

## **‘Belief’ and Belief**

Eric Marcus  
Auburn University

(Forthcoming in the *European Journal of Philosophy*)

Our interest in understanding belief stems from our being creatures who think. To study thought is to study what we are. And philosophers have (mostly) settled on belief—either on its own or insofar as it figures in knowledge—as our basic cognitive state. But how to determine what counts as a belief for the purpose of the investigation? The *scope question*, as I will call it, might seem to have an easy answer: our topic is the state we call ‘belief’. But between paradigmatic belief—the conscious rational state that partly constitutes knowledge—and the purely fanciful mental states of toasters and alarm clocks, there is the following (no doubt less-than-exhaustive) motley of ‘belief’ states: conscious cognitions of humans, such as guesses, that are understood by their bearers to be likely false; person-level states that unconsciously guide the intelligent behavior of rational creatures, such as implicit biases; conscious cognitions of non-human animals (hereafter simply ‘animals’); and sub-personal states that carry information from one place in living bodies to another. Is a theory of belief a theory of all of these states or only some of them? And how should such a question be decided?

My route to answering this question begins (in section one) with an examination of a recent argument to the effect that belief is invariably *weak*: one can believe, even rightly, while acknowledging the possibility, even the probability of error. I show that although one kind of state we refer to using the term ‘belief’ fits this profile, other ‘belief’ states do not. This conclusion is interesting in its own right, but my purpose here is to recruit weak belief to help answer the scope question. When ‘belief’ is heard in a weak

sense, it attributes a state that only a rational creature can be in. I will use this observation as a starting point for an argument that the study of (our) belief should not be constrained by the requirement that the illuminated state be held in common with any non-rational being. In fact, ‘belief’, even in its non-weak sense, does not pick out in its application to animals what it typically does when applied to adult humans. More generally, I aim to support what is sometimes called the Transformative Theory of Rationality, the view that rationality does not merely add powers or complexity to the non-rational mind but transforms it into a different kind of mind altogether.

## **I. Two Senses of ‘Belief’**

According to Grice, “senses are not to be multiplied beyond necessity,” a maxim he calls ‘Modified Ockham’s Razor’.<sup>1</sup> Following others, I will call it *Grice’s Razor*.<sup>2</sup> Its point is to enjoin the student of language to find non-semantic explanations of intuitively different uses of the same word. Both Ockham’s and Grice’s Razors are trivial in themselves—everyone can sign on to them—but virtually impossible to enforce without begging some question or another. ‘Bank’, it is agreed by all, has a different sense in ‘that bank is a perfect place to put your money’ and ‘that bank is a perfect place to picnic on a sunny day’; and the same sense in ‘that bank gives free checking to all of its customers’ and ‘this bank is insured by FDIC’. The tricky cases involve polysemy, where there is a difference in meaning without equivocation.

---

<sup>1</sup> Grice (1989).

<sup>2</sup> Davis (1988) and Hazlett (2007).

Aristotle emphasized long ago that a word such as ‘healthy’ is “said in many ways”: it applies to animals, plants, complexions, activities, foods, organizations, atmospheres, relationships, attitudes, and appetites, among other things.<sup>3</sup> But this is not accidental or *mere* homonymy, as is exemplified by the two senses of ‘bank’. When a word is polysemous, it has several non-accidentally related meanings. And the relatedness of the meaning is a matter not of etymology but, as Robyn Carston puts it, “of speakers’ mental representations and apprehension of distinctness of words and of relatedness of senses.”<sup>4</sup> For the distinct senses exemplified by the same word to exhibit polysemy, it is not enough that the word makes distinct contributions to meaning across contexts. That could be the result of pragmatic inferences. (“I love you” sometimes means “please turn out the light” but that has nothing to do with polysemy.) Polysemy is present when the same word is properly understood as making distinctive *semantic*, i.e., pre-pragmatic-inferential, contributions. But when a word is habitually used to generate a specific pragmatic implication, it can, over time, become a new sense.

Carston again:

Languages change over time and one of the most flexible and dynamic aspects of this is changes in the senses expressed by (substantive) words, which entails the conventionalisation of at least some derived senses.<sup>5</sup>

---

<sup>3</sup> Aristotle (350 B.C.E), ch. 1.

<sup>4</sup> Carston (2020), 110.

<sup>5</sup> Carston (2020), 121.

Because of the inevitability of the conventionalization of pragmatically derived senses, polysemy is ubiquitous. Given this fact (as I now will take it to be), it would be a mistake to operate with the assumption that if the semantic value of two tokens of a term ‘x’ are distinct, it must be because they are *merely* homonymous. The semantic value of ‘healthy’ varies, Grice’s maxim notwithstanding. And ‘belief,’ I will argue in this section, is like ‘healthy’: it is polysemous.

I begin by discussing a non-exclusive distinction between two senses of ‘belief,’ i.e., two interpretations of ‘belief’ that arise in different contexts and on which the word takes different semantic values: the ‘knowledge-minus’ sense and the ‘weak’ sense.<sup>6</sup> Both uses are common in ordinary language and both have figured prominently in philosophical discussions. There is pressure from both directions to eliminate one or the other as not in fact a sense of ‘belief’ at all. I will argue in what follows that we should resist the pressure from both directions and embrace polysemy. Ultimately, my interest here is not in the words at all, but in what we use them to say. However, the analysis of this difference in meaning will be crucial as we bore down on object-level questions in section two. I’ll begin with a quick characterization.

