

Knowledge is a mental state (at least sometimes)

Penultimate draft. Forthcoming in *Philosophical Studies*.

Adam Michael Bricker (University of Cologne)

Abstract:

It is widely held in philosophy that knowing is not a state of mind. On this view, rather than knowledge itself constituting a mental state, when we know, we occupy a belief state that exhibits some additional non-mental characteristics. Fascinatingly, however, new empirical findings from cognitive neuroscience and experimental philosophy now offer direct, converging evidence that the brain can—and often does—treat knowledge as if it is a mental state in its own right. While some might be tempted to keep the metaphysics of epistemic states separate from the neurocognitive mechanics of our judgements about them, here I will argue that these empirical findings give us sufficient reason to conclude that knowledge is at least sometimes a mental state. The basis of this argument is the epistemological principle of neurocognitive parity—roughly, if the contents of a given judgement reflect the structure of knowledge, so do the neurocognitive mechanics that produced them. This principle, which I defend here, straightforwardly supports the inference from the empirical observation that the brain sometimes treats knowledge like a mental state to the epistemological conclusion that knowledge is at least sometimes a mental state. All told, the composite, belief-centric metaphysics of knowledge widely assumed in epistemology is almost certainly mistaken.

Keywords: Knowledge-first; Experimental; Epistemology; Theory of Mind; Mentalizing; Knowledge Attribution

1. Introduction: Knowledge as a Mental State

Contrasting with paradigm mental states like belief, desire, and pain, it is widely held in philosophy that knowing is not a state of mind in its own right. Instead, most epistemologists take knowledge to be what might be described as a “composite” (Nagel 2013, 275) or “metaphysical hybrid” (Williamson 2000, 50), on which being in the state of knowledge requires both being in the mental state of belief and fulfilling some additional non-mental conditions. On this picture, while being in a mental state (i.e. belief) is *necessary* for knowing that *p*, there is no mental state for which being in that state is *sufficient* for knowing that *p*. This position has been famously challenged by both Williamson (2000) and Nagel (2013), who maintain that knowledge is in fact a mental state in its own right. That is, there is some mental state for which being in that state is both *necessary and sufficient* for knowing that *p*. Nevertheless, while this position has been widely discussed (see e.g. Greenough and Pritchard 2009; Gendler and Hawthorne 2013, chapters 10-13; McGlynn 2014; Carter et al. 2017), it has attracted little additional support beyond its two main proponents. As summarized by Nagel, “the current climate in philosophy is largely hostile to the idea that knowledge is a mental state” (2013, 274).

My aim in this paper is to advance a new line of support for the claim that knowledge is a mental state. This support will come by way of new behavioral and neurocognitive findings—including our first direct look into how the brain attributes knowledge—together with an epistemological parity principle that facilitates an inference from how the brain attributes knowledge to what knowledge is actually like. While I don't want to say that we can always infer the nature of knowledge from how the brain attributes it, I will argue that we have compelling reason to think that, in the case of these specific empirical findings, we can make such an inference. As, in this particular case, the brain attributes knowledge like a mental state, we can then conclude that knowledge, in this particular case, is in fact a mental state.

Going into this argument, however, it is important that we don't misunderstand the precise nature of the support it lends to the thesis that knowledge is a mental state. Crucially, this argument alone only gives us reason to think that there is a mental state for which being in that state is *sufficient* for knowing that *p*. Put another way, this argument can presently only support that claim that knowledge is at least sometimes a mental state, not that it is always a mental state. This follows directly from the empirical nature of the argument. Since we currently only know how the brain attributes knowledge in a small range of cases, we cannot eliminate the possibility that in certain other cases the brain attributes knowledge like a composite. In such a case, being in some type of mental state would remain necessary for knowing that *p*, but what type mental state this is might vary. For this reason, we should refrain from concluding on the basis of this argument alone that knowledge is a mental state, full stop.

Nevertheless, if one is already sympathetic to the idea that knowledge is a mental state—in the sense that being in some mental state is both necessary and sufficient for knowing—then my argument for sufficiency of course lends additional support to the idea as a whole. Similarly, if one is of the conviction that knowledge must display a unitary metaphysics,¹ my argument again supports the broader claim that knowledge is a mental state. While I am hesitant to make such a commitment, finding it far too easy to imagine discovering cases in which the brain attributes knowledge like a composite, I do recognize that this is the natural conclusion if we think that no knowledge states might be more mental than others.

Furthermore, even without any firm commitment to the idea that knowledge is always a mental state, the observation that it is at least sometimes a mental state is sufficient to undermine the widely held composite metaphysics of knowledge.² Whatever the exact metaphysical composition of knowledge ultimately proves to be, at this point we can be confident that a unitary composite framework, on which knowledge is always a composite of belief with additional non-mental features, is insufficient. It is in this precise manner that the argument I'll present here offers a motivation for moving away from the idea that knowledge is just a species of belief.

In developing this argument, this paper is structured in the following way: First, I'll start with a discussion of the neurocognitive mechanisms responsible for our judgements about knowledge, which provide compelling evidence that the brain often attributes simple perceptual knowledge like it is a mental state (§2). Next, I'll examine Nagel's argument that the mechanics of mental state attribution provide a reliable guide to the metaphysics of knowledge, highlighting a key shortcoming of this approach (§3). After this, I'll present my

¹ That is, different knowledge states cannot have varying metaphysical compositions. If knowledge is sometimes just a mental state, then knowledge is only ever just a mental state.

