

# True belief about knowledge

Adam Michael Bricker

Adam.Michael.Bricker@gmail.com

## Abstract

Here I pose a challenge to realism about knowledge, the view that facts about knowledge are non-trivially mind-independent, adapting an evolutionary debunking argument from metaethics. In brief: Our beliefs about knowledge are the products of innate knowledge-representing capacities with a deep and well documented evolutionary history, and, crucially, this history indicates that such capacities are indifferent to whether there are any mind-independent facts about knowledge. Instead, knowledge-representing capacities are likely just a byproduct of processing limitations on primate cognition. This presents an explanatory challenge for the knowledge realist—How is it, then, that we nonetheless happen to have many true beliefs about knowledge? I'll argue that, without abandoning the non-naturalism that characterizes much of contemporary epistemology, realism struggles to provide a compelling answer. In contrast, *evolutionary anti-realism*, the view advocated for here, meets the explanatory challenge head-on. Facts about knowledge are grounded in facts about what our innate knowledge-representing capacities classify as knowledge, so it's no surprise that we have many true beliefs about knowledge. While none of this shows that realism about knowledge is false, these considerations give us *ceteris paribus* reason to prefer an anti-realist approach.

**Keywords:** Metaepistemology, theory of mind, naturalized epistemology, knowledge realism, knowledge anti-realism, evolutionary debunking

**Word count: 10,646**

## **1. Introduction: Facts about knowledge**

It's spring, and the snow is melting. I can watch as it recedes, almost in real time, in the heat of an unrelenting Nordic sun. I can see that the thermometer outside reads 10 °C in the shade. Sufficiently acquainted with the chemical sciences, I know that the snow is melting. It is a fact that I know that the snow is melting. Our world is populated with many such facts about instances of knowledge. It is a fact that I know that today is Friday; it is a fact that I know that my desk is covered in dust; and it is a fact that I know that our household is running low on pickles. There are also many facts about knowledge itself, which we infer from facts about instances of knowledge. It is a fact, for example, that knowledge is factive—S knowing that p entails p.

It's commonly thought that such *epistemic facts*—facts about knowledge, understanding, rationality, etc.—are non-trivially mind-independent.<sup>1</sup> Whether it's true that I

---

<sup>1</sup> At least two trivial senses of mind-dependence will be excluded from discussion here: (1) Facts about the epistemic status of S's belief that p will trivially depend on facts about S's belief that p. (2) Facts about the epistemic status of some of S's beliefs about minds and the mental, particularly factive attitudes, may be trivially mind-dependent in virtue of the dependence relation that factive attitudes stand in with reality. Going

know that the snow is melting is entirely independent of facts about human attitudes about knowledge, truth, rationality, and so on. This view, *metaepistemological realism*, is widely held by epistemologists—so widely held that it is uncommon to see it defended. As put by Carter and Sosa, the view is “plausibly the default presupposition of many first-order debates in mainstream epistemology, even though—as a metaepistemological position—it is not often explicitly argued for” (2022, §1.1). As we’ll use the term, *realism about knowledge*, the target of this paper, is one central component of metaepistemological realism: facts about knowledge are non-trivially mind-independent. *Anti-realism about knowledge* will then refer to the broad class of views that reject this realist default, maintaining either that facts about knowledge are mind-dependent or simply denying that there are facts about knowledge at all.

This paper canvasses support for a novel version of the anti-realist thesis, dubbed “evolutionary anti-realism,” which adopts a response-dependent framework alongside the empirical insight that our attitudes about knowledge have a deep evolutionary history codified in dedicated knowledge-representing systems.

EVOLUTIONARY ANTI-REALISM ABOUT KNOWLEDGE: Facts about whether S knows that p are grounded in facts about whether a properly functioning theory of mind system would under normal (or perhaps ideal) conditions classify S as knowing that p.

On this view, the fact that I know that the snow is melting is grounded in the fact that a properly functioning *theory of mind system*—the innate, evolution-guided neurocognitive system responsible for intuitive judgements about knowledge—would under normal (or perhaps ideal) conditions classify me as knowing that the snow is melting. And the fact that

---

forward, this paper will understand mind-independence and mind-dependence exclusively in the non-trivial sense.

knowledge is factive is grounded in the fact that theory of mind only classifies S as knowing that p if p is represented as true.<sup>2</sup> Accordingly, there are no mind-independent facts about knowledge, only those grounded in facts about human social cognition. Or, put another way, knowledge cuts across any mind-independent mental and epistemic categories. On this moderate form of anti-realism,<sup>3</sup> there may be mind-independent facts about (e.g.) beliefs, perceptions, and memories. And there may even be mind-independent facts about (e.g.) epistemic justification, rationality, and understanding. But facts about knowledge don't number among these mind-independent mental and epistemic facts. Crucially, however, this isn't to say that knowledge has no mental or epistemic components. There still may be (and almost certainly are) mind-independent constraints on knowledge. At an absolute minimum, knowledge requires belief (or something like it) and truth. But the fact that knowledge has these mind-independent constraints is itself mind-dependent. Similarly, this view in no way entails that there are any mind-dependent *constraints* on knowledge.<sup>4</sup> S can have knowledge without any agent, real or hypothetical, thinking about S or S's knowledge.

In what follows, I'll discuss this view in detail, working out an argument premise-by-premise over the course of the paper. But first, I want to note that evolutionary anti-realism is a more modest proposal than those we often find under the banner of *metaepistemological anti-realism*, which tend to rely on premises that many in epistemology will find

---

<sup>2</sup> Note that this framing in terms of some counterfactual, mind-dependent procedure parallels other versions of anti-realism in metaethics, particularly constructivism (e.g., Street 2010).

<sup>3</sup> Here I'll use *moderate* to indicate an anti-realist view about knowledge that can be largely decoupled from other anti-realist commitments.

<sup>4</sup> This assumes that the mind-dependence of knowledge itself—whether the existence of knowledge depends on some mental attitudes—is a separate question from the mind-dependence of facts about knowledge. This is a reasonable assumption to make (see Searle 1996), although it's certainly non-trivial (for an alternate take, see Khalidi 2015, 111).

unappealing. First, many such views are blunt philosophical instruments, with anti-realism about knowledge falling out as a downstream consequence of some broader anti-realist commitments, like metaethical anti-realism. Olson (2014) and Streumer (2017) advance a kind of epistemic error theory, on which normative reasons for belief do not exist, through a general error theory of all normative claims (Streumer) or by way of metaethical error theory (Olson). And one of the central arguments cited by Chrisman for epistemic expressivism, roughly the view that knowledge attributions are non-descriptive expressions of the satisfaction of epistemic norms, is that such a view is required for ethical expressivism (2012). Relatedly, we find something structurally analogous in the characterization of the constructivist framework critiqued in Boghossian's *Fear of Knowledge*, with "constructivism about knowledge" understood as constituted by upstream claims about truth, justification, and/or rational reflection being socially constructed (2005, 22-3). In each case, anti-realism about knowledge is but one node of a broader anti-realist network, restricting its appeal to only those comfortable with such commitments.

Additionally, there are also versions of anti-realism about knowledge motivated by the same anti-invariantist considerations cited by epistemic contextualists—the skeptical challenge to invariantism (e.g., Chrisman 2007) and putatively context-sensitive patterns of ordinary knowledge attribution (e.g., MacFarlane 2010). Here I won't venture to offer any critique of such viewpoints or arguments, but simply point out that both these underlying strategies are widely questioned in epistemology. There is significant disagreement regarding whether eschewing invariantism offers a viable anti-skeptical strategy (for an overview, see Rysiew 2021, §4.2); it has been suggested that apparent context-sensitivity is the product of unreliable cognitive heuristics and biases (Williamson 2005; Nagel 2010; Gerken 2017); and empirical findings call into question the extent to which ordinary knowledge attributions are context-sensitive in the first place (Gerken 2017; Rose et al. 2019). For all those wary of such

contextualism-style arguments, then, any anti-realist account motivated along these lines will be a non-starter.