We often appeal to what the agent knows in explaining something that he did. We might do this directly, as in “he checked the attic, knowing that it was their typical

---

<sup>6</sup> I describe the distinction as ‘non-exclusive’ not simply out of caution, but because there are self-evidently many other senses. Here are instances of a few: “I believe him,” “I believe in him”, “He’s such a phony, I don’t think he actually believes anything.”

hiding place” or indirectly, as in “he checked the attic since it was their typical hiding place.” Under very similar circumstances but where we take the agent to be wrong about their typical hiding place, we might say instead “he checked the attic, believing it to be their typical hiding place.” ‘Belief’ is commonly used in this manner to explain the action of an epistemically deficient agent. One would say “she only meant to scare him, as the gun had no bullets—or so she believed” only if one thought (or was pretending to think that) the gun *did* have bullets. In speaking of her believing something, we are attributing ‘belief’ in the knowledge-minus sense: ‘belief<sub>K-</sub>’ as I will write it. Williamson captures this idea in his analysis,

**Belief<sub>K-</sub>:** S believes<sub>K-</sub> iff for all S knows, S knows that p<sup>7</sup>

Especially when used in the first-person, ‘belief’ is commonly used to signal a lack of certainty, as in “I believe he’s run out of bullets; I’m going to make a break for it.” In using ‘belief’ this way, I communicate an awareness that my evidence left me short of what would be required to *know* that his gun was out of bullets, but I nonetheless express an intention to act on that ground. When I believe that p in this sense, I reason from the premise that p in thought and action, despite knowing that I don’t know that p. We also sometimes use belief in this sense to describe someone else, e.g., “he believed that the gun was not loaded, but he wasn’t sure.” I will write this as belief<sub>w</sub>.

**Belief<sub>w</sub>:** S believes<sub>w</sub> iff S (i) grounds their thought and action on p despite (ii) believing<sub>K-</sub> that she doesn’t know that p.

---

<sup>7</sup> Williamson (2000), ch. 18.

These concepts of belief are taken up by theorists of different sorts. Since Davidson, 'belief-desire' explanation has become a shorthand for the explanation of intentional action that gives an agent's reasons for acting. He holds:

Whenever someone does something for a reason, therefore, he can be characterized as (a) having some sort of pro attitude toward actions of a certain kind, and (b) believing (or knowing, perceiving, noticing, remembering) that his action is of that kind.<sup>8</sup>

Why didn't Davidson use the label *knowledge* to cover the second category, so that this kind of explanation might have become famous as 'knowledge-desire' rather than 'belief-desire' explanation? Because Davidson held that the primary reason was the cause of an action, and he, like most others at the time, would have rejected the idea that anything is added to the causal efficacy of belief in virtue of its truth. And the same goes for perceiving and remembering; they are (or so it was generally thought) just belief plus some causally irrelevant stuff.

Others are explicit about the idea that, in explaining action, we should abstract from making any commitments about the world 'outside' of the agent. Fodor, for example, defends the doctrine of Methodological Solipsism, according to which psychology should only be concerned with the formal or non-semantic properties of mental states, on the grounds that only its formal properties could figure in causal explanations:

---

<sup>8</sup> Davidson (2001), Essay 1.

Since, on that assumption, knowledge is involved with truth, and since truth is a semantic notion, it's going to follow that there can't be a psychology of knowledge (even if it is consonant with the formality condition to hope for a psychology of belief).<sup>9</sup>

Stephen Stich defends the very similar Principle of Psychological Autonomy and does so on more or less the same grounds. The thought behind the principle is that

we expect a psychological theory which aims at explaining behavior to invoke only the "purely psychological" properties which are shared by a subject and its [atom-for-atom] replicas....What knowledge adds to belief is psychologically irrelevant.<sup>10</sup>

These passages reflect the fact that the dominant sense of 'belief' in one prominent strand of philosophy of mind corresponds to belief<sub>K</sub>.

John Hawthorne, Daniel Rothschild, and Levi Spectre (henceforth HRS) have recently argued that *belief is weak*.<sup>11</sup> The basis of their argument concerns facts such as the following. Whereas a sentence such as

(a) It's raining but I'm not sure it's raining

---

<sup>9</sup> Fodor (1980), 64.

<sup>10</sup> Stich (1978), 574.

<sup>11</sup> Hawthorne, Rothschild, and Spectre (2016). See also Dorst 2019, Rothschild 2020, and Holguín (forthcoming). See also McGlynn (2013) and Whiting (2013) for earlier work along the same lines.

is “bad-sounding”,

(b) “I believe it’s raining, but I’m not sure it’s raining”

sounds fine. They use this and related contrasts to argue that the evidential standards for assertion are higher than those required for belief. Belief is thus weak, i.e., “the evidential standards that are required for belief are very low.”<sup>12</sup> The strength of evidence in favor of *p* required for someone to rationally *believe* it is such that they might do so while at the same time acknowledging that they don’t possess evidence strong enough to *know* that *p*.

The idea that ‘belief’ can be used to signal uncertainty is not new.<sup>13</sup> What is new is the proposal that such uses reveal *the* ordinary sense of ‘belief’ and thus reflect the weakness of *the* state of belief. But this is wrong.

When we attribute ‘belief’ in belief<sub>K</sub>- contexts (i.e., contexts that give rise to the interpretation of ‘belief’ indicated in the sketch of belief<sub>K</sub>.) we portray the attributee as failing to know. When we attribute beliefs in belief<sub>W</sub> contexts (i.e., contexts that give rise to the interpretation of ‘belief’ indicated in the sketch of belief<sub>W</sub>) we portray the attributee as not taking themselves to know. But of course it doesn’t follow from the fact that subject doesn’t know that the subject thinks they don’t know. From the fact that I describe someone as pulling the trigger because they “believed the gun was not loaded”, it doesn’t follow that they harbor doubts that the gun was not loaded, i.e., it does not

---

<sup>12</sup> Hawthorne, Rothschild, and Spectre (2016), 1394.

<sup>13</sup> See, e.g., Benton (2011), Benton and Turri (2014), and van Elswyk (2018).



follow that *they* might say anything like “I believed the gun was not loaded [but I wasn’t sure].”