² Far too many accounts of knowledge assume this composite metaphysics to exhaustively recount here. Some of the most influential come from Goldman (e.g. 1986), Kornblith (2002), Pritchard (e.g. 2005; 2012), and Sosa (e.g. 1999; 2007). See Ichikawa and Steup (2018) for an overview of many of these accounts.

own argument that the way in which the brain processes knowledge attributions should inform how we think about the metaphysics of knowledge, an argument which derives from the principle of neurocognitive parity (§4). Finally, I'll close by discussing some of the questions left open by my present account, along with its most significant implications (§5).

2. Knowledge Attribution in the Brain

If we are to infer the nature of knowledge from how the brain computes knowledge attributions, as is the goal of this paper, we'll need at least two key ingredients. One of these ingredients is some sort of epistemological principle to facilitate this inference, which will be the topic of the following two sections. First, however, in this section I want to provide the other key ingredient, an empirical perspective on how the brain computes knowledge attributions. This perspective will situate pioneering behavioral findings from Phillips et al. (2018) and my own EEG study (Bricker 2020)—the first to apply neuroimaging techniques to knowledge attribution—within a broader context offered by experimental philosophy and developmental and comparative psychology (see especially Phillips et al. 2020). The main takeaway from these studies is simple: At least for cases of ordinary perception, the brain does not appear to attribute knowledge like a composite state. Instead, at the neurocognitive level, this knowledge is attributed like a mental state in its own right. Here I'll limit discussion to the neurocognitive properties of knowledge attribution, reserving the epistemological implications for the following sections.

To start, it might be helpful to note that the neurocognitive perspective I'll offer here differs from the biases-and-heuristics cognitive psychology of knowledge attribution favored by some epistemologists in the past decade (e.g. Nagel 2010; Gerken 2012). I have a number of reasons for preferring the neurocognitive perspective. The first is simply a matter of generality. While the biases-and-heuristics approach focuses on narrow bands of knowledge attribution, especially odd cases in which something might be going wrong, here I'm interested in the mechanisms that govern ordinary, everyday cases in which everything seems to go right. My second reason is the more practical matter of what empirical evidence we presently have available to us. While the studies I'll discuss (primarily Phillips et al. 2018 and Bricker 2020) provide us with a direct look into the neurocognitive mechanisms of *knowledge* attribution, both the Nagel and Gerken accounts were largely based on biases observed empirically for *belief* attribution. This seems to have derived from the tendency of psychology-oriented philosophers of knowledge attribution to assume that “knowledge involves belief or a belief-like attitude” (Gerken 2012, 148). However, as we'll see shortly, empirical evidence now suggests that this assumption is mistaken. Belief attribution seems to recruit neurocognitive processes (i.e. self-perspective inhibition) in ways that knowledge attribution does not. Accordingly, we cannot safely assume that the mechanisms of belief attribution generalize to knowledge attribution—direct empirical evidence from studies about knowledge attribution is needed. As is illustrated by a number of recent accounts of knowledge attribution (Bricker 2020; Phillips et al. 2020; Westra and Nagel 2021), all of which eschew the belief-derived biases-and-heuristics approach, the knowledge-first program likely represents the future of this research. Finally, the neurocognitive perspective also offers the opportunity for *converging evidence* in a way that isn't easily replicated by cognitive psychology. For example, as we'll see below, electrophysiological measurements, reaction time data, and attribution reports all point in the same direction, which allows us to make more confident conclusions about the mechanisms of knowledge attribution. In short, for all these reasons, contrasting with certain previous philosophical approaches to

knowledge attribution, our starting point here will be the neurocognitive perspective: What can cognitive neuroscience tell us about how we think about the mental states of others?

As it happens, there is a wealth of research in cognitive neuroscience regarding how the brain processes the mental states of others, a cognitive capacity often referred to as “Theory of Mind” (ToM; for an overview, see Heleven & Van Overwalle 2018). This research focuses primarily on belief states and has revealed a number of distinct neurocognitive processes that contribute to belief attribution. Here I will focus on two such processes: (i) perspective taking and (ii) self-perspective inhibition. First, in order to form accurate judgements about what S believes, it is necessary to both take S’s perspective and inhibit one’s own self-perspective (Heleven & Van Overwalle 2018). At the intuitive level, this is easy to make sense of—What S believes is of course a function of information available on her perspective, and self-perspective information might interfere with the process of taking S’s perspective, especially when the information available on the two perspective differs. Less obvious at the intuitive level, however, is that these two processes are neurocognitively distinct. Not only have neuropsychological studies observed a double dissociation between self-perspective inhibition and perspective taking (Samson et al. 2004; Samson et al. 2005), but a number of fMRI studies indicate distinct neural correlates for the two processes, with perspective taking being broadly associated with the temporoparietal junction (Schuwerk et al. 2014; Özdem et al. 2019) and self-perspective taking the inferior frontal gyrus (van Der Meer et al. 2011; Hartwright et al. 2012; Hartwright et al. 2015). Fascinatingly, it even appears that self-perspective inhibition is distinct from more general inhibitory control in the brain (Hartwright et al. 2015). In short, we have good reason to think that belief attribution recruits (at least) two distinct neurocognitive process: perspective taking and self-perspective inhibition.

Although cognitive neuroscience has revealed a good deal about how the brain processes belief attributions, as well as mental state attribution generally, until quite recently, surprisingly little attention has been directed towards the direct study of *knowledge* attribution. While comparative and developmental psychology have long studied the respective phylogeny and ontogeny of knowledge representation (for an excellent review, see Phillips et al. 2020), the processes underlying the computation of knowledge attributions in the adult human brain have gone comparatively overlooked. Accordingly, we are only beginning to understand the extent to which the neurocognitive mechanics of belief attribution are shared by knowledge attribution.