By eschewing all this excess philosophical baggage, one attractive feature of evolutionary anti-realism is that mind-dependence for knowledge can be decoupled from any additional anti-realist assumptions. One can be a realist about both truth and non-knowledge epistemic facts while still being an anti-realist about knowledge. And while the argument that I'll present is adapted from metaethics, it neither assumes nor entails any form of metaethical anti-realism. Additionally, we might also avoid any of the contentious assumptions about skeptical problems or the context-sensitivity of knowledge attribution that characterize contextualism-adjacent versions of anti-realism about knowledge. In this way, evolutionary anti-realism is positioned to enjoy a wider appeal than those alternatives currently on the market.

Here's the plan for the rest of the paper—We'll start by outlining the argument, transposing an “explanatory challenge” to non-naturalist realism from metaethics to metaepistemology (§2). Next comes a premise-by-premise defense of the argument, in which we observe that our beliefs about knowledge are the products of evolutionary forces (§3), which are causally indifferent to whether there are mind-independent truths about knowledge (§4), so the non-naturalist realist struggles to explain how we have many true beliefs about knowledge (§5). We'll close with a quick word on further reasons to prefer evolutionary anti-realism, highlighting its capacity to underwrite the expansive use of empirical techniques in epistemology (§6).

## **2. Evolutionary debunking and the explanatory challenge**

This section adapts the structure of an evolutionary debunking argument from metaethics to one in support of evolutionary anti-realism about knowledge. As an initial disclaimer, keep in mind that evolutionary debunking arguments have been widely debated over the past two decades, and there's extensive disagreement not only regarding whether they're successful, but even how exactly they're supposed to work in the first place (see Joyce 2006; Street 2006; Enoch 2010; Kahane 2011; Clarke-Doane 2012; Vavova 2015; De Cruz 2016; Levy and Levy 2020). In what follows, I'll adopt just one of many possible options—an explanatory challenge interpretation (especially Enoch 2010) of a Street-style argument against realism about evaluative truths (Street 2006), explicitly restricted as a challenge to only non-naturalist realism. This argument will maintain, roughly, that given the evolutionary history of our beliefs about knowledge, an anti-realist approach can better explain how these beliefs are often true than can realism about knowledge. As we'll observe, such an argument is limited not only in that the metaepistemological naturalist, who takes facts about knowledge to be causal (e.g., Kornblith 2002), has the means to circumvent this challenge, but because, even as a challenge to non-naturalist realism, it's far from conclusive. At most, the argument offers tentative reason to prefer evolutionary anti-realism, all while leaving open the possibility that even this tentative reason might be undermined by some “third factor” explanation (see Enoch 2010; Skarsaune 2011; Schechter 2018). Accordingly, my aim here is not anything like definitively refuting realism about knowledge. Rather, in giving us *any* reason to favor a moderate anti-realist alternative, my hope is to initiate a broader reappraisal of the epistemologically common practice of uncritically assuming (non-naturalist) realism as something of an implicit default.

Let's start then with the basics. Street's 2006 argument is built on the central empirical claim that the contents of our evaluative judgements are to a significant degree influenced by evolutionary forces that are indifferent to any mind-independent evaluative

truths—our (non-epistemic<sup>5</sup>) beliefs about “what one should or ought to do, about what is good, valuable, or worthwhile, about what is morally right or wrong, and so on” (110) are largely driven by the evolutionary advantages of having cognitive mechanisms that generate such beliefs:

[T]endencies to make certain kinds of evaluative judgements rather than others contributed to our ancestors’ reproductive success not because they constituted perceptions of independent evaluative truths, but rather because they forged adaptive links between our ancestors’ circumstances and their responses to those circumstances, getting them to act, feel, and believe in ways that turned out to be reproductively advantageous. (127)

As I’ll understand the argument, these empirical considerations set up an explanatory challenge for the realist: Assuming that many of our evaluative beliefs happen to be true, how is it that evolution-guided belief-forming mechanisms, which are indifferent to any mind-independent evaluative truths, have happened to supply us with such true beliefs? (see especially Enoch 2010; also Skarsaune 2011; Schechter 2018). Enoch frames this challenge in terms of explaining the “correlation” between our normative judgements and normative truths (2010, §3)—if our normative beliefs correlate with normative truth, what explains this correlation? As noted by Schechter (2018, §6), analogous challenges appear throughout philosophy, wherever non-naturalist realisms can be found.<sup>6</sup> This is because such views have an inherent structural disadvantage in explaining fact–belief alignment: If our beliefs in some domain are neither caused by the facts (non-naturalism) nor constitutive of the facts (realism),

---

<sup>5</sup> See Street 2006, footnote 2.

<sup>6</sup> E.g., the Benacerraf-Field challenge to mathematical Platonism (Benacerraf 1973; Field 1989, 25), Schechter’s counterpart argument for logical truths (2010), and Nozick’s challenge re. metaphysical necessity (2001, ch. 3).



then there's no ready-made explanation for how those beliefs happen to be true. However, while this means that the challenge might be formulated in the absence of any phylogenetic considerations, the (putative) evolutionary origin of evaluative belief sharpens the challenge, boxing in the kind of explanation the realist is expected to provide. How is it that *evolution specifically* has supplied us with true beliefs about non-causal, mind-independent facts?<sup>7</sup>

In principle, the realist may be able to meet this challenge with a “third-factor” explanation, which accounts for both evaluative facts and our beliefs about these facts with a common, shared origin (Enoch 2010; Skarsaune 2011; Schechter 2018). We'll circle back to this in §5, finding that, in the case of knowledge, the most obvious third-factors come up short.

Perhaps more immediately relevant, one might also in principle sidestep the challenge entirely by taking up a naturalist stance. As Street herself notes (2006, 112 & 131), the argument works best against *non-naturalist* value realism, on which “evaluative facts or truths are not reducible to any kind of natural fact, and are not the kinds of things that play a role in causal explanations” (2006, 111). While Street does also attempt to extend the challenge to naturalist realism (2006, §7; 2008; for one response, see Copp 2008), here I'll refrain from doing so, restricting the critique to only non-naturalist realism about knowledge, on which facts about knowledge are non-causal. As we'll observe below (§5), if one grants that facts about knowledge play a causal role in the explanation of patterns of animal behavior (Kornblith 2002), other mental attitudes (Logins 2021), or perhaps even some individual behaviors in humans (Williamson 2000; Nagel 2013), then the challenge might be

---

<sup>7</sup> Notice that this explanatory challenge interpretation differs from versions of the argument that question the justificatory status or truth-tracking of our beliefs about knowledge (e.g., Kahane 2010; De Cruz 2016; see also Enoch 2010, §2; Clarke-Doane and Baras 2021). The claim here is not that, given realism, we have no justified/sensitive/safe evaluative beliefs.

straightforwardly addressed: Our beliefs about facts about knowledge are often true because they are often caused by facts about knowledge. Nevertheless, as many philosophers are reluctant to commit to facts about knowledge playing any role in causal explanation—Logins describes causal inefficacy as the “standard view” for knowledge (2021, 1338), and Kornblith’s naturalism has been roundly criticized on a number of fronts (e.g., Roth 2003; BonJour 2006; Millar 2007, §2; Pernu 2009; Hannon 2019, §1.5)—non-naturalist realism is plausibly viewed as the dominant strain of realism about knowledge amongst mainstream epistemologists. The explanatory challenge, then, might be understood as pressing non-naturalist realists into a hard choice between abandoning realism or adopting naturalism. Here I won’t attempt to adjudicate between these two options, but instead simply grant non-naturalism as a given.