There is such a thing as acting in light of what’s true (as one sees it). These are the circumstances in which we say things like: “She drove all the way to Son of a Butcher because they have good lamb.” But if she doesn’t know they have good lamb on account of the fact that they have no lamb, we say instead something like: “She drove all the way to Son of a Butcher because she believed they had good lamb.” But we might use the same formulation if she went merely because she thought it *likely* that they had good lamb. This reveals an ambiguity. Upon hearing the relevant utterance we may be unsure about which sense of belief is at issue, leading us to ask a follow-up question such as “if she wasn’t sure, why did she drive all that way?”. And if the answer is: “Oh, she was completely certain, just wrong”, then I know what was meant was belief<sub>K</sub>. But if the answer is: “I guess she was willing to the chance. She really loves lamb!”, then we know what was meant was belief<sub>w</sub>. I hold, then, pace HRS, that there are at least two senses of ‘belief’.

HRS will point out that my objection to their approach depends on my having altered their definition of weak belief. The weakness of a belief, for them, is a matter of its low-threshold evidential requirement; for me, it is a matter of the subject’s consciousness of the weakness of their evidence. And this difference will be crucial for the argument to follow. By what right do I make this alteration? And why should we consider any subsequent discussion to bear on the topic of HRS’s essay?

I answer as follows. My conception of weak belief fits the linguistic data at least as well as theirs and is better positioned than theirs to make sense of the explanatory uses to which we put the concept. It fits the data just as well since the sentences that

provide our grasp on the idea of weak belief are simply variations on an example in which someone says either of themselves or another that they ‘believe’ something but aren’t sure. The fine-soundingness of such sentences might either be due to (a) belief’s being such that little evidence is required to be in it or (b) it being perfectly conceivable that someone asserts the relevant proposition in a hedged or qualified manner. I take HRS to use the term ‘weak belief’ to designate the state expressed by this class of statements. But it remains to be seen precisely what state this is.

Consider now the following example: Tweedledee believes<sub>K</sub> that if he stands at the edge of the cliff, the fence will prevent him from falling over the edge. Nonetheless, even though he really wants to see the amazing view, he doesn’t go to the edge, because he can’t rid himself of (what he recognizes to be) an irrational fear of falling.<sup>14</sup>

Tweedledum, by contrast, merely believes<sub>w</sub> that if he stands at the edge of the cliff, the fence will prevent him from falling over the edge. We might say of him: “Tweedledum believes that he won’t fall, but he’s not sure.” Since he merely believes<sub>w</sub> that he won’t fall, he doesn’t go to the edge—even though he really wants to see the amazing view. Sometimes, reluctance to act derives from an irrational fear; other times it derives from weak belief. When it is the latter, what explains such reluctance includes a recognition that one’s evidence is weak and thus that hedging would be required if one were to assert the relevant proposition. This is my conception of weak belief. But it’s not clear what, on HRS’s conception, grounds this distinction between the actions of our two

---

<sup>14</sup> This is a variation on the example that Tamar Gendler uses to introduce the notion of an ‘alief’, understood as an arational state that prompts one to act as if one holds a certain belief that one does not in fact hold. See Gendler (2008)

characters. Weak belief, as HRS interpret it, can't do so since belief's (supposedly) low threshold of evidence all by itself cannot constitute the subject's felt need to hedge when asserting the believed proposition. If they're right, then the explanation "He believed that the gun was unloaded, but since he wasn't sure, he double-checked" would be missing a crucial element. Properly supplemented, it would be "He knew he believed that the gun was unloaded, but since he wasn't sure, he double-checked." But the latter is not a more complete version of the former. It's gibberish. The role weak belief plays in explaining behavior depends on it including a consciousness of the weakness of one's evidence, which is how I will understand the concept in what follows.

Let us recall where we are in the argument. The point of this section is to argue that the term 'belief' is polysemous; it has (at least) two senses:  $\text{belief}_K$  and  $\text{belief}_W$ . This result is important, first, because it complicates the investigation into the nature of belief. We cannot circumscribe our topic as whatever the term 'belief' refers to, since 'belief' does not pick out a single state. The philosophical upshot of this discussion will be the topic of the next section. In the meantime, I consider further objections:

HRS also argue that belief is weak by appealing to the phenomenon of 'neg-raising':

When saying that you don't believe  $p$  you often suggest that you believe not- $p$ , i.e. that you disbelieve  $p$ . This is called neg-raising because the negation seems to be read further on in the sentence: not as negating the mental state verb but as negating its content. Neg-raising occurs with all the belief-like mental state expressions we have discussed so far, 'believe', 'think', and 'is of the opinion that'. Neg-raising does not,

however, occur with ‘know’: saying ‘John doesn’t know p’ doesn’t suggest that John knows not-p.<sup>15</sup>

If one denies that John believes that p, it suggests that John believes that not-p. But if one denies that John knows that p, it does not suggest that John knows that not-p. This observation is unimpeachable. But what follows? Their argument postulates that the best explanation for this observed difference is that ‘belief’ is weak, whereas ‘know’ is strong. To be convincing, other competing explanations would have to be considered. There is no space to develop an alternative here, but a close examination of what happens when one negates a knowledge claim shows that if, anything, the phenomenon points in precisely the opposite direction. Ordinarily, to say of someone that he doesn’t know that p is to suggest that they have no idea that p and so do not even believe it. If the speaker emphasizes ‘know’ (“he doesn’t *know* that p”) it suggests that he nonetheless *believes* that p. But negating a knowledge claim *never* suggests that the subject might be in a state that *he* could felicitously express by saying “p, or so I believe”, i.e., that he is in a state of uncertainty about p. When we say that John doesn’t *know* that p (with that emphasis), we imply that he *believes<sub>K</sub>* that p. Although this is not knowledge, it is a state whose measure is epistemic. The shift from ‘know’ to ‘believe’ when a knowledge-claim is negated is *not a* shift from a state whose measure is epistemic to one whose measure is more forgiving. Being wrong cannot change the measure of a state. Yet this is precisely what HRS must say, given that, on their view, knowledge is strong and belief is weak.

---

<sup>15</sup> Hawthorne, Rothschild, and Spectre (2016), 1399.