An important early step in this direction comes from my own recent neurocognitive study (Bricker 2020), which used EEG to record neural activity during both belief and knowledge attribution. This study employed simple cartoons as stimuli, which varyingly depicted states of belief, knowledge, and ignorance, and was specifically designed to directly investigate whether the brain attributes knowledge like a composite or mental state. Fascinatingly, in no uncertain terms, the results of the study strongly indicate that knowledge is in fact attributed like a mental state by our neurocognitive systems, a conclusion which was supported by two independent lines of evidence: (1) Reaction time data showed no significant differences between belief and knowledge attribution conditions. If the brain computed knowledge attributions like knowledge were a composite—i.e. taking belief attribution to be a stage of knowledge attribution—we would of course expect the computation of knowledge attributions to take longer. (2) Knowledge attribution doesn’t display the same strong self-perspective inhibition characteristic of belief attribution. Were belief attribution a stage of knowledge attribution, we would of course expect the same strong self-perspective inhibition for knowledge attribution. As I concluded at the time, these results “provide quite compelling behavioral and electrophysiological evidence in favor of the thesis that knowledge is attributed like a mental state” (2020, 8).

The findings of this study converge with previous behavioral findings reported by Phillips et al. (2018), which observed faster response times for knowledge attribution vs. belief attribution. Using short vignettes featuring an epistemic agent in some simple scenario, this study asked participants to evaluate whether statements about the agent's beliefs or knowledge were true or false. Phillips et al. reported faster response times for judgements about knowledge vs. belief not only in cases of attribution, but also in negative evaluations of the respective states. Again, this observation is flatly inconsistent with a belief computation stage for judgements about knowledge. As put by the authors, "knowledge assessment may often not depend on prior evaluations of belief" (2018, 2263).

It would of course be a bit hasty to make any definitive conclusions on the basis of a pair of empirical studies, and I'll be the first to concede that additional neurocognitive research into knowledge attribution is sorely needed. However, it would be equally mistaken to take these studies as isolated data points. Instead, the assertion that the brain attributes knowledge like a mental state is best understood as a direct confirmation of what we've long had reason to suspect. Not only do previously reported knowledge attribution reports (Myers-Schulz and Schwitzgebel 2013; Murray et al. 2013) indicate that knowledge attribution can occur in the absence of a belief-attribution stage, but as noted by both myself (Bricker 2020) and Nagel (2013), those working in the ToM research program often already include knowledge in their list of paradigm mental states. Nagel goes further, offering a comprehensive account of evidence from both developmental and comparative psychology that points in this direction (2013, §3)—Most notably, both young human children (e.g. Hogrefe et al. 1986) and certain non-human primates (e.g. Kaminski et al. 2008) display a knowledge-attributing capacity that outstrips their belief-attributing capacity. More recently, Phillips et al. have similarly argued from a wide range of evidence—spanning comparative and developmental psychology, experimental philosophy, and cognitive science—that knowledge representation is more basic than belief representation, going so far as to advocate for a "shift of focus" in Theory of Mind research towards a more knowledge-centric approach (2020, 1). Together, all this constitutes a compelling, convergent case that the brain attributes knowledge like a mental state in its own right.

As exciting as all this is, however, at this point we need to put the brakes on any excess enthusiasm to ensure that things don't get out of hand. None of the available evidence suggests, in any way, that the neurocognitive systems of human adults *exclusively* attribute knowledge like a mental state. Developmental and comparative psychology doesn't directly tell us anything about everyday knowledge attribution in human adults; experimental philosophy only offers evidence from a handful of highly specific cases (Myers-Schulz and Schwitzgebel 2013; Murray et al. 2013); and the only cognitive (Phillips et al. 2018) and neurocognitive (Bricker 2020) evidence presently available exclusively involves the attribution of simple perceptual knowledge. To be clear, there isn't presently any neurocognitive evidence that the brain sometimes also attributes knowledge like a composite. However, we cannot at this time exclude this due to a simple lack of research into the neurocognitive mechanics of knowledge attribution. Accordingly, here we need to be careful to refrain from making any general claims that go beyond what is presently supported by extant empirical evidence. While the brain does attribute knowledge like a mental state, at least in cases of ordinary perceptual knowledge, we cannot confidently conclude at this time that it always attributes knowledge like a mental state.

With this final word of moderation, we've reached the end of the empirical portion of this paper. Now comes the challenging question, which constitutes our primary concern here—Can we infer that (at least) perceptual knowledge actually can be a mental state on the basis of how our ToM systems attribute knowledge? I of course think that we can, but this inference is anything but trivial. In the next section, I'll first consider Nagel's argument that

the metaphysics of knowledge applied by our knowledge-attributing cognitive systems should be treated at broadly correct, before highlighting one of this argument's key shortcomings. Following that, I'll introduce my own, quite different argument for the same sort of conclusion.

3. Nagel's Argument

In the previous section, I discussed some exciting new empirical findings regarding knowledge attribution. These findings directly confirm what indirect empirical evidence already indicated—the brain can and often does attribute knowledge to others like knowledge is a mental state. On this basis, I want to argue against the composite view of knowledge favored by many epistemologists. That is, I think we might reasonably conclude that knowledge is (at least sometimes) a mental state on the grounds that our knowledge-attributing systems treat it like one. As it so happens, an argument of this form has already been advanced by Nagel (2013, §4), which will be our focus for this section. Ultimately, I will conclude that Nagel's argument commits to an empirical assumption that is not generally defensible—the assumption that the accuracy of a cognitive process is evidence for the veracity of the metaphysics it applies.³ For this reason, I don't think that this argument presents the best path forward. Before discussing this shortcoming in detail, however, I first need to say a bit about the argument itself.