With the parameters set, here then is the argument:

1. CAUSATION: The human capacity to represent others as having knowledge, which (a) has been largely guided by evolutionary processes, (b) is responsible for our beliefs about knowledge.
2. ADVANTAGE: There are significant evolutionary advantages of representing others as having knowledge, which are causally independent of any mind-independent facts about knowledge.
3. NON-ERROR THEORY: We have many true beliefs about knowledge (and comparatively few false beliefs).
4. EXPLANATION: If CAUSATION and ADVANTAGE, then evolutionary anti-realism better explains NON-ERROR THEORY than does non-naturalist realism.

From 1-4, it follows that evolutionary anti-realism offers a better explanation of how we have many true beliefs about knowledge than does non-naturalist realism.

The rest of the paper will be dedicated to a systematic defense of this argument, laying out the case for CAUSATION (§3), ADVANTAGE (§4), AND EXPLANATION (§5) in turn. However, before that, two quick, clarifying remarks: First, notice that here we're interested in "beliefs about knowledge" generally, which is taken to include both *intuitive* and *reflective* beliefs. Intuitive beliefs about instances of knowledge—immediate, unreflective judgements about whether S knows that p in some particular case—are understood as being directly generated by evolution-driven capacities. Reflective and theory-driven beliefs about knowledge itself—beliefs like *knowledge is fallible* (e.g., Cohen 1988) or *knowledge is subject to both modal and virtue-theoretic constraints* (e.g., Pritchard 2012)—are then understood as being largely guided by intuitive beliefs. Notice also that NON-ERROR THEORY is simply assumed. It's not the case that there are no facts about knowledge (a more radical anti-realism); nor is it the case that, while there are facts about knowledge, we have few-to-no or merely chance-level true beliefs about these facts (a kind of metaepistemological skepticism). Within the context of mainstream epistemology, I don't think that this is an especially controversial assumption to make, and adopting either radical anti-realism or metaepistemological skepticism just to avoid moderate anti-realism strikes me as a real baby-with-the-bathwater proposition.

### **3. CAUSATION**

CAUSATION is an empirical premise composed of two claims about the human capacity to represent the knowledge of others: (a) the development of this capacity was largely guided by evolutionary processes, and (b) this capacity is primarily responsible for our beliefs about

knowledge, including both unreflective intuitive judgements as well as reflective and theory-driven judgements. This section will defend both these claims in turn, arguing that the empirical evidence in this particular case is remarkably strong. Comparative psychologists working on theory of mind have put considerable effort into understanding the evolutionary history of knowledge-representing capacities (see Krupenye and Call 2019; Krupenye 2021), and it's widely recognized that these capacities are directly responsible for our intuitive beliefs about knowledge (e.g., Nagel 2012; Bricker 2020; Phillips et al. 2021a). This quiets any worries, as have been raised for the metaethical version of evolutionary debunking (e.g., Levy and Levy 2020), that the causal premise may be unsupported by the available empirical evidence. Moreover, as our reflective judgements about knowledge are largely guided by intuitive attitudes, we find that both our reflective and unreflective beliefs about knowledge are ultimately the products of evolution-driven knowledge-representing capacities.

Let's start with the basics. Humans have a deep, innate capacity to represent the mental states of others—their beliefs, desires, perceptions, confidence, etc.—referred to collectively as “theory of mind,” “mentalizing,” or “mindreading” (for discussions, see Apperly 2011; Mahy et al. 2014; Phillips and Norby 2021). Theory of mind is a staple of social cognition, complete with specialized neural mechanisms (e.g., van der Meer et al. 2011; Hartwright et al. 2015) and a hallmark developmental trajectory starting in early childhood (see, e.g., Fabricius et al. 2010). And, crucially, theory of mind is responsible for our capacity to represent others as having *knowledge*. To some, particularly in epistemology, it might be surprising to hear that the knowledge of others is represented by the same system that supports representations of others' mental states. After all, many epistemologists doubt that knowledge is a *sui generis* mental state, instead taking it to be a kind of metaphysical composite state, composed of both mental (belief) and nonmental components (truth + epistemic constraints). Nevertheless, empirical findings consistently show not only that adult

humans deploy theory of mind in computing knowledge representations, but that these representations are computed as if knowledge is its own kind of mental state, not a mental-nonmental hybrid (Phillips et al. 2018; Bricker 2020). This view is bolstered by developmental findings demonstrating that the capacity to attribute knowledge is a primary milestone in the development of theory of mind, which precedes the capacity to represent belief (for discussion, see Phillips et al. 2021a, §4.2). Some have taken these kinds of findings as evidence that knowledge is in fact a mental state in its own right (Nagel 2013; Bricker 2021). However, the more epistemologically conservative view is that the ontology of theory of mind doesn't match the metaphysical structure of knowledge, even if it supplies us with the correct verdict for individual cases (for discussions along these lines, see Ichikawa and Jenkins 2017; Gerken 2021). On either interpretation, however, the empirical evidence is clear—The human capacity to represent others as knowing is a central component of our shared theory of mind capacities.

The evolutionary history of such capacities is a major research program in comparative psychology (for reviews, see Krupenye and Call 2019; Krupenye 2021; Phillips et al. 2021a, §4.1). The modern use of “theory of mind” to refer to the representations of others' mental states even has its origins in the study of non-human primate cognition, and, from the beginning, knowledge was taken to number among the paradigm states represented by theory of mind (Premack and Woodruff 1978). If we turn to this research, a number of points are immediately clear—Not only do our theory of mind capacities have long evolutionary histories, but knowledge representation is a central and early milestone in the evolutionary history of theory of mind. A summary timeline from Krupenye, for example, puts knowledge representation as emerging roughly 30-40 million years ago, with belief representation coming only 16-20 million years ago (2020, 109). That the capacity for knowledge representation developed early, before that of belief representation, is indicated by

a key set of evidence—non-human primates whose lineage diverged from that of humans in the relatively distant past, like rhesus macaques and chimpanzees, display an ability to represent the knowledge of others that significantly outstrips their ability to represent others' beliefs (e.g., Kaminski et al. 2008; Horschler et al. 2019). Fascinatingly, this holds not only for false beliefs, but Gettier-style true beliefs that fall short of knowledge.<sup>8</sup> All told then, we have little reason to doubt that our innate capacity to represent others as knowing, a core component of theory of mind, is largely the product of evolutionary processes.

It's no great leap to think that our everyday, intuitive judgements about cases of knowledge are underwritten by this core representational capacity—Whether we intuitively believe that S knows that p is primarily a matter of whether our theory of mind systems represent S as knowing that p. Nagel has suggested precisely this in her influential defense of the method of cases in epistemology (2012; see also Gerken 2017, chapter 5), and we even have direct neural evidence that laypeople rely on core theory of mind systems when forming explicit judgements about what others know (Bricker 2020, §4.3). However, perhaps the most compelling demonstration of the connection between theory of mind and the intuitive judgements of interest to epistemology comes from just how closely their epistemic profiles match. This was recently discussed at length by Jonathan Phillips and company, who note that at least four epistemologically central features of knowledge are identifiable in the knowledge representations supported by theory of mind (2021a, §2): (i) knowledge is factive, (ii) knowledge is different than true belief, (iii) S knowing whether p is consistent with

---

<sup>8</sup> For example, the Kaminski et al. study (2008) observed that Chimpanzees were significantly more successful at predicting an agent's actions on the basis of simple knowledge from memory (i.e., the agent saw that an object was placed under a box, where it remained) than a true belief that was made true as a matter of luck (i.e., while the agent wasn't looking the object was moved from under the box, but then returned to under that same box as a matter of luck).