Here's another way of making this point: To believe<sub>K</sub> is to exercise a capacity to know. And one use of 'believe' is to attribute the exercise of this capacity while remaining neutral on its success. This use postulates a state that is not weak. Thus, the particular way that knowledge fails to exhibit the phenomenon of neg-raising provides evidence that one sense of 'belief' is believe<sub>K</sub>.

HRS also consider and find wanting the view that there is "an ambiguity in the term 'believe'."<sup>16</sup> They do not exactly target my view, as they envisage the two senses as weak and strong, whereas I postulate a distinction between a weak state and one that is indistinguishable (for the subject) from knowledge. Still, their argument raises a challenge that must be addressed. On the targeted view,

'believe' has at least two meanings: it has a weak sense in which it means something like being of the opinion that or suspecting and a stronger meaning in which attributes full belief.... However, the examples above show that this cannot be right. For if 'believe' had two senses it should be possible to attribute belief in the weaker sense but deny it in the stronger sense. But we have seen that we cannot say of a single person that they are of the opinion that p but lack the belief that p. Thus, 'believe' is not ambiguous: it is simply weak.<sup>17</sup>

---

<sup>16</sup> Hawthorne, Rothschild, and Spectre (2016), 1399.

<sup>17</sup> Hawthorne, Rothschild, and Spectre (2016), 1399.

But it is not hard to envision a plausible scenario in which the polysemy of ‘belief’ can be made explicit. Here is a dialogue in which both senses are in play and in which a misunderstanding is corrected by disambiguating them.

Prosecutor: When you fired the gun at the victim, were you trying to kill him?

Jana: No, I was just trying to scare him. I believed [?] the gun was empty.

Prosecutor: But if you merely *believed* [believed<sub>w</sub>] it was empty, shouldn’t you have double-checked?

Jana: No, I didn’t think I had to check! I thought I *had* just meticulously examined it, but it turned out I had examined a different gun. I believed [believe<sub>K</sub>] it was empty, but I was wrong.

This is a genuine ambiguity, one that is resolved only by clarifying which sense of ‘belief’ was at issue.

A different sort of objection will come from the *believe<sub>K</sub>-purist*. HRS are *belief<sub>w</sub>-purists*; they say that everything we call ‘belief’ is weak. The former will acknowledge that, as noted above, ‘belief’ can signal hedging. But they ask: why think that these hedging uses signify a new mental sort of mental state, rather than the same attitude towards a more complex content? On the *believe<sub>K</sub>-purist* view, Jana doesn’t believe in *any* sense that the gun is out of bullets, though she does *believe<sub>K</sub>* something like this: that the reasons she has, despite being insufficient for knowledge that the gun is out of

bullets, are nonetheless good enough to act on in the circumstances.<sup>18</sup> Why think this involves a new kind of belief?

I reply as follows: Explaining why someone believes or acts as they do by appealing to what they know is crucially different from explaining why someone believes or acts as they do by appealing to what, from their point of view, may or likely be true. In the paradigm case of acting from knowledge, we take the source of our knowledge to open up the world to us. The world is in view. But the less certainty we ascribe to an agent, the less apt it is to describe them as acting “because p” in the relevant sense; instead we are moved to say that they are acting on the basis of a calculation or hypothesis or ‘belief’ about p. Their reasons do not get them all the way to an obstructed cognitive connection to the fact on which they act.

Imagine this: I know that I locked the door, having checked it three times. But a doubt insinuates itself into my consciousness. I go downstairs to quadruple-check. I consider it highly likely to be locked, but I’m no longer certain. The technical term having been given the meaning it has, we could explain my behavior this way: I went downstairs because I merely believed<sub>w</sub> that it was locked. I certainly did *not* go downstairs because I believed<sub>K</sub>- it was locked. If I had believed<sub>K</sub>- it was locked, there would have been no felt need to check. The difference between explaining an action by appeal to believe<sub>w</sub> and explaining an action by believe<sub>K</sub>- is the difference between what can and what can’t make sense of my trip back downstairs.

The believe<sub>K</sub>- purist will point out here that we don’t need a technical term to understand this difference. We already have words and phrases that serve the same

---

<sup>18</sup> Thanks to a referee for raising this line of objection.

purpose: suspicion, educated guess, hunch, informed speculation, etc. Everything said above about the way in which explaining someone's action by citing  $\text{believe}_W$  is defective in comparison to the paradigm case can equally be said about the states picked out by the just-listed phrases. But so what? As defined,  $\text{belief}_W$  designates a heterogeneous class of mental states in their action-guiding role. They have in common that all fall short of a representation of what is; they represent what may be as the ground of action. As such, it's a useful category for philosophical analysis, elucidating the paradigm case (in which we act on certain knowledge) through contrast. But the category is not merely a philosophical artifice. When people say things like "I checked the attack because, although I wasn't sure, I believed the kids were hiding there", they reveal that they were acting on either an educated guess, hunch, speculation, suspicion, etc. This is a genuinely distinctive use of the ordinary term 'believe', one that indicates uncertainty in the provenance of the action.

The  $\text{belief}_K$ -purist will point out here that, by the definition above,  $\text{believe}_W$  is reducible to  $\text{believe}_K$ . Recall:

S  $\text{believes}_W$  iff S (i) grounds their thought and action on p despite (ii)  $\text{believing}_K$ -that she doesn't know that p

They might conclude from this reducibility that any explanation that can be given in terms of  $\text{belief}_W$  can, without loss of anything important, be given in terms of  $\text{belief}_K$ . This inference does not strike me as self-evidently valid. Be that as it may, a subject's  $\text{believing}_W$  describes a qualitatively different relation to that fact or would-be fact on which one acts than does  $\text{believe}_K$  simpliciter. So although I agree with the  $\text{belief}_K$ -purist that  $\text{believe}_K$  is the fundamental sort of belief, I would insist that  $\text{believe}_W$



corresponds to a distinct use of the word ‘belief’, one that indicates not just uncertainty but an understood-by-the-agent-to-be-defective cognitive relation to the grounds of their action. And it is precisely the way in which the concept of knowledge itself figures in the possession of a weak belief that sparks my interest here. Weak belief is only possible for rational creatures. In the next section, I will extend this claim. Given what belief is *in humans*, belief of any sort entails the possession of the concept of knowledge. No such concept is required for an animal belief of any sort.