The context from which Nagel's argument originates isn't entirely unlike the situation in which we find ourselves now.⁴ After detailing empirical findings indicating that our ToM systems attribute knowledge like a mental state, Nagel notes that this alone might not be considered sufficient to grant that knowledge actually is a mental state:

Even if our natural mindreading⁵ systems parse agency with the help of the epistemic mental state concepts of knowledge and belief rather than belief alone, one might worry that mental reality could fail to correspond to our natural way of tracking it. (2013, 301)

This, of course, is none other than the very challenge that the present paper seeks to address. Nagel's strategy for solving this problem is somewhat tricky to parse, but it centers on the idea that knowledge-based explanations generally facilitate more accurate predictions about the actions of others than belief-based explanations. The key contrast Nagel draws is between (1) "intuitive" (2013, 301) knowledge-based predictions, which rely on "trust in our natural instincts" (2013, 302), and (2) belief-based predictions, which rely on "reflection" and more "cognitive effort" (2013, 302). As the belief-based cognitive strategy "generates an awkwardly large space of possibilities to be contemplated, and our capacity to reason accurately about all those possibilities is limited" (2013, 302), we might understand why the knowledge-based strategy—which doesn't entail this "serial contemplation of various possibilities" (2013, 302)—might "[enable] us to make better predictions" (2013, 302) about

³ Note that this shortcoming has been previously noticed by McGlynn (2014, 186).

⁴ There are of course some key contextual dissimilarities as well. Most notably, Nagel didn't have direct neurocognitive evidence available to her. Additionally, and perhaps for this reason, the relatively short text dedicated to this present argument indicates that Nagel considered the empirical component more demanding. Her argument from empirical findings to the metaphysics of knowledge, presented under the header "Conclusion," reads almost like an afterthought. My attitude is almost diametrically opposed. Especially given the empirical findings discussed above (§2), establishing what our knowledge-attributing systems do is the easy part. The epistemologically interesting bit is establishing the conditions on which the brain's metaphysics of knowledge ought to be taken seriously.

⁵ Mindreading is roughly synonymous with ToM here.

the actions of others. On the basis of this putatively greater accuracy of the knowledge-based strategy—which I’ll not dispute here⁶—Nagel concludes that knowledge really is, ontologically, a mental state.

Nagel’s conclusion, crucially, rests on the key assumption that this accuracy generally constitutes evidence that the correct mental state ontology is being applied by our knowledge-attributing systems. Otherwise, it wouldn’t be clear why the effectiveness of treating knowledge like a mental state in predicting action has any bearing on whether knowledge is actually a mental state. However, this assumption is, at a minimum, misguided. In order to understand why, let’s first note that what it appears Nagel has done is appeal to something like a two-systems account (see Evans 2008) for cognitive strategies of action prediction. Here, the knowledge-based strategy resembles a type-I process: fast, intuitive, and heuristic in the strict sense that it systematically excludes some of the available information. In contrast, the belief-based strategy resembles a type-II process: reflective, requiring more cognitive effort, and incorporating more available information. This heuristic characterization of the knowledge-based strategy, however, should immediately raise red flags for the key assumption of Nagel’s argument. Especially when dealing with cognitive heuristics, not only is the accuracy of a cognitive strategy for predicting Xs not generally associated with the application the correct metaphysics underlying X, it is in fact often associated with application of the *incorrect* metaphysics underlying X.

This frequently inverse relationship between the predictive accuracy of a cognitive strategy and the veracity of the metaphysics applied by that strategy is clearly illustrated by the “less-is-more effects” displayed by a number of cognitive heuristics (see Gigerenzer & Gaissmaier 2011). While not usually discussed in such terms, a heuristic displaying a less-is-more effect is generally characterized by a high degree of accuracy in some task predicting Xs, often matching or even outperforming far more complex statistical methods, despite not applying anything close to the correct metaphysics of X. For example, the recognition heuristic—on which one of two alternatives is selected on the basis of being recognized—has been demonstrated to produce surprisingly accurate predictions about victories in tennis tournaments (Serwe & Frings 2006; Scheibehenne & Bröder 2007) and performance of investment portfolios (Ortmann et al. 2008). However, it would of course be a mistake to make any inferences about recognizability being constitutive of the metaphysics of athletic prowess, the soundness of investment, or whatever concepts we might take to *actually explain* the respective observations the recognition heuristic successfully predicts. Something similar holds for a number of other cognitive heuristics,⁷ and the point here is quite clear: Especially where heuristics are involved, we simply cannot reasonably assume that predictive accuracy (e.g. for human action) is associated with the correct underlying metaphysics (e.g. for the mental states underlying human action). For this reason, at least as it stands, I don’t think that Nagel’s approach represents the best way forward.

One way we might address this shortcoming in Nagel’s argument would be to double down on the cognitive psychology—We might ask whether knowledge attribution actually displays the characteristics of heuristic decision making, which could then serve as a guide to whether we attribute knowledge by applying the actual metaphysics of knowledge. However, especially for the reasons discussed at the start of section two, I think that our energy can be spent more efficiently by applying an entirely different strategy: identifying an epistemological principle that allows us to plausibly infer the metaphysical composition of knowledge on the basis of the neurocognitive mechanisms of knowledge attribution. In the next section, my goal is to do just that.