“egocentric ignorance,” on which the knowledge evaluator doesn’t know whether p, and (iv) knowledge is modality-general, the same kind of state regardless of the basis on which it’s formed. Crucially, these features are visible even in the knowledge representations of non-human primates and young human children (Phillips et al. 2021a, §4.1.1 & §4.2.3): The knowledge representations of non-human primates, for instance, are factive, distinct from true belief representations, compatible with egocentric ignorance, and not tied to any particular modality. The natural explanation for all this is that the innate knowledge representations supplied by theory of mind (factive, not true belief, compatible with egocentric ignorance, modality-general) are largely responsible for our unreflective beliefs about knowledge (factive, not true belief, compatible with egocentric ignorance, modality-general). In short, the intuitive beliefs about knowledge central to epistemology are in all likelihood directly continuous with the evolutionary history of theory of mind.

This then leaves the question of reflective beliefs about knowledge, which might take the form of either beliefs about knowledge as such (e.g., *knowledge is fallible*) or beliefs about instances of knowledge driven not by intuition, but some theoretical concerns. Here I’d suggest that the process of reflecting on knowledge is—and probably must be—largely driven by and built upon the content of our intuitive beliefs. First, as Street points out in her argument, rational reflection isn’t an independent check on evaluative belief, but rather an operation dependent upon it (2006, 124; original emphasis):

[R]eflection must always proceed from some evaluative standpoint; it must work from some evaluative premises; it must treat some evaluative judgements as fixed, if only for the time being, as the assessment of other evaluative judgements is undertaken. In rational reflection, one does not stand completely apart from one’s starting fund of evaluative judgements: rather, one *uses* them, reasons in terms of them, holds some of them up for examination in light of others.

For rational reflection about knowledge, our starting fund of beliefs is none other than those intuitive beliefs supplied by theory of mind. Beyond these intuitive beliefs, and reflective beliefs that trace back to intuition, it's unclear how we might theorize about knowledge at all.<sup>9</sup> Moreover, we might observe that the reliance on intuitive judgements in the formation of reflective beliefs is especially pronounced in the case of epistemology. Entire theories of knowledge are proposed and defended almost exclusively on the basis of our intuitive beliefs (Pritchard 2012 is a particularly transparent, representative example; and for a more general overview, see Ichikawa and Steup 2018), and core theoretical constraints used in reflective argument are often straightforwardly derived from our intuitive beliefs. A fallibilist about knowledge, for example, might defend fallibilism on anti-skeptical grounds (see Cohen 1988), with the central premise of this argument, anti-skepticism about knowledge of the external world, deriving from our intuition-grounded belief that we all have lots of knowledge about the world around us. There are admittedly rare cases in which epistemologists might reflectively discount intuitive judgements about knowledge outright, but this usually only happens for "shifty patterns" that don't fit in with the vast majority of our intuitive beliefs about knowledge (Gerken 2017, chapter 2). That is, our main body of intuitive belief, assumed to be correct, serves as the basis for rejecting a recalcitrant subset of intuitions. Here, as with epistemology more generally, it's clear that our innate, intuitive beliefs about knowledge serve as the foundation for reflection and theorizing. By all accounts, both our intuitive and reflective beliefs about knowledge trace back to the long evolutionary history of our capacity to represent others as knowing.

---

<sup>9</sup> While the naturalist can maintain that psychology itself may supply a means of investigating and reflecting upon knowledge that could sever the ties to intuitive judgement and theory of mind, such an option isn't available on the non-naturalism at issue here.



#### 4. ADVANTAGE

ADVANTAGE, like CAUSATION, is an empirical claim about knowledge representation—There are significant evolutionary advantages to representing others as having knowledge, which are independent of any mind-independent facts about knowledge. In order to support this premise, we'll first walk through a couple of the most important roles of knowledge representation, action prediction and facilitating learning from others, after which we'll discuss why the cognitive simplicity of knowledge representation makes it better suited to these roles than the available alternatives. The crucial observation here is that all this is explanatorily decoupled from any truths, in the realist sense, about knowledge, or even whether it's a mind-independent fact that knowledge exists at all. Imagining that there are no mind-independent facts about knowledge for our knowledge-representing capacities to capture, the same needs would exist, and the neurocognitive profile of knowledge representation would still make it a prime candidate for fulfilling those needs.

Likely the most basic and primordial function of knowledge representation is the facilitation of predictions about the actions of others. A chimpanzee, for example, might predict whether a competitor for food will or will not beat them to a meal on the basis of whether this competitor is represented as knowing where the food is (e.g., Kaminski et al. 2008). On this kind of simple predictive framework, deployed by both non-human primates and young human children, “knowing leads to acting correctly,” e.g., going to and eating the food, whereas “not knowing leads to acting incorrectly,” e.g., leaving a perfectly good meal up for grabs (Fabricius et al. 2010, 1402). This role in action prediction is central to both the general characterization of theory of mind capacities in comparative psychology and how we trace the course of its evolutionary development (see Krupenye 2021). Moreover, even for adult humans, predicting action remains an important role of knowledge representation.

Colloquially, we frequently articulate what we anticipate that others will do in terms of what we take them to know, and experimental evidence indicates that knowledge representation, not belief representation, may be primarily responsible for this predictive function in human adults (Turri 2017). Put together, there is little doubt that knowledge representation plays a key role in the prediction of others' actions.

A second important function of knowledge representation is the facilitation of learning about the world from others. For example, if S\* represents S as knowing where some food is, this makes S a potential source of information about the location of the food, even if S\* is unaware of this information. This use of knowledge representation in learning is already observed in non-human primates (see Phillips et al. 2021a, §6.2). For example, a study from Krachun et al. (2009) observed that chimpanzees and bonobos were able to locate food based on the behaviors of those that they represented as knowing where the food was, even when they otherwise had no indication of the food's location. However, the learning function of knowledge representation is especially pronounced in human social interaction. During the course of everyday life, we frequently seek others out to learn about topics that they know about but we do not. If I want to know about the special features that distinguish Cuban Spanish from other varieties of the Spanish language, for example, by far the easiest thing to do is seek out someone who knows about Cuban Spanish—either through learning resources produced by an expert or simply asking a Cuban person. This informant-identifying role of knowledge is widely recognized in epistemology, notably occupying a central place in the influential account of knowledge developed by Craig,<sup>10</sup> who goes so far as to suggest that this is why we have the concept of knowledge in the first place (1990). While Craig's

---

<sup>10</sup> The popularity of Craig's approach extends from the old-school analysis of knowledge (e.g., Pritchard 2012) to the empirical study of knowledge representation (e.g., Phillips et al. 2021a), and only seems to be growing in recent years. Hannon (2019), for instance, offers a sustained defense and update of the program.

genealogical approach is interested in how the concept of knowledge would emerge in the imaginary state of nature, not the actual evolutionary history of knowledge representation, there are strong parallels to the view I'm advocating for here. Importantly, in both cases, one of the central roles of knowledge representation—perhaps even *the* central role—is flagging others as good informants.

Notice that the value of the action-predicting and informant-tagging functions of knowledge representation are flatly unrelated to any mind-independent facts about knowledge. Regardless of what, if anything, does and does not constitute knowledge, it is uncontroversially advantageous for humans and their evolutionary ancestors to be able to predict the actions of others and effectively pick out good informants. The crucial question is whether knowledge representation's suitability to fulfilling these roles depends on any mind-independent facts about knowledge. Looking now to why knowledge representation plays these roles in the first place, we quickly find that this can be explained without reference to any such mind-independent knowledge facts. Knowledge representation plays these crucial roles because knowledge representations are cognitively easy to compute. They are ultimately the product of the *processing limitations of primate cognition*. Assuming that there were no mind-independent facts about knowledge at all, neither these processing limitations nor the benefits of action prediction and informant flagging would change. To illustrate this, we'll compare knowledge representation to both more complex, non-factive representations, and simpler, modality-specific factive representations.