I began this section with a discussion of Grice’s razor, which admonishes us against multiplying meanings unnecessarily. I have argued the cognitive state of a subject who acts in light of (their false belief that) *p*, who does not doubt that *p* and so would not add the hedge, and that of someone who hedges, saying “I believe that *p*, but...” are distinct. The mere fact that we use the word ‘belief’ in both cases should carry no more weight than it does in the case of ‘health’. Polysemy is, after all, ubiquitous.

Note too that there is an analogous instance of polysemy in the practical realm, one that also has roots traceable to ordinary usage as well as a fairly well-entrenched philosophical one. The words ‘want’ and (especially) ‘desire’ are ordinarily used to denote a state that is connected to the anticipation of pleasure, but are also sometimes used to describe the possession of motivation of any sort.<sup>19</sup> We might even call the broader sense of ‘want’ the *action-minus* sense, as we often describe someone who has vainly acted on intention to *x* by *y*-ing as “*y*-ing because they wanted to *x*,” where the relevant sense of want is consistent with but does not entail that it was the prospect of pleasure in *x*-ing that supplied the motivation. Philosophers routinely now use ‘desire’

---

<sup>19</sup> See Marcus (forthcoming).

in the action-minus sense, fittingly and for the same reason, as the partner to the knowledge-minus sense of 'belief' in the label 'belief-desire' psychology.

It might be argued that before we can accept that we are really dealing with multiple senses of the same word, we need a story about how the two uses are connected. What is called health in the case of food, appetites, and attitudes are what conduces to the health of an individual organism. What is called health in ecosystems and organizations is an analogue to biological homeostasis. But its focal use is to describe the state of an individual organism. A story of the same structure can be harder to see in the case of 'belief'.

In fact, it is not hard to construct. However the demand for such a story is overstated, given how common it is for words to exhibit this kind of semantic drift. That what was at one point (let us suppose) a semantically simple expression has spawned wide-ranging referential tendrils is entirely unexceptional. But here, in any case, is a just-so story about the connection between the two senses of 'believe'. The association between people described as believing<sub>K</sub>- and being mistaken led to a use of 'belief' in which the speaker uses 'belief' to acknowledge the possibility of their own error. I use 'belief' in describing the mistaken views of others; when I worry that I am mistaken, I then turn it on myself. But in this new context, I use the term to *express* the doubt of the believer himself (me!), whereas when I describe another as believing<sub>K</sub>-, I am not attributing to them any doubt at all. And this creates a divergence in meaning. The original use expresses the doubt of the belief-attributor about whether the believer knows. The new use of the word is self-reflexive; it expresses the doubt of the believer herself as to whether she knows. But again: even if this particular story is wrong, it

would not undermine the main idea: words just do come to possess different senses as the momentum and logic of their uses in various settings carries them ever farther apart.

The now-common impulse to pragmatize words into semantic homogeneity blinds one to real differences between things. Not because sameness of sense entails sameness of reference, but because indulging the impulse makes us insensible in our theorizing to the differences between what we are saying about, e.g., a person in describing them using (what are in fact) expressions with distinct senses. The health of an attitude and the health of an animal are not the same thing, nor is it *really* only animals who are healthy. Were one to insist that ‘belief’ invariably meant belief<sub>w</sub>, it would follow that Davidson, Fodor, Stich, and others had either been inadvertently been arguing for the view that the central cognitive state in the explanation of action was compatible with viewing its object as highly unlikely to be true or, even worse, had inadvertently been using the word wrongly and therefore failed even to address the topic of belief.

My focus in this section has been HRS’s claim that belief is weak. I have argued that there is a use of ‘belief’ whose purpose is to express weakness and there is another sense whose purpose is to prescind on the question of epistemic success. The more general lesson of the foregoing—and an important one for our purposes—is this: when we profess an interest in ‘Y’s, the question of what Ys *are* for the purpose of the investigation remains underdetermined. So what about an investigation into belief?

## **II. Human and Animal Belief**

We are interested in what *belief* is, as a chemist might be interested in what a *liquid* is, or as a microbiologist might be interested in what a *virus* is. Before beginning any such investigation, one needs a fairly definite conception of the scope of the phenomenon. Sometimes there is no debate whatsoever about which things are Ys. In other cases, one might begin with some paradigmatic cases and wind up surprised at what investigation turns up: a few things that we thought were Ys are not, and a few things we would never have thought were Ys are. But to get going at all, one must take a stand, at least provisionally, on exactly what one's theory is supposed to explain.

Consider now the Transformative Theory of Rationality.<sup>20</sup> According to this doctrine, the distinctive rationality of human beings should be understood not simply as a new suite of (rational) cognitive abilities, but as a transformation of an old suite of (non-rational) cognitive abilities. The Rationalist, as I will call the advocate of the Transformative Theory, accepts that both humans and animals learn about the world on the basis of perception and act in light of what they've learned. Both think, but what thought comes to is different in the two cases. Human rationality transforms what thinking is. So, on this theory, 'belief' does not pick out the same state in the rational and the non-rational. To understand what we use the term to describe in humans, we must (at least temporarily) put aside the non-rational phenomenon.

The Rationalist cannot be refuted simply by noting that we use the same words to describe the minds of humans and animals. After all, as we saw in section one, the word 'belief' picks out different things in different contexts. It is also not enough to point to empirical findings of significant commonalities in the cognitive processing carried about

---

<sup>20</sup>See Boyle (2011) and (2012), McDowell (1994).

by human and non-human brains. For these results cannot establish the identity of thought in humans and animals until we have already settled on an analysis of our use of 'belief' in the two contexts that renders them so much as eligible for identification via empirical investigation.