⁶ Although note that Phillips et al. have argued precisely the opposite, maintaining that knowledge attribution is “poorly suited to predicting other agents’ behavior” (2020, §6.1).

⁷ At a minimum, the fluency and “one reason” heuristics (Gigerenzer & Gaissmaier 2011).

4. Neurocognitive Parity

My aim with this section is to argue in support of an epistemological principle, which I've opted to call "neurocognitive parity." This principle, I will maintain, is quite a modest proposal. It strikes me as highly intuitive, and I struggle to anticipate how one might plausibly object to it. Crucially, however, this principle is invaluable for our purposes, as it provides a straightforward basis on which we might infer that knowledge is a mental state on the basis that our neurocognitive knowledge-attributing systems treat it like one. In arguing in support of this principle, this section will begin by considering a familiar case—that of the potentially biased cognitive mechanism for knowledge attribution—in which I think we can already see something like neurocognitive parity at work. After this, I'll provide my preferred formulation for this principle, along with an account of why it seems so plausible. Finally, I'll close by discussing how it straightforwardly allows us to conclude, on the basis of the empirical findings outlined in section two, that knowledge is indeed a mental state.

Let's start by imagining a case in which the content of some class of judgements about knowledge appears to have major implications for how we theorize about the structure of knowledge. Perhaps, in the content of these judgements, knowledge is attributed in ways that are inconsistent with some dominant theory of knowledge, suggesting some mistake in this theory. However, imagine further that epistemologists cannot agree on whether this judgement content actually reflects the structure of knowledge, and therefore cannot agree on whether we need to adjust our theories of knowledge to accommodate it. Perhaps some epistemologists don't share in this judgement content, others only weakly, while still others hold this dominant theory so sacrosanct as to preclude the legitimacy of this judgement content. In short, regardless of the reason, it isn't clear whether this class of judgement content actually reflects the structure of knowledge. Now, let's posit that these judgements form via a cognitive process that is in some way faulty, so that, in the epistemic vignettes associated with this class of judgements, the judgement mechanics clearly don't reflect the structure of knowledge.⁸ Instead, the cognitive processes that produce this judgement content are sensitive to some entirely different feature of these cases.

Provided that the mechanics via which these judgements form do not reflect the structure of knowledge, I now want to ask whether we think that the *content* of these judgements reflects the structure of knowledge, thereby giving us reason to revise our dominant theory of knowledge to accommodate this content. At the risk of overplaying my hand, I'd submit that the appropriate response here is, "No. Quite obviously no." Not only is my own personal response a quite definitive *no*, I struggle to imagine why anyone would take the content of some epistemic judgement to tell us something about the structure of knowledge when the mechanics that produced this judgement are known to be faulty, not themselves reflecting the structure of knowledge but some other feature in the case. Moreover, it doesn't appear that I am alone here. As it happens, we see this sort of intuition reflected in a number of prominent arguments that have been advanced in epistemology. While differing on the purportedly faulty mechanics, these arguments have all maintained that the contents of certain judgements (particularly those associated with contextualism) don't reflect the structure of knowledge, all on the basis that the underlying judgement mechanics fail to do so: Hawthorne (2004) and Williamson (2005) have both argued this on the basis of the availability heuristic, Nagel via egocentric bias (2010), and Gerken from "epistemic focal bias" (2013). While we have reason to question whether any of these putatively faulty mechanics are actually realized in knowledge attribution (discussed in §2;

⁸ It is important to not confuse this sense of "faulty" with unreliability. As discussed in the previous section, it could still be that these faulty mechanisms generally produce correct judgements despite not reflecting the structure of knowledge.

see also Bricker 2020, §4.4), this isn't the point. Instead, the point here is that, *were* the mechanics to fail to reflect the structure of knowledge in any of the suggested ways, this would clearly give us ample reason to conclude the same about the content they produce.

I'd suggest that our intuition here is a reflection of an epistemological parity principle, a principle I'll choose to express in the following way:

NEUROCOGNITIVE PARITY: The contents of judgement J about knowledge reflect the structure of knowledge only if the mechanics of the neural and cognitive processes responsible for J also reflect the structure of knowledge.

With the principle of neurocognitive parity now laid out in this way, I want to clarify some of the terms I've been using from the onset of this section. First, the *content* of an intuitive judgement about knowledge simply refers to what some particular case is judged to be, knowledge or not knowledge. To take a familiar example, in the original Gettier cases (1963), the content of our judgements is that S (i.e. Smith or Jones) doesn't know. Conversely, the *mechanics* referred to here are the empirical properties of knowledge-attributing processes, like those discussed in section two. For example, the mechanics we are presently interested in are core Theory of Mind systems with reduced self-perspective inhibition—critically with no belief-attribution stage—which we can then express as the brain attributing knowledge “like a mental state in its own right.” Next, *structure* here refers both the metaphysical composition of knowledge states (Are they composite or entirely mental?) as well as additional properties we might think are associated with knowledge states (e.g. Are they factive? Are they safe? etc.). Finally, likely the most important idea to nail down here is what it means for some component of our judgements about knowledge—either its content or mechanics—to *reflect* the structure of knowledge. Although I'm not sure that I can provide a complete account of this reflection relation, roughly speaking, it just means standing in the right sort of evidential relation with the concept of knowledge to infer the structure of knowledge from properties of judgement content/mechanics. Moreover, as I think we have a strong intuitive sense of what it means to reflect (or not reflect) the structure of knowledge, I'm not sure that a complete account is entirely necessary. After all, we freely apply this idea when we make arguments about the structure of knowledge on the basis of our judgement content in some case, and I don't think it's especially cryptic to claim that, “The way in which knowledge is attributed in the brain reflects the structure of knowledge.” Nevertheless, to perhaps clarify things a bit further, I can offer something of a heuristic test for reflection—Do you think that our theories of knowledge should be expected to accommodate some given judgement's contents or mechanics? If so, then you likely think that those contents/mechanics reflect the structure of knowledge.