In principle, there is no reason that non-factive, belief-like representations cannot fulfil the same basic roles as knowledge representations. In fact, especially in adult humans, they often do. Representations of others' beliefs are particularly valuable in action prediction (see, e.g., Baker et al. 2009; Southgate and Verneti 2014), being useful in a wider range of cases than knowledge (e.g., false belief). And we might also expect that belief representations

could in principle be used in flagging good informants, perhaps (e.g.) supporting reasoning about *why* the believer believes what she does. It's not that the capacity to represent non-factive states like belief emerges relatively late in evolutionary development because such representations cannot reliably support the same critical functions as knowledge representation; they emerge later because earlier primate cognition cannot reliably support them. A defining feature of non-factive, belief-like representation is that it demands significantly more cognitive and neural resources than factive, knowledge-like representation. As summarized by Westra and Nagel, non-factive theory of mind is “cognitively effortful in ways that make it an unlikely basis for everyday social interaction,” whereas factive theory of mind is “fast and efficient” (2021, 2), allowing it to serve as something like the default mode of mindreading (see also Gordon 2021 for a complimentary view). This is largely down to the fact that representing non-factive mental states requires the management of independent perspective information, which is often inconsistent with one's own perspective (see Leslie 1987). When I represent S as believing that the earth is flat, for example, I hold separate representations of the earth as being round—my own primary representation of reality—and a representation of S's representation of the earth being flat.

Keeping these two distinct representations decoupled, and managing conflicts between them, requires a good deal of cognitive effort—so much so that even adult humans can struggle with it (Birch and Bloom 2007). This cognitive effort is associated with an increased demand for neural resources (Bricker 2020) and is largely supported by identifiable areas in the prefrontal cortex (van der Meer et al. 2011; Hartwright et al. 2015), which may even be specialized beyond more general inhibitory control (Hartwright et al. 2015). It's easy to understand, then, why knowledge representation appears earlier in our evolutionary history than belief representation. Belief representation has neural and cognitive requirements that our more distant evolutionary ancestors were likely ill-equipped to handle (Povinelli and

Preuss 1995; see also Smaers et al. 2017). Knowledge representation, in contrast, doesn't require any management of decoupled representations or inhibition of the self-perspective (Bricker 2020; Phillips et al. 2021b; Westra and Nagel 2021; see also Gordon 2021).

Accordingly, for primates with relatively limited processing capacity, it is an ideal candidate for supporting action prediction and learning from others.

Moving on, there is also no reason in principle why the core functions of knowledge representation couldn't be supported by modality-specific factive representations—e.g., seeing that *p*, hearing that *p*, and remembering that *p*. The problem with such a strategy, however, is quite straightforward. Each of these representations on their own is limited in scope. It's far more useful to have them all, not just one. But it would be unwieldy and inefficient to have to work through a series of modality-specific representations, first computing whether *S* sees that *p*, then whether *S* hears that *p*, then whether *S* remembers that *p*, and only getting to action prediction/informant flagging when a sufficient number of these representational options have been serially exhausted. Evidence from other epistemic judgements confirms, entirely unsurprisingly, that having to string multiple evaluations together into a composite judgement comes at a steep cognitive cost (Bricker 2020). The more efficient option is quite clearly a single kind of representation, which is widely useful regardless of modality.

In short, knowledge representation has an optimal cognitive profile for the early evolutionary history of theory of mind—more powerful than single factive representations tied to particular modalities and more efficient than a disjunction of many modality-specific representations, but without the cognitive and neural demands that put non-factive representations out of reach. Notice, moreover, that this cognitive profile helps to explain not only the evolutionary history of knowledge representation, but its continued dominance in the mentalizing of adult humans. Far from some vestigial capacity, we continue to rely heavily

on knowledge representation in daily life because the efficiency-power tradeoff frequently favors factive representation over non-factive representation (see especially Westra and Nagel 2021). While humans are certainly capable of the inhibitory control and perspective management required of belief representation, this demands costly neural resources (Bricker 2020), and we can still struggle to avoid interference between perspectives (Birch and Bloom 2007). All told, the modality-general, factive format of knowledge representation makes it particularly well suited for supporting action prediction and flagging good informants, with processing requirements attuned to the cognitive and neural limitations of both human and non-human primates.

The final observation to make here is that mind-independent facts about knowledge—as opposed to facts about mere (reliable, true) belief (see §5)—are explanatorily superfluous in this story. While the structure of knowledge representation tells us plenty about both the social needs and cognitive limitations of primates, it’s causally independent of whether there are, in the realist sense, any facts about knowledge at all. It could be, for instance, that knowledge doesn’t actually play a causal role in action, as non-naturalist epistemologists think (discussed above), with non-factive and modality-specific representations better reflecting actual mental ontology. Or it could be that what makes S a good informant has only to do with more fine-grained epistemic properties of her beliefs—perhaps rationality, basing on evidence, or reliability. In a word, it could be that the category “knowledge,” with its deep origins in the primordial theory of mind, cuts across all mind-independent distinctions in both the mental and epistemic domains, with the only facts about knowledge being those grounded in facts about our representations of knowledge. All the same, human and non-human primates would find the same advantages in developing action-predicting and informant-flagging capacities, capacities that would need to be supported by the same neurocognitive systems subject to the same processing limitations. Independent of any putative mind-

independent facts about knowledge, knowledge-representing capacities are evolutionarily advantageous.

## 5. EXPLANATION

The final premise of the argument posits that, if CAUSATION and ADVANTAGE, then evolutionary anti-realism better explains true belief about knowledge (NON-ERROR THEORY) than does non-naturalist realism. In support of this claim, let's start with the explanation offered by evolutionary anti-realism: If facts about knowledge are grounded in facts about the theory of mind systems that support knowledge representation, systems which are largely responsible for our beliefs about knowledge, then it's entirely unsurprising that at least many of those beliefs should be true (and comparatively few should be false). All we have to grant is that, at least much of the time, our theory of mind systems are indeed functioning properly. Consider S's belief that I know that the snow is melting. The fact that I have this knowledge is grounded in the fact that a properly functioning theory of mind system would under appropriate conditions represent me as knowing in this case. So, if anyone then represents me as knowing this through employing properly functioning theory of mind capacities—assuming an everyday case in which ordinary, mindreading-conducive conditions hold—the resulting belief will be true. In this way, evolutionary anti-realism provides an entirely straightforward account of how many of our beliefs about knowledge are true. Whenever our theory of mind systems function properly in ordinary, success-conducive conditions, our beliefs are produced by systems that recapitulate the facts that ground the truth of those beliefs.

Contrast this with the kinds of explanations available on knowledge realism. As mentioned above, a naturalist could seek to circumvent EXPLANATION by maintaining that facts about knowledge play some causal role in our beliefs about knowledge. Roughly, facts about knowledge are natural facts out there in mind-independent reality—something like facts about water or childbirth or snakes—and there is a survival benefit to tracking these facts. Thus, we’ve evolved the capacity to represent knowledge in virtue of the advantages of tracking knowledge. As this stands at odds with ADVANTAGE, the question then becomes whether the empirical evidence is more in line with the naturalist or evolutionary anti-realist story of knowledge representation. All this, however, can be set aside when considering the primary target of this paper, *non-naturalist* realism about knowledge, which rejects the suggestion that facts about knowledge play meaningful roles in causal explanation. Instead, in what is arguably the default view in epistemology, only facts about *belief* (rather than facts about both belief and knowledge) are taken to play roles in the causal explanation of individual human action (*contra* Williamson 2000) and group animal behavior (*contra* Kornblith 2002).