When we say that a dog is going to the door because he believes Jones is home from work and we say that a human is going to the door because he believes Jones is home from work, is 'belief' being said in the same way? According to the Rationalist, it is more like when we describe both a living rat and a rat doll as *having a tail*. Even if the rat doll has a tail that is stuffed entirely with rat meat, even if it is simply an actual rat's tail that has been glued to the doll, it would still not *have a tail* in the same sense as a rat has a tail, any more than an ordinary rat would have two tails were one glued to his back. Part of what it is to have a tail in the ordinary sense is for the relevant item to be the functioning organ of a living animal. Similarly, according to the Rationalist, part of what it is to have a belief, in the sense in which humans have a belief, is to be in a state that actualizes distinctively rational cognitive capacities. If the Transformative Theory is correct, no amount of similarity in the underlying cognitive processing is sufficient to render a state the sort of belief possessed by a human, any more than the internal similarity of a glued-on tail to that of a live rat can make the possession of the glued-on tail the same as the possession of a tail in the ordinary sense.

Of course, the difference between a rat's and a doll's tail is far more salient than the difference between human and dog belief. And so it is relatively easy (in comparison to the corresponding claim about 'belief') to stomach the idea that 'tail' in its application to living things and dolls is an instance of polysemy. To get beyond what has become a discipline-wide, reflexive dismissal of the very idea, we need to consider the

Transformative Theory's underlying rationale. Here's a feature that pretty much everyone now agrees characterizes belief in both humans and animals: it is governed by a standard. We squabble about how exactly to characterize it, but minimally it has this much to do with truth: false beliefs are bad qua beliefs. There is also widespread acceptance of the idea that part of what it is to hold a belief is for that belief to have practical ramifications. What a believer believes should influence their actions in a manner that befits the governing truth-standard. We can call this the truth-in-action standard. To believe that p yet fail to act as if p were true is to act deficiently. Again, there are loads of details to fill in and many of these details are the subject of philosophical dispute. But between these two standards, it might seem as if we have a core notion of belief and a core notion of how belief fits with behavior, one that is broad enough to match how we think of both humans and animals and yet narrow enough (at least with a bit more tinkering) to exclude other kinds of mental states. This might seem to justify identifying the beliefs of the rational and non-rational alike, at which point we can begin to discuss similarities and differences in how the common state is 'implemented', the mechanisms that realize it, and so forth.

But now here's a difference, one that is also widely accepted in the abstract. Unlike animals, we *understand* the standards: they are *norms*. We recognize that our beliefs are deficient insofar as they are false and that our actions are deficient insofar as they fail to reflect our beliefs. And the power of holding our beliefs (or candidate-beliefs) and actions (or candidate-actions) up for scrutiny in light of these norms reflects that understanding and manifests itself in the (often less than perfect) articulation of both the belief and the norms. Furthermore, insofar as we are capable of such scrutiny, we stand in a distinctive relation to our beliefs. We can determine what to believe by

exercising these sorts of rational capacities. This is unlike any relation that a non-rational creature stands in to their own beliefs since they (by hypothesis) do not have such capacities. And it is unlike any relation that any other person stands to my beliefs. Your assessment of my beliefs (or candidate beliefs) and actions (or candidate actions) has no tendency, all by itself, to determine that I hold a belief or perform an action. But my own assessment does.

How do we decide whether this is a difference in what beliefs *are* in humans and animals or whether it is just a difference in how the very same state is cognitively situated in the minds of the human and the animal? The Rationalist holds that the word ‘belief’ picks out different states in humans and animals. The two uses of the terms are unified and differentiated according to the sketch above. Given that this is not accidental homonymy, we might also say that, according to the Rationalist, human and animal belief are two ways of being a belief.<sup>21</sup> To say that a creature believes that p, in paradigmatic cases, is (more or less) to say that they would be incorrect if p were false and that they are bound to act as if p were true. The difference between human and animal senses of ‘belief’ corresponds to a difference in what correctness and boundness amount to, as well as to a corresponding difference in believers’ connection to the contents of their beliefs, and perhaps a difference in the contents themselves. Most centrally, the Rationalist holds that the truth-standard and the belief-in-action standard ‘govern’ the relevant state in different senses.

---

<sup>21</sup> Using terminology from Ford (2011), we might describe human and animal beliefs as *determinate* species of the same genus, i.e., species (such as lion and red) that are explanatorily prior to their genus (e.g., animal and color).

According to the Transformative Theory, human belief—i.e., what ‘belief’ refers to when typically applied to cognitively unimpaired adult humans—is an exercise of conceptual understanding, an exercise of the ability to grasp proposition. This ability is reflected in competence with the corresponding linguistic forms. Human belief, on this view, is also marked by an understanding of the act of judgment itself: knowledge that if  $p$  is true,  $\sim p$  must be false. I understand that insofar as I believe  $p$ , I am committed to accepting whatever follows from  $p$  and committed to rejecting anything incompatible with  $p$ . To believe *anything* in the sense in which human adults do depends upon the possession of the concepts of incompatibility and entailment. In believing that  $p$ , I already hold myself to a standard of truth, thus the applicability of the truth-norm is a function of my belief always already being my attempt (so to speak) to believe what’s true. The dynamics of rational psychology, according to the Rationalist, are a function of our own understanding of what makes for truths and falsehoods. But the governance of the animal belief—governance in virtue of which it is a belief—is not a function of the animal’s understanding of propositions, the norm of truth, the very idea of incompatibility, the very idea of entailment. For they do not have any grasp of such things.<sup>22</sup>

---

<sup>22</sup> It is worth emphasizing that the Transformative Theory itself is silent about who is and who isn’t rational. It just postulates two different sorts of cognitive states, one sort had by rational creatures and one sort had by non-rational creatures. The theory ultimately thus does not depend any empirical assumption about the (non) rationality of actual animals.



