's beliefs. The account is dispositional because [Believe, Judge] and [~Believe, ~Judge] assert mere tendencies of those who believe that p and do not believe

that p, respectively. It is reliabilist because it endorses a reliabilist account of epistemic justification. Finally, the account is extrospective in that it claims that we attain peculiar and privileged knowledge of our beliefs by thinking about non-mental matters. I have defended [Believe, Judge], [~Believe, ~Judge], and [USE] in ways that are compatible with the neutrality that is distinctive of extrospective accounts of self-knowledge. The result is a dispositional reliabilist extrospective account of peculiar and privileged knowledge of belief.

*Mississippi State University*

---

[i] Not all self-knowledge is peculiar and privileged. I might know, for example, that I *feel anxious* around my neighbor on the basis of my (observed) behavior around him.

[ii] Ryle (1949) is one such exception. Carruthers (2011) offers an updated defense of Ryle's position.

[iii] Recent defenders of this theory include Armstrong (1981), Goldman (2006), Lycan (1996), and Nichols and Stich (2003).

[iv] Recent defenders of this theory include BonJour (2003), Chalmers (2003), Fumerton (1995), and Gertler (2012).

[v] Given this metaphorical "looking through," such accounts are often called *transparency accounts*. But note that Carruthers (2011) uses this label in a different, non-standard way.

[vi] Recent defenders of extrospective accounts include Byrne (2005, 2011), Dretske (1994), Fernández (2013), Gordon (2007), and Moran (2001).

[vii] Some beliefs might be either inaccessible or accessible only via the kind of third-personal reasoning used to know about others' beliefs.

[viii] Unless otherwise noted, all references to Byrne in this section are to his 2005.

[ix] Byrne (2011, p. 206n6) admits that strictly speaking this is too strong, for there will be a temporal gap between the recognition that p and the formation of the higher-order belief. Because this gap will be extremely short, he believes the possibility can be set aside. I too will set aside this possibility in what follows.

[x] This label, but not the idea, is from Byrne (2011, p. 206). There, he offers the following rule as one that is self-verifying, but not strongly self-verifying: if p, believe that one knows that p.

[xi] Carruthers (2010) writes the following, which I largely agree with: I take judgments to be events of belief-formation … Judgments are a kind of active, occurrent, mental event, which when stored give rise to dormant, standing-state, beliefs (p. 78).

[xii] See Williamson (2001) for a discussion of safety.

[xiii] This is not to say that philosophers cannot engage in this project.

xiv I thank Brie Gertler for making clear the importance of this distinction.

xv See, for example, Peacocke (1999), Zimmerman (2006), Gendler (2008), Schwitzgebel (2010), Gertler (2011), and Paul (2012).

xvi Schwitzgebel (2013) defends a dispositionalist account of belief on which the dispositions constitutive of belief include not only behavioral dispositions, but also what he calls *cognitive* and *phenomenal* dispositions.

xvii Schwitzgebel (2010) describes some nice cases intended to illustrate this alleged contingency.

xviii While my account does not depend upon [Assumption_J], it is compatible with it.

xix I will say more about this in section 3.3.

xx Assuming that the holding of contradictory beliefs is possible, this second conjunct is not redundant.

xxi The above three conditions are from Gertler (2011). She takes these conditions to be jointly sufficient for one's having an explicit standing belief.

xxii The clause in (b') concerning new evidence is meant to distinguish S's implicitly believing that p from S's being merely disposed to believe that p by virtue of having ready access to evidence for p. See Gertler (2011).

xxiii Perhaps implicitly believing that p requires the condition that one's assent be *quick*. I have omitted such a condition to simplify the discussion. I believe this is harmless since I am concerned with cases where any such requirement for speed will be satisfied.

xxiv Byrne (2005) describes the rule DOORBBELL: if the doorbell rings, believe that there is someone at the door (p. 94). To follow (or merely try to follow) DOORBELL on

a particular occasion is simply to come to believe that there is someone at the door because one has correctly (or incorrectly) judged that the doorbell is ringing.

xxv I omit this parenthetical in the remainder of this subsection for the sake of readability.

xxvi Notice that I do not claim that S will form the belief that he does not believe that p. I claim only that S will not form the belief that he believes that p. Using BEL is thus not a way to settle the matter of whether one believes that p: one can end up with no belief either way. Still, as I argue below, BEL is a reliable belief-forming method. I thank an anonymous referee for alerting me to this point.

xxvii If, as Byrne claims, BEL is self-verifying, then the higher-order belief would be true. Gertler's complaint is thus not that BEL can lead to false beliefs.

xxviii While [USE] is unique to my account, Gertler (2011, p. 132) presses a somewhat similar worry when discussing Byrne's account.

xxix I do not count information available only via such information-gathering activities as accessing the internet or opening a book as being available in one's "immediate environment." Because people do not engage in such activities when responding to inquiries about their minds, only new information available in one's immediate environment (as I am using the term) is relevant to the use of BEL.

xxx That people use sentences with mentalistic terms without intending to speak of the states/event/properties to which those terms refer should be uncontroversial. Contexts can easily be imagined where, for example, the utterance "I want a beer" is used merely to ask for a beer. While in such a context the speaker likely wants a beer, it does not follow that S's utterance is a report about S's mind/wants.

xxxi Exceptions include beliefs held for pragmatic reasons.