As discussed in section 2, assuming this kind of non-naturalism puts the realist in a bind. Not only does non-naturalist realism stand at a general structural disadvantage in explaining widespread true beliefs about putatively non-causal, constitutively mind-independent facts, but the empirical details of belief about knowledge only exacerbate this disadvantage, constraining the kind of explanation the metaepistemological non-naturalist realist is expected to provide. In brief, if our beliefs about knowledge are largely the product of evolutionary forces (per CAUSATION) and these forces are explanatorily independent of any putative mind-independent facts about knowledge (per ADVANTAGE), then (assuming NON-ERROR THEORY) how is it that *evolution* has happened to nonetheless guide us to many mind-independently true beliefs about knowledge? Therein lies the core of the



explanatory challenge. While here I won't claim that such a challenge cannot in principle be met, even the most promising strategy we'll identify faces limitations that the anti-realist explanation does not.

How then might the non-naturalist realist seek to meet the explanatory challenge? If we return to metaethics, we find one particular strategy is repeatedly cited as especially promising—the “third-factor” explanation (Enoch 2010; Sarksaune 2011; Schechter 2018). Roughly, there is some third factor that explains both moral facts and moral belief. For instance, on Enoch's sketch of third-factor explanations, the third factor is simply *survival itself* (2010, §5.3): Survival is typically good (presumably), and evolutionary forces select for survival, in part through the belief that survival and survival-promoting practices are good. Accordingly, we might thereby explain how we have at least some true beliefs about what is good without resulting to either anti-realism or naturalism. Adapting this strategy to the case of knowledge, the challenge is then to specify some third factor TF such that both:

(i) TF typically<sup>11</sup> supplies knowledge.

and

(ii) Evolutionary forces have resulted in the belief that TF is knowledge.

---

<sup>11</sup> Here I'll understand *typically* to mean that it is unusual—at least in ordinary, everyday life—to encounter an instance of TF that is not also an instance of knowledge. This stipulation is required to ensure that NON-ERROR THEORY is satisfied. If many ordinary cases of TF do not constitute knowledge, then this risks the possibility that the belief that TF is knowledge could generate too many false beliefs about knowledge in proportion to true beliefs for NON-ERROR THEORY to hold.

Sources of knowledge are perhaps the best candidates for this kind of third factor. Memory, introspection, reasoning, testimony, and—especially—perception are all plausibly viewed as typically supplying knowledge; and it’s difficult to deny that evolutionary forces have resulted in beliefs that instances of memory, introspection, reasoning, testimony, and—again, especially—perception are knowledge (see especially Phillips et al. 2021a). One might quibble over the extent to which some of these sources really are typically knowledge (perhaps especially memory, testimony, and/or reasoning), as well as whether this allows us to explain a sufficient number of belief–fact correlations (e.g., all those cases where the sources don’t provide knowledge). However, here I’ll set these concerns aside to grant that sources of knowledge provide a solid path towards a metaethics-style third-factor explanation. Instead, I want to take issue with the broader strategy here.

Notice that, at its core, this approach seeks to explain many instances of belief–fact alignment by positing new correlations, which are themselves striking and unexplained: it just so happens that perception is typically knowledge, *and* we independently believe that perception is knowledge; it just so happens that memories are typically knowledge, *and* we independently believe that memories are typically knowledge; etc. We haven’t eliminated unexplained correlations as much as push them up a level of abstraction. To be fair, Enoch recognizes this problem, maintaining that “explanatory progress” has nonetheless been made, even if we do so via a suite of other unexplained correlations (2010, 434). I’ll leave it to the reader to decide how convincing this is, as, regardless of whether this is a good explanation, I think we can probably do better. Let’s then eschew the metaethical formula in preference of a more thoroughly epistemological third factor.

I'd suggest that a more compelling third-factor account comes from the seemingly benign assumption that there are many reliable true beliefs<sup>12</sup> (RTBs) in the world, and facts about RTBs are natural facts that play meaningful roles in the causal explanation of successes and failures in humans and non-human organisms. How an agent will behave with respect to some domain—particularly whether she displays a pattern of successful behavior in that domain—is at least partially a function of whether that agent has RTBs about that domain. As there are clear survival advantages in predicting who is likely to display these kinds of successful behavior patterns, not the least of which is identifying good informants, our evolutionary ancestors, the story goes, developed representational capacities to track RTBs. And for many of the very same reasons cited in §4, knowledge representation was an ideal candidate for tracking RTB in others—knowledge is easier for primate brains to represent, and there is a very tight correlation between RTB and knowledge. In this way, we can explain how many of our beliefs about knowledge happen to be true. Again, none of this requires saying that knowledge is mind-independently causal, or that knowledge is RTB. All the non-naturalist realist needs to grant is that knowledge and RTB are closely correlated, and that RTB is mind-independently causal.<sup>13</sup> In fact, here the non-naturalist *cannot* grant that RTB is knowledge, because that would just be naturalism.

I'm happy to acknowledge that all this may well be correct. There probably are RTBs in the natural world, which knowledge-representing capacities emerged in part to track, as RTB-representation would have been too cognitively demanding. However, none of that can fully close the kind of explanatory gap we observed above for other third factors. In

---

<sup>12</sup> As well as, strictly speaking, reliable true representational states that don't necessarily rise to the level of *bona fide* belief.

<sup>13</sup> While one could dispute the claim that reliability is generally mind-independent (Brandom 1998, §V; for a response, see Kornblith 2002, §2.8), here we'll grant that it's probably fine.

particular, it is still a remarkable, unexplained coincidence that this process of ersatz RTB tracking happened to give us true beliefs about *knowledge*, something else, which isn't RTB, isn't causal at all, but also isn't mind-dependent. One might appeal to the fact that RTB and knowledge are closely correlated, but like we saw above with knowledge and its canonical sources, this isn't enough. Just as with perception or memory, it's still remarkable and unexplained that RTB just so happens to be knowledge, *and* we independently believe that RTB is knowledge. We may have managed to bundle many unexplained instances of fact–belief correlation into a single ur-correlation, but, I'd suggest, we're left with a lingering sense that something quite important has gone unexplained. Moreover, I'd further suggest that this RTB account is just as plausibly interpreted as a non-realist story, perhaps even more so: RTB is the real thing, and evolution invented a similar enough category to facilitate efficiently tracking RTBs most of the time. None of this gives us any reason to expect that the invented category is also mind-independently real, and it should probably even make us more doubtful.

Like everything else in this paper, such doubt is by no means conclusive. I can't rule out the possibility of a clever solution to this explanatory problem, just as I can't rule out there being some other third factor that avoids the issues of those imagined here. Instead, the point is just that there are non-trivial limitations facing third-factor explanations of true belief about knowledge, underscoring the explanatory asymmetry between evolutionary anti-realism and non-naturalist realism. While evolutionary anti-realism provides a conceptually tight, empirically grounded account of how we have many true beliefs about knowledge, the third-factor strategy of non-naturalist realism is more difficult to execute and struggles with major open questions. Again, it could still be that non-naturalist realism is preferable overall *despite* not being able to easily explain how we have many true beliefs about knowledge. The point here is simply that, despite the widespread acceptance of non-naturalist realism in

contemporary epistemology, we can identify at least some considerations that favor at least some version of moderate anti-realism. There's at least one thing—explaining true belief about knowledge—that anti-realism about knowledge is better suited to than its non-naturalist competition.