Taking all this together, the principle of neurocognitive parity then says that, if *what* we judge about knowledge in some case stands in the right sort of evidential relation with the concept of knowledge to tell us something about its (e.g.) metaphysical composition, then the cognitive and neural details of *how* this judgment forms also stands in the right sort of evidential relation with the concept of knowledge to tell us something about its metaphysical composition. Put a bit more simply, if what we judge is potential evidence for the nature of knowledge, so too is how the brain produces those judgements. This might be the easiest way to think about Neurocognitive Parity—an *if what, then how* principle. To whatever extent we take the *what* of knowledge attribution to potentially indicate the (e.g.) metaphysical composition of knowledge, we must also grant the same for the *how*. To be clear, neurocognitive parity of course doesn't say that the mechanics of knowledge attributions always reflect the structure of knowledge, just as we don't want to say something similar about the content of every knowledge attribution. The point of neurocognitive parity is

simply that when the content of judgements about knowledge can evidence the structure of knowledge, so too can the neurocognitive mechanics responsible for that content.

It is my contention that something like neurocognitive parity is overwhelmingly plausible. Its plausibility arises from the simple fact that the contents of judgements about knowledge are inextricably linked with the neurocognitive processes that produce them. If a judgement about knowledge is non-accidentally indicative of the structure of knowledge—indicative in a way that allows us to make reliable inferences about knowledge’s metaphysical composition—this evidential value must be in virtue of the processes by which it was generated. Otherwise, it is unclear how we might explain this evidential potential in the first place.⁹ Allowing for the quite benign supposition that intuitive judgements about knowledge are indicative of the structure of knowledge in virtue of the processes that produce them,¹⁰ we might then ask how it is that these processes encode the correct structure of knowledge into a judgement’s content. The answer, I would submit, is simply that these processes themselves reflect the structure of knowledge. To the extent that the content of our epistemic judgements evidences the structure of knowledge, it is the mechanics of the neurocognitive processes that guide these judgements which are responsible for this. That is to say, the mechanics must, in some way, reflect the actual structure of knowledge. This is why, as we saw at the beginning of this section, we so readily withdraw the evidential status of the *content* of epistemic judgements upon realizing that the mechanics of those judgements don’t actually reflect the structure of knowledge. Apart from those mechanics reflecting the structure of knowledge, it just isn’t clear why we might think the same for the content they produce.

For this reason, I am quite confident that we might reasonably accept the principle of neurocognitive parity as a rather mild expression of the mechanisms by which the evidential potential of judgements about knowledge arises in the first place. To be clear, I don’t want to argue here that neurocognitive parity holds for every conceivable judgement about knowledge, if for no other reason than this claim is quite strong, and I don’t have to make it. Instead, the safer thing to do is just take neurocognitive parity as something like the default state for our judgements about knowledge. In the absence of some case-specific reason to think otherwise, we should be entirely comfortable with the idea that judgement content reflects the structure of knowledge only if judgement mechanics do. Beyond this, and until such time as a putative case-specific reason for eschewing the default might be identified, I’m not sure I have anything further to add on neurocognitive parity.

With this understanding of the rationale behind neurocognitive parity in mind, we can now apply the principle to the mechanics of knowledge attribution discussed in section two, thereby completing the argument that knowledge is (at least sometimes) a mental state. The structure of this argument is quite straightforward and might be expressed in the following way:

(P1) There is a class of judgements *C* about cases of simple perceptual knowledge for which the following two conditions hold:

⁹ Indeed, the idea that the reliability of our ToM capacities explains the evidential potential of the judgements they produce is the core idea of Nagel’s influential defense of the use of intuitive judgements in epistemology (2012).

¹⁰ There is a potential worry here that the capacity for different substrates to realize the same judgement content, which is at least conceptually plausible, might undermine this assumption. However, I would note that everything I say here is entirely compatible with different sets of physical structures realizing the same judgement content/mechanics. For example, it need not be the case that just the TPJ (or even just brains) support belief-stage-free knowledge-attributing processes for us to think that the content of these judgements reflects the structure of knowledge in virtue of not having a belief-attribution stage. In short, it is the functional, not structural, characteristics that are important here. I am thankful to an anonymous reviewer for raising this point.

(P1a) The mechanics of judgements in class C treat knowledge as if it is a mental state. [Empirical]

(P1b) The contents of judgements in class C reflect the structure of knowledge.