So how does section one's discussion of the two senses of 'belief' bear on the Transformational theory? I begin my answer with the following observation: the concept of  $\text{belief}_K$  arises in the context of the explanation of someone *else's* behavior (or of one's own behavior at a previous time), for reasons connected to Moore's Paradox. If I take myself to be acting *because p*, then I take myself to know that p. I am not neutral on the question of whether I know that p, and hence would not prescind from the question by saying "I am x-ing because I believe that p"—unless I was trying to mislead the audience about my state of certainty. One cannot bracket the truth of p while at the same time taking oneself to act in light of the fact that p. The concept of  $\text{belief}_K$  arises in virtue of a perceived discrepancy between what an agent thinks they know and what a (distinct) explainer of the action thinks they know. Thinking you don't know that p, I say that you are x-ing because you (merely)  $\text{belief}_K$  that p.

The concept of  $\text{belief}_W$ , on the other hand, arises in the first instance in the context of self-attribution. A person puts forward something as true, but with a qualification: "or so I believe", or something to that effect. This reflects consciousness that one falls short of an epistemic requirement, knowledge (or at least  $\text{belief}_K$ ) that one fails to know. In attributing  $\text{belief}_W$  to another, one is saying that they could only honestly assert that p if they at the same time acknowledge their own epistemic shortcomings.

The concept of  $\text{belief}_W$ , on this analysis, has its roots in doxastic self-consciousness. One is aware that one believes that p and also that one's believing does not measure up to the standards of knowledge. To believe weakly thus entails the exercise of the very concepts that confer on one the ability to assert. It is inextricable from the capacity for propositional articulation.

Is this analysis correct? It may seem vulnerable to an obvious and fatal problem: that we also attribute beliefs<sub>w</sub> to non-rational creatures, e.g.: “Fido believes his lost bone is in the neighbor’s yard, but he’s not sure. If he were more confident, he would be willing to brave their rottweiler.” But this is not weak belief. If dogs were capable of weak belief, there would be a distinction between a reluctance to act that had its source in weak belief and a reluctance to act that had its source in fear. This is the lesson of Tweedledum and Tweedledee. But we can make no such distinction in the case of a dog. There are a variety of possible reasons why a dog might refrain from encroaching on its neighbor’s territory, but they do not include consciousness of the weakness of the evidence favoring the presence of a buried bone. We can explain the lack of confidence of a rational creature by appealing to belief<sub>w</sub>. No such explanation for lack of confidence is available in the case of the non-rational. They are *not* believers<sub>w</sub>.

If weak belief includes propositional doubt, which in turn depends on the doxastic self-scrutiny that language enables, what would follow about belief<sub>k</sub>? And what are the implications for the scope question?

Consider that weak belief is a *weakening* of something else: the holding p true to which one does *not* append doubt about p. What must belief<sub>k</sub> be like if it is weakenable? Specifically, given that weak belief is, as just argued, articulable, could *unweakened* belief fail to be articulable? Intuitively, it’s hard to see how. Belief<sub>w</sub> is expressed this way: ‘p is true...but might not be’. Belief<sub>k</sub> is expressed this way: ‘p is true’. The latter is equally articulable and articulable without qualification. This holding true, insofar as it is not weak, does *not* as such put one in mind to say “or so I believe”. And it is judged according to epistemic standards. It is belief<sub>k</sub>. One is already

exercising the capacity to believe<sub>K</sub>- whenever one believes<sub>W</sub>, just as the capacity to ride a bike (on two wheels) is exercised whenever one does a wheelie.

Then again, some might say that belief<sub>K</sub>- should be understood, at least in its so-called dispositional mode,<sup>23</sup> as a *potential* for articulation. However, there is only so little actuality such a potentiality can have. To doubt what one believes is to doubt *one's own* view. And 'view', in the relevant sense, is the person-level endorsement of what is equally suited to be the object of person-level doubt. Furthermore, as Robert Audi emphasizes, whatever a dispositional belief is, it cannot be a mere disposition to believe, i.e., the propensity to acquire something to which the norm of truth applies, but must be something to which the norm applies directly.<sup>24</sup> A belief<sub>W</sub>, even one that is not *on* one's mind, must be conceptually demanding if it is to be someone's back-burnered sense that although they are willing to act on p and put it forward in a qualified manner, they do not know it. And insofar as the difference between belief<sub>K</sub>- and belief<sub>W</sub> is just the presence of doubt, we should say that a belief<sub>K</sub>- on the mind's back burner is no less an exercise of the *person's* cognitive capacities than those that are on their minds. It must be *their* undoubted take on the world. Thus we should just say that the sort of believing that people do is *tout court* conceptually demanding.

'Belief' in the sense of belief<sub>W</sub> has no application to the non-rational. And while 'belief' in the sense of belief<sub>K</sub>- does have application to the non-rational, the state that it picks out in rational creatures is one that actualizes the ability to articulate. *Rationality itself is a dimension of polysemy*. It follows that an investigation into the nature of

---

<sup>23</sup> I cast doubt on the very idea of dispositional belief in Marcus (2021), ch. 1.

<sup>24</sup> Audi (1984).

human belief cannot exclude from the data a feature of belief simply because it is not shared by animals. The scope of a theory of *our* beliefs does not extend to those of the non-rational. Here, as in other areas of philosophy, we do better when we resist the temptation to highest-common-factor-ism.

This picture of the difference between rational and non-rational belief also explains why some—most prominently Ernest Sosa—are drawn to distinguish between animal knowledge and a more demanding sort of knowledge.<sup>25</sup> Animals don't themselves apply evidential norms in coming to believe or disbelieve. Often, *we* say whether *they* know according to whether their cognitive processes conform to epistemic standards. But an animal's beliefs do not include a grasp of such standards. We are not only believers, we are possessors of the concept of belief. And thus our own beliefs, unlike the beliefs of animals, reflect our application of epistemic standards to ourselves. We can put the point this way: Whereas the norms governing animal belief are *external* to their beliefs, the norms governing human beliefs are *internal* to our beliefs themselves. When we ascribe belief to a person, we attribute to them the cognitive wherewithal to bring that belief to consciousness and subject it to the scrutiny of the governing norm of truth. Human belief is itself an exercise of the very cognitive capacities that operate when such beliefs are made explicit and scrutinized. Our beliefs themselves (often) reflect having followed and not merely conformed to epistemic norms. What transforms belief is doxastic self-consciousness, in which belief (ideally) conforms to the norms of belief precisely through an understanding of those norms. Under these conditions, it is also a distinctively rational form of knowledge. An

---

<sup>25</sup> See, e.g., Sosa (1991).

epistemology that distinguishes between human and animal knowledge fits best with a philosophy of mind that makes a parallel distinction among beliefs.