As a final word on this argument, I want to quickly point out how exactly we've made good on the promise, introduced in section 1, to avoid any additional anti-realist commitments for mental or epistemic facts. One might wonder, for instance, whether this argument motivates anti-realism for all truths about mental states that we believe on the basis of our theory of mind capacities—beliefs, desires, intentions, confidence, and so on. However, this worry can easily be circumvented by granting that facts about mental states are natural facts, playing meaningful roles in the causal explanation of everyday behaviors and actions. As this naturalism about mental states is widely held in philosophy, particularly epistemology, there's no reason that anti-realism about the mental needs to be a downstream consequence of the argument I've advanced here. And notice, moreover, that there is an equally fundamental barrier to generalizing to other epistemic facts, like facts about justification or rationality. Theory of mind doesn't support our beliefs about (non-knowledge) epistemic status, so there wouldn't be a comparable empirical case to be made for CAUSATION were we to expand its scope beyond knowledge.

## **6. Conclusion and future development**

What explains true belief about knowledge? Non-naturalist realism, arguably the default view amongst epistemologists today, struggles to offer a comprehensive response. While the realist might appeal to different kinds of third-factor accounts, these still face explanatory limitations. In contrast, evolutionary anti-realism, on which facts about knowledge are

grounded in facts about the theory of mind systems responsible for knowledge representation, offers a compelling response to this explanatory challenge. The same neurocognitive systems responsible for our beliefs about knowledge are also responsible for their truth.

It's important, however, that we don't take things too far. Not only does this argument only provide narrow support for anti-realism about knowledge, which could in principle be overridden by competing considerations, but notice too that nothing offered here provides substantial reason to prefer evolutionary anti-realism over other moderate anti-realist accounts of knowledge. For example, a kind of moderate constructivism about knowledge, on which (roughly) facts about knowledge are grounded in human attitudes about knowledge, rather than facts about the neurocognitive systems that produce these attitudes, fits nicely into the rationale I've outlined here.<sup>14</sup> Beyond meeting the explanatory challenge, such a view could also avoid the undesirable features of other anti-realist metaepistemologies discussed in section 1. And while I'm not sure that constructivism might also underwrite the methodological advantages of evolutionary anti-realism sketched below, all this is really quite beside the point. The primary aim of this paper is to challenge non-naturalist realism about knowledge by pointing out that a moderate anti-realist alternative can explain our true beliefs about knowledge in a way that non-naturalist realism cannot. The possibility that other formulations of moderate anti-realism might also be capable of meeting the explanatory challenge is entirely at ease with this objective.

Finally, as a matter of speculation, I'd suggest that there are a number of promising routes for further developing the approach proposed here. Evolutionary anti-realism might, for instance, help account for why knowledge seems like something of a misfit in both the mental and epistemic domains, with the challenge of specifying how knowledge relates to

---

<sup>14</sup> This isn't at all surprising, as Street motivates this kind of constructivism about evaluative truths with her debunking arguments (2006, §10).

both other mental states (especially belief) and other epistemic statuses proving so difficult to resolve. However, I want to close by highlighting one especially important methodological consequence of evolutionary anti-realism. The past couple of years have seen an explosion of interest in understanding the neural and cognitive mechanisms underlying our knowledge-representing capacities (especially Bricker 2020; Gordon 2021; Phillips et al. 2021a; Phillips et al. 2021b; Westra and Nagel 2021; Rakoczy and Proft 2022). On a standard, realist approach, it's not clear how this emerging knowledge-centric theory of mind program might inform epistemological questions about the nature of knowledge, as facts about knowledge need not have any robust connection with facts about how knowledge is represented. However, on evolutionary anti-realism, facts about knowledge are grounded in facts about the very knowledge-representing systems of interest to this empirical research program. Accordingly, by understanding the architecture underlying knowledge representation, we can gain direct insight into the nature of knowledge itself. Put another way, the empirical investigation of knowledge representation can be understood as directly continuous with the philosophical investigation of the nature of knowledge, opening up the possibility for significant advancements in how questions about knowledge are pursued. Behavioral measures (e.g., reaction times, eye tracking) and even neural techniques (e.g., EEG, fMRI) could be deployed in the direct service of epistemology. Ironically, non-naturalist anti-realism may be capable of delivering on naturalist realism's largely unfulfilled promise of a fundamentally empirical investigation of knowledge itself (especially Kornblith 2002).

## References

- Apperly, I. (2011). *Mindreaders: the cognitive basis of “theory of mind.”* Psychology Press.
- Baker, C., Saxe, R., & Tenenbaum, J. B. (2009). Action understanding as inverse planning. *Cognition*, 113(3), 329–349.
- Benacerraf, P. (1973). Mathematical Truth. *The Journal of Philosophy*, 70(19), 661–679.
- Birch, S. & Bloom, P. (2007). The Curse of Knowledge in Reasoning about False Beliefs. *Psychological Science*, 18(5), 382–386.
- Boghossian, P. (2005). *Fear of knowledge*. Oxford
- Bonjour, L. (2006). Kornblith on Knowledge and Epistemology. *Philosophical Studies*, 127(2), 317–335.
- Brandom, R. B. (1998). Insights and Blindspots of Reliabilism. *The Monist*, 81(3), 371–392.
- Bricker, A. M. (2020). The neural and cognitive mechanisms of knowledge attribution: An EEG study. *Cognition*, 203, 104412–.
- Bricker, A. M. (2021). Knowledge is a mental state (at least sometimes). *Philosophical Studies*.
- Carter, J. A. & Sosa, E. (2022). Metaepistemology. In *The Stanford Encyclopedia of Philosophy* (Spring 2022 Edition), Edward N. Zalta (ed.), URL = <https://plato.stanford.edu/archives/spr2022/entries/metaepistemology/>.



Chrisman, M. (2007). From Epistemic Contextualism to Epistemic Expressivism. *Philosophical Studies*, 135(2), 225–254.

Chrisman, M. (2012). Epistemic Expressivism. *Philosophy Compass*, 7(2), 118–126.

Clarke-Doane, J. (2012). Morality and Mathematics: The Evolutionary Challenge. *Ethics*, 122(2), 313–340.

Clarke-Doane, J., & Baras, D. (2021). Modal Security. *Philosophy and Phenomenological Research*, 102(1), 162–183.

Cohen, S. (1988). How to be a Fallibilist. *Philosophical Perspectives*, 2, 91–123.

Copp, D. (2008). DARWINIAN SKEPTICISM ABOUT MORAL REALISM. *Philosophical Issues*, 18(1), 186–206.

Craig, E. (1990). *Knowledge and the State of Nature: An Essay in Conceptual Synthesis*. Oxford University Press.

De Cruz, H. (2016). Numerical cognition and mathematical realism. *Philosophers' Imprint*, 16(16), 1–13.

Enoch, D. (2010). The epistemological challenge to metanormative realism: how best to understand it, and how to cope with it. *Philosophical Studies*, 148(3), 413–438.

Fabricius, W., Boyer, T. W., Weimer, A. A., & Carroll, K. (2010). True or False: Do 5-Year-Olds Understand Belief? *Developmental Psychology*, 46(6), 1402–1416.

Field, H. (1989). *Realism, mathematics and modality*. Oxford: Blackwell.

Gerken, M. (2017). *On Folk Epistemology: How we Think and Talk about Knowledge*. Oxford University Press.

Gerken, M. (2021). Representation and misrepresentation of knowledge. *Behavioral and Brain Sciences*, 44, e153–e153.

Gordon, R. M. (2021). Simulation, Predictive Coding, and the Shared World. In *The Neural Basis of Mentalizing* (pp. 237–255). Springer International Publishing.

Hannon, M. (2019). *What’s the Point of Knowledge? A Function-First Epistemology*. Oxford University Press.

Hartwright, C., Apperly, I. A., & Hansen, P. C. (2015). The special case of self-perspective inhibition in mental, but not non-mental, representation. *Neuropsychologia*, 67, 183–192.

Horschler, D., Santos, L. R., & MacLean, E. L. (2019). Do non-human primates really represent others’ ignorance? A test of the awareness relations hypothesis. *Cognition*, 190, 72–80.