(P2) The contents of judgements about knowledge reflect the structure of knowledge only if the mechanics of those judgements also reflect the structure of knowledge. [Neurocognitive Parity]

(C1) The mechanics of judgements in class C reflect the structure of knowledge. [P1b and P2]

(P3) If the mechanics of some class of judgements (i) reflect the structure of knowledge and (ii) treat knowledge as if it is a mental state, then knowledge is a mental state for the cases that are the subject of those judgements. [Reflection]

(C2) For the cases of simple perceptual knowledge that are the subject of judgements in class C, knowledge is a mental state. [P1a, P3, and C1]

In this way, neurocognitive parity facilitates the straightforward inference from attributing knowledge like a mental state (in these cases) to knowledge being a mental state (in these cases). Since this style of argument is mostly in the setup, there isn't a ton to unpack here. (P1a) is just a straight-ahead expression of the empirical evidence discussed back in section 2—Both studies (Phillips et al. 2018; Bricker 2020) document a class of judgements about knowledge, the mechanics of which treat knowledge like a mental state, and the plausibility of these direct findings is bolstered by a wealth of indirect evidence indicating that knowledge representation is evolutionarily and developmentally more basic than belief representation. Similarly, (P2) is just a direct application of neurocognitive parity, and the move from (P1b) and (P2) to (C1) is an uncomplicated, core inference directly supported by the principle. Since much of this section has been devoted to motivating neurocognitive parity, I don't think I have anything to add here. Perhaps the move from (C1), (P1a), and (P3) to (C2) is a bit less transparent, but note that the sort of incorporation into our theories of knowledge expressed by (P3) is central to how I'm characterizing the reflection relation operative in (P1b). To say that some contents or mechanics reflect the structure of knowledge is just to say they tell us something we want our accounts of knowledge to accommodate. Thus, if some knowledge-reflecting judgement mechanics indicate that knowledge is a mental state, we can conclude that knowledge is in fact a mental state.

This then just leaves (P1b), which I think is the only part of this argument that requires additional support. The operative question here is whether we think the contents of the judgements made by participants, particularly in the EEG study (Bricker 2020),¹¹ reflect the structure of knowledge. I would submit that we have ample reason to think that they do. First, these are simple, ordinary, paradigmatic intuitive judgements about perceptual knowledge. This is clearly not a case in which “shifty patterns” of knowledge attribution (Gerken 2017, ch. 2) might lead us to question whether judgement content actually reflects

¹¹ I've opted to focus on my own study here, since it took care to report a wealth of relevant details re. whether participants were making judgements whose contents reflected the structure of knowledge. However, I do think that much of what I say also holds for the Phillips et al. (2018) study.

the nature of knowledge. Next, a number of different aspects of the study specifically indicate that its participants were reliable evaluators of the knowledge of others: (i) A pre-experiment procedure was used to ensure that, when completing knowledge attribution tasks, participants understood knowledge in a literal sense (2020, 5); (ii) participants with low judgement accuracy were excluded from the analysis, and mean error rates for participants included in the analysis were extremely low ($\leq .001$; 2020, 9); (iii) electrophysiological evidence, especially an ERP component known as the “late slow wave,” indicated that participants’ core ToM systems were indeed active during knowledge attribution, ruling out the possibility of a non-ToM heuristic response strategy (2020, 10-11). In short, these were by all accounts ordinary, reliable judgements about simple perceptual knowledge produced by competent knowledge evaluators. While the contents of these judgements aren’t all that epistemologically interesting, they nonetheless appear to be the kind that reflect the structure of knowledge.

For many, especially those of an experimentalist persuasion, this is likely sufficient. However, for some more traditional philosophers there may be a lingering worry—As these judgements are those of laypeople, not epistemological experts, perhaps their contents do not actually constitute evidence about the metaphysics of knowledge.¹² This worry, however, might be quieted by noting the nature of the neurocognitive mechanisms of interest here (§2). These are not obscure biases or heuristics that are likely to differ between epistemologists and laypeople. Instead, these mechanisms constitute a core neurocognitive system responsible for judgements about knowledge. Accordingly, it is likely that these same mechanisms guide both experts and laypeople in (at least) their simple, ordinary judgements about perceptual knowledge. While it is plausible that education in epistemology might inoculate one from certain cognitive biases, it is much less plausible that this education reconfigures the core neurocognitive architecture of one’s entire ToM system. This then allows us to understand that, when epistemologists make judgements about knowledge under similar conditions, it is quite likely that the mechanisms underlying these judgements also treat knowledge like a mental state. In short, in cases of both philosophers and competent laypeople, we can be quite confident that there is in fact a class of judgements about knowledge for which both (P1a) and (P1b) are true.

With this final element in place, we might understand how the argument I’ve presented here straightforwardly applies neurocognitive parity, allowing us to conclude that knowledge is (at least sometimes) a mental state on the basis of empirical facts about the mechanics underlying knowledge attribution—Given that judgement content in these cases reflects the structure of knowledge, it follows directly from neurocognitive parity that the mechanics underlying these judgements also reflect the structure of knowledge. As these mechanics clearly treat knowledge like a mental state, we can then conclude that knowledge is in fact a mental state in such cases.

5. Implications and Open Questions

Before concluding this paper, I want to talk a bit about the precise implications of the argument I’ve presented above, on which neurocognitive parity is applied to our current understanding of how the brain computes knowledge attributions. Crucially, I want to be especially sure that we don’t go overboard and draw conclusions that the present argument is not in a position to support. Accordingly, here I’ll discuss three of the most important implications of the argument I’ve presented, along with the most likely ways in which one

¹² See, e.g., the “expertise defense.” (Williamson 2011; Nado 2014)

might overlook open questions and take each of these implications a little too far: (1) Again, my argument only supports the conclusion that *perceptual* knowledge is *often* a mental state, and I want to resist the urge to make the more general conclusion that knowledge is always a mental state. (2) While my argument highlights a fundamental problem with accounts of knowledge on which knowing is conceptualized as a species of belief, we cannot conclude on this basis anything like knowledge being conceptually prior to belief. (3) The theoretical framework provided by neurocognitive parity, along with the emergence of a knowledge-centric Theory of Mind research program, offers important methodological implications regarding the use of the tools of psychology and neuroscience to address fundamental questions in epistemology. Nevertheless, it is not presently clear the extent to which this method might generalize, and we have reason to suspect that it might be limited to questions of knowledge and belief. Additionally, I also wish to be clear that (4) it isn't obvious that my argument undermines the "entailment thesis" that knowing that *p* entails believing that *p*. In what follows, I discuss each of these points in turn.