REFERENCES

Armstrong, David. 1981. <u>The Nature of Mind and Other Essays</u> (Ithaca, NY: Cornell
University Press).

BonJour, Laurence, and Ernest Sosa. 2003. <u>Epistemic Justification: Internalism versus
Externalism, Foundations versus Virtues</u> (Malden, MA: Blackwell Publishing
Ltd.).

Byrne, Alex. 2005. "Introspection," <u>Philosophical Topics</u>, vol. 33, pp. 79-104.

------. 2011. "Transparency, Belief, Intention," <u>Proceedings of the Aristotelian Society
Supplementary</u>, vol. 85, no. 1, pp. 201-221.

Carruthers, Peter. 2010. "Introspection: Divided and Partly Eliminated," <u>Philosophy and
Phenomenological Research</u>, vol. LXXX, no. 1, pp. 76-111.

------. 2011. <u>The Opacity of Mind: An Integrative Theory of Self-Knowledge</u> (Oxford:
Oxford University Press).

Chalmers, David. 2003. "The Content and Epistemology of Phenomenal Belief," in
<u>Consciousness: New Philosophical Essays</u>, eds. Quentin Smith and Aleksandar
Jokic (Oxford: Clarendon Press), pp. 220-272.

Dretske, Fred. 1994. "Introspection," <u>Proceeding of the Aristotelian Society</u>, vol. 94, pp.
263-278.

Evans, Gareth. 1982. <u>The Varieties of Reference</u>, ed. John McDowell (Oxford: Oxford
University Press).

Fernández, Jordi. 2013. <u>Transparent Minds: A Study of Self-Knowledge</u> (Oxford: Oxford
University Press).

Fumerton, Richard. 1995. <u>Metaepistemology and Skepticism</u> (Lanham: Rowman and
Littlefield).

Gendler, Tamar Szabó. 2008. "Alief and Belief," <u>Journal of Philosophy</u>, vol. 105, pp.
634–663.

Gertler, Brie. 2011. "Self-Knowledge and the Transparency of Belief," in <u>Self-
Knowledge</u>, ed. Anthony Hatzimoysis (Oxford: Oxford University Press), pp.
135-145.

------. 2012. "Renewed Acquaintance," in <u>Introspection and Consciousness</u>, eds. Declan
Smithies and Daniel Stoljar (Oxford: Oxford University Press), pp. 93-128.

Goldman, Alvin. 2006. <u>Simulating Minds: The Philosophy, Psychology, and
Neuroscience of Mindreading</u> (Oxford: Oxford University Press).

Gordon, Robert. 2007. "Ascent Routines for Propositional Attitudes," <u>Synthese</u>, vol. 159,
pp. 151-165.

Lycan, William G. 1996. <u>Consciousness and Experience</u> (Cambridge, MA: Bradford
Books/MIT Press).

Moran, Richard. 2001. <u>Authority and Estrangement – An Essay on Self-Knowledge</u>
(Princeton, NJ: Princeton University Press).

Nichols, Shaun, and Stephen P. Stich. 2003. <u>Mindreading: An Integrated Account of
Pretense, Self-Awareness, and Understanding of Other Minds</u> (Oxford: Oxford
University Press).

Paul, Sarah K. 2012. "How We Know What We Intend," <u>Philosophical Studies</u>, vol. 161,
no. 2, pp. 327-346.

Peacocke, Christopher. 1999. <u>Being Known</u> (Oxford: Oxford University Press).

Ryle, Gilbert. 1949. <u>The Concept of Mind</u> (Chicago: The University of Chicago Press).

Schwitzgebel, Eric. 2010. "Acting Contrary to Our Professed Beliefs, or the Gulf Between Occurrent Judgment and Dispositional Belief," <u>Pacific Philosophical Quarterly</u>, vol. 91, pp. 531-553.

------. 2013. "A Dispositional Approach to the Attitudes: Thinking Outside of the Belief Box," in <u>New Essays on Belief: Constitution, Content, and Structure</u>, ed. Nikolaj Nottelmann (London: Palgrave Macmillan), pp. 75-99.

Williamson, Timothy. 2001. <u>Knowledge and Its Limits</u> (Oxford: Oxford University Press).

Zimmerman, Aaron. 2006. "The Nature of Belief," <u>Philosophical Studies</u>, vol. 128, pp. 337-379.