Ichikawa, J. & Steup, M. (2018). The Analysis of Knowledge. In *The Stanford Encyclopedia of Philosophy* (Summer 2018 Edition), Edward N. Zalta (ed.), URL = <https://plato.stanford.edu/archives/sum2018/entries/knowledge-analysis/>.

Ichikawa, J. J. and Jenkins, C. S. I. (2017). *On Putting Knowledge “First.”* In *Knowledge First*. Oxford University Press.

Joyce, R. (2006). *The evolution of morality*. MIT Press.

Kahane, G. (2011). Evolutionary Debunking Arguments. *Noûs*, 45(1), 103–125.

Kaminski, J., Call, J., & Tomasello, M. (2008). Chimpanzees know what others know, but not what they believe. *Cognition*, 109(2), 224–234.

Khalidi, M. (2015). Three Kinds of Social Kinds. *Philosophy and Phenomenological Research*, 90(1), 96–112.

Kornblith, H. (2002). *Knowledge and its place in nature*. Clarendon.

Krachun, C., Carpenter, M., Call, J., & Tomasello, M. (2009). A competitive nonverbal false belief task for children and apes. *Developmental Science*, 12(4), 521–535.

Krupenye, C. (2021). The Evolution of Mentalizing in Humans and Other Primates. In *The Neural Basis of Mentalizing* (pp. 107–129). Springer International Publishing.

Krupenye, C. & Call, J. (2019). Theory of mind in animals: Current and future directions. *Wiley Interdisciplinary Reviews. Cognitive Science*, 10(6), e1503–n/a

Leslie, A. (1987). Pretense and Representation: The Origins of “Theory of Mind.” *Psychological Review*, 94(4), 412–426.

Levy, A. & Levy, Y. (2020). Evolutionary Debunking Arguments Meet Evolutionary Science. *Philosophy and Phenomenological Research*, 100(3), 491–509.

Logins, A. (2021). Persistent burglars and knocks on doors: Causal indispensability of knowing vindicated. *European Journal of Philosophy*.

MacFarlane, J. (2010). Relativism and knowledge attributions. In *The Routledge Companion to Epistemology*. Bernecker, S., & Pritchard, D. (Eds.). Taylor & Francis.

Mahy, C., Moses, L. J., & Pfeifer, J. H. (2014). How and where: Theory-of-mind in the brain. *Developmental Cognitive Neuroscience*, 9(C), 68–81

Millar, A. (2007). THE STATE OF KNOWING. *Philosophical Issues*, 17(1), 179–196.

Nagel, J. (2010). Knowledge ascriptions and the psychological consequences of thinking about error. *Philosophical Quarterly*, 60(239), 286–306.

Nagel, J. (2012). Intuitions and Experiments: A Defense of the Case Method in Epistemology. *Philosophy and Phenomenological Research*, 85(3), 495–527.

Nagel, J. (2013). Knowledge as a mental state. In *Oxford studies in epistemology volume 4*.

Nozick, R. (2001). *Invariances: the structure of the objective world*. Cambridge, Mass: Belknap Press of Harvard University Press.

Olson, J. (2014). *Moral Error Theory: History, Critique, Defence*. Oxford University Press.

Pernu, T. (2009). Is Knowledge a Natural Kind? *Philosophical Studies*, 142(3), 371–386.

Phillips, J., Buckwalter, W., Cushman, F., Friedman, O., Martin, A., Turri, J., Santos, L., & Knobe, J. (2021a). Knowledge before belief. *Behavioral and Brain Sciences*, 44, e140–e140.

Phillips, J., Buckwalter, W., Cushman, F., Friedman, O., Martin, A., Turri, J., Santos, L., & Knobe, J. (2021b). Actual knowledge. *The Behavioral and Brain Sciences*, 44, e177–e177.

Phillips, J., Knobe, J., Strickland, B., Armary, P., & Cushman, F. (2018). Knowledge before belief: Response-times indicate evaluations of knowledge prior to belief. In *Proceedings of the fortieth annual conference of the cognitive science society*.

Phillips, J. & Norby, A. (2021). Factive theory of mind. *Mind & Language*, 36(1), 3–26.

Povinelli, D. & Preuss, T. M. (1995). Theory of mind: evolutionary history of a cognitive specialization. *Trends in Neurosciences (Regular Ed.)*, 18(9), 418–424.

Premack, D., & Woodruff, G. (1978). Does the chimpanzee have a theory of mind? *Behavioral and Brain Sciences*, 1(4), 515-526.

Pritchard, D. (2012). Anti-luck Virtue Epistemology. *The Journal of Philosophy*, 109(3), 247–279.

Rakoczy, H. & Proft, M. (2022). Knowledge before belief ascription? Yes and no (depending on the type of “knowledge” under consideration). *Frontiers in Psychology*, 13, 988754–988754.

Rose, D., Machery, E., Stich, S., Alai, M., Angelucci, A., Berniūnas, R., ... Zhu, J. (2019). Nothing at Stake in Knowledge. *Noûs*, 53(1), 224–247.

Roth, P. (2003). Review of Hilary Kornblith, *Knowledge and Its Place in Nature*. *Notre Dame Philosophical Reviews* 12.

Rysiew, P. (2021). Epistemic Contextualism. In *The Stanford Encyclopedia of Philosophy* (Spring 2021 Edition), Edward N. Zalta (ed.), URL = <https://plato.stanford.edu/archives/spr2021/entries/contextualism-epistemology/>.

Schechter, J. (2010). THE RELIABILITY CHALLENGE AND THE EPISTEMOLOGY OF LOGIC. *Philosophical Perspectives*, 24(1), 437–464.

Schechter, J. (2018). Explanatory Challenges in Metaethics. In *The Routledge Handbook of Metaethics*. Routledge

Searle, J. (1996). *The construction of social reality*. Penguin.

Skarsaune, K. O. (2011). Darwin and moral realism: survival of the fittest. *Philosophical Studies*, 152(2), 229–243.

Smaers, J., Gómez-Robles, A., Parks, A. N., & Sherwood, C. C. (2017). Exceptional Evolutionary Expansion of Prefrontal Cortex in Great Apes and Humans. *Current Biology*, 27(10), 1549–1549.

Southgate, V. & Vernetti, A. (2014). Belief-based action prediction in preverbal infants. *Cognition*, 130(1), 1–10.

Street, S. (2006). A Darwinian Dilemma for Realist Theories of Value. *Philosophical Studies*, 127(1), 109–166.

Street, S. (2008). REPLY TO COPP: NATURALISM, NORMATIVITY, AND THE VARIETIES OF REALISM WORTH WORRYING ABOUT. *Philosophical Issues*, 18(1), 207–228.

Street, S. (2010). What is Constructivism in Ethics and Metaethics? *Philosophy Compass*, 5(5), 363–384.

Streumer, B. (2017). *Unbelievable Errors: An Error Theory about All Normative Judgements*. Oxford University Press.

Turri, J. (2017). Knowledge Attributions and Behavioral Predictions. *Cognitive Science*, 41(8), 2253–2261.

van der Meer, L., Groenewold, N. A., Nolen, W. A., Pijnenborg, M., & Aleman, A. (2011). Inhibit yourself and understand the other: Neural basis of distinct processes underlying Theory of Mind. *NeuroImage*, 56(4), 2364–2374.

Vavova, K. (2015). Evolutionary Debunking of Moral Realism. *Philosophy Compass*, 10(2), 10–4116.

Westra, E., & Nagel, J. (2021). Mindreading in conversation. *Cognition*, 210, 104618–104618.

Williamson, T. (2000). *Knowledge and its limits*. Oxford University Press.

Williamson, T. (2005). Contextualism, subject-sensitive invariantism and knowledge of knowledge. *The Philosophical Quarterly*, 55(219), 213–235.