First, as I've frequently repeated from the beginning of this paper, neurocognitive parity alone does not at this time support the conclusion that all knowledge states are mental states. Due to the presently limited body of available cognitive and neurocognitive evidence, the argument I've presented here can only support the conclusion that perceptual knowledge is often a mental state in ordinary cases, and even then, it cannot rule out perceptual knowledge sometimes being a composite in certain other cases. To be clear, I do recognize that there is likely a strong impulse among many epistemologists to maintain that knowledge must display a unitary metaphysical composition—Regardless of whether knowledge might be a mental or a composite state, it certainly cannot be the case that some knowledge states are composites whereas other knowledge states are mental. Indeed, I acknowledge that, if one is committed to such an assumption, one might then conclude on the basis of the argument I've presented here that knowledge is a mental state in its own right, full stop. Moreover, there is likely also an impulse among some knowledge-first epistemologists to think that the burden of proof now falls squarely on those who think that knowledge is a composite. Until such time as there is any evidence that the brain attributes knowledge like a composite, given the body of evidence pointing in the opposite direction (see §2), we can assume that it always attributes knowledge like a mental state. However, while I'm not strictly unsympathetic to either of these ideas, it is more reasonable at this point to view all this as an open empirical question. As it is all too easy to imagine that at least some small proportion of knowledge attributions might be computed like composites, while nonetheless appearing to reflect the structure of knowledge, I'm unconvinced of the wisdom of just assuming a unitary metaphysics *a priori*. Ultimately, it remains to be seen whether knowledge from sources other than perception also constitute mental states, as well as whether perceptual knowledge is exclusively a mental state. While I am not aware of any case in which a dualistic mental-composite metaphysics of knowledge has been defended, or even examined, now might be the time for epistemologists to start acclimating to the idea.

Were the method I advocate for here to ultimately lead us to a dualistic mental-composite metaphysics for knowledge, this of course would bring with it a series of major epistemological commitments. Perhaps most significantly, it would mean that there is no unity in kind for knowledge states at the level of the states themselves. Instead, multiple classes of states, with distinct metaphysical compositions, would fall under the category of knowledge. While this might at first seem a bit radical, I would submit that this picture fits nicely into a Craig-style approach to knowledge (1990), which identifies the central, unifying force behind knowledge with the role it plays for the knowledge *attributor*—i.e. identifying good informants. Taking such an approach, we might perhaps argue that, as both mental and composite states can play this central role in identifying good informants, it is only

reasonable to expect that the mechanics of our knowledge-attributing systems might be geared towards picking out both kinds of states. This sort of argument would echo a similar move from Pritchard, who has maintained that distinct anti-luck and ability conditions on knowledge states emerge from its singular role in identifying good informants (2012, §4). For those who insist that there must be some attributor-independent unity to knowledge states themselves, perhaps due to some special functional role played by knowledge for the *knower*, this would likely be an unacceptable cost of my account. However, at least for those who think this way on the basis of taking knowledge to be a natural kind (e.g. Kornblith 2002), the very idea that we can make inferences about the structure of knowledge from judgement content—a basic assumption of much of analytic epistemology that I explicitly commit to here—will be objectionable.

Additionally, were knowledge to display a dualistic metaphysics, which brought together otherwise disunified kinds, this would likely support a view on which the attribution or representation of knowledge states is in some sense metaphysically prior to knowledge states themselves. Such a view might take the form of social constructivism, on which knowledge is constituted by what ideal knowledge-attributing agents would judge to be knowledge, or perhaps alternatively a non-constructivist conferralism (see Sveinsdottir 2008), on which the property of knowing is conferred by (at least some) actual judgements of actual knowledge-attributing agents. While I don't think this sort of anti-realist picture would be an especially negative consequence for my account—after all, the phylogeny of knowledge-attributing capacities lets us tell a very natural story about how such capacities might have emerged in the absence of any attributor-independent knowledge states—I recognize that some might view this consequence less favorably. To the extent that this sort of anti-realist metaphysics of knowledge is undesirable, I must then concede that this would be a limitation of my account. However, again, it is important to reiterate that all this is purely speculative—we presently have no empirical evidence indicating that knowledge displays any sort of dualistic mental-composite or associated anti-realist metaphysics.

Conversely, one of the major implications that quite definitively follows from my argument is that accounts of knowledge that take knowledge to be a composite of belief with non-mental components—which is to say, most of them (see footnote 2)—are flatly mistaken. As we've observed here, at least some cases of ordinary perceptual knowledge are not species of belief, but instead mental states in their own right, and traditional belief-centric accounts are unable to accommodate this result. However, it is important that we don't go overboard with our conclusions here. As a number of commentators on Nagel's argument have pointed out, this style of argument from knowledge attribution does not support the conclusion that knowledge is conceptually prior to belief (Butterfill 2013; Roessler 2013; Rysiew 2013).

Moving on, perhaps the most significant long-term implications here are methodological. As I've argued above, the principle of neurocognitive parity can serve as an invaluable tool for inferring the nature of knowledge from