### III. Conclusion

‘Belief’ is not said in the same way in connection to humans and animals. Or so I have argued. Why care? Because we have no clear idea of what it is that we are trying to illuminate until we understand the specific difference between what we are saying about humans and animals in using the language of ‘belief’. Human belief that *p* requires: a (propositional) grasp of *p*; an understanding that one must believe anything that follows from *p* and must not believe anything incompatible with *p* (so long as one doesn’t surrender the original belief); and an ability to avow the belief, i.e., to say knowledgeably, but not on the basis of evidence, that one holds the belief.

Furthermore, these requirements are not fully separable. It is only insofar as we grasp propositions and possess doxastic self-knowledge that we can be responsive to entailment and incompatibility among candidate beliefs. And our ability to avow our beliefs is itself a function of the fact that epistemic standards are (unlike for the non-rational) internal to our beliefs.<sup>26</sup> This emerges in their transparency: I can avow my beliefs simply by thinking about what’s true. Human belief (whether belief<sub>K</sub> or the belief<sub>W</sub>) is a representation of the world in which understanding, normative sensitivity, and self-knowledge are co-actualized. How precisely to understand this unity is the central task for a theory of human belief, but it doesn’t so much as show up as a

---

<sup>26</sup> See Marcus (2021) for a detailed account of belief along these lines.

desideratum until one has distinguished it from mere animal belief. I hope, then, to have supplied a corrective to the current trends in the philosophy of mind that make it impossible to understand our own beliefs and so ultimately ourselves.<sup>27</sup>

---

<sup>27</sup> Thanks to audiences at Utrecht and Southampton Universities, and especially to Ram Neta.

r

## Bibliography

Audi, Robert 1984. Dispositional Beliefs and Dispositions to Believe. *Noûs* 28 (4) 419-434.

Aristotle, 350 B.C.E. *Categories*.

Benton, Matthew A. 2011. Two More for the Knowledge Account of Assertion. *Analysis* 71: 684–687.

Benton, Matthew A., and Turri, John. 2014. Iffy Predictions and Proper Expectations. *Synthese* (191): 1857–1866.

Boyle, Mathew. 2011. ‘Making up Your Mind’ and the Activity of Reason. *Philosopher’s Imprint* (11): 1–24.

Boyle, Mathew. 2012. Essentially Rational Animals. In *Rethinking Epistemology*, edited by Guenther Abel and James Conant. Berlin, Germany: Walter de Gruyter, 395–428.

Carston, Robyn. 2021. Polysemy: Pragmatics and sense conventions. *Mind & Language* (36):108–133.

Davidson, Donald. 2001. *Essays on Actions and Events*. Oxford: Oxford University Press.

Davis, Wayne. 1998. *Implicature*. Cambridge: Cambridge University Press.

Dorst, Kevin. 2019. Lockeans Maximize Expected Accuracy. *Mind*, 128 (509): 175–211.

Fodor, Jerry A. 1979. Methodological solipsism considered as a research strategy in cognitive psychology. *Behavioral and Brain Sciences* 3 (1):63-73.

Ford, Anton. 2011. Action and generality. In Anton Ford, Jennifer Hornsby & Frederick Stoutland (eds.), *Essays on Anscombe's Intention*. Harvard University Press.

Frisson, S. & Pickering, M. 2001. Obtaining a figurative interpretation of a word: Support for underspecification. *Metaphor and Symbol* (16): 149–171.

Gendler, Tamar. 2008. “Alief and Belief.” *Journal of Philosophy* 105, 634-663.

Grice, H. P.. 1989. *Studies in the Way of Words*. Cambridge: Harvard University Press.

Hawthorne, John; Rothschild, Daniel and Spectre, Levi. 2016. Belief is weak. *Philosophical Studies* 173 (5):1393-1404.

Holguín, Ben (forthcoming). Thinking, Guessing, and Believing. *Philosophers' Imprint*:1-34.

Marcus, Eric. (forthcoming). “Knowing What You Want,” forthcoming in *Forms of Knowledge*, ed. Lucy Campbell. Oxford: Oxford University Press.



Marcus, Eric. 2021. *Belief, Inference, and The Self-Conscious Mind*. Oxford: Oxford University Press.

McDowell, John. 1994. *Mind and World*. Cambridge, MA.: Harvard University Press.

McGlynn, Aidan 2013. Believing Things Unknown. *Noûs* 47 (2):385-407.

Rothschild, Daniel. 2020. What It Takes to Believe. *Philosophical Studies*, 177 (5): 1–19.

Ruhl, Charles. 1989. *On monosemy: A study in linguistic semantics*. Albany, NY: SUNY Press.

Sosa, E. (1991). *Knowledge in perspective*. Cambridge: Cambridge University Press

Stalnaker Robert. 2006. On logics of knowledge and belief. *Philosophical Studies* 128(1):169–199.

Stich, Stephen P. (1978). Autonomous psychology and the belief/desire thesis. *The Monist* 61 (October):573-91.

van Elswyk, Peter. 2018. *In the Mood for Assertion*. Ph.D. thesis, Rutgers University, New Brunswick.

Whiting, Daniel (2013). Nothing but the Truth: On the Norms and Aims of Belief. In Timothy Chan (ed.), *The Aim of Belief*. Oxford University Press.

Williamson, Timothy. 2000. *Knowledge and its Limits*. Oxford University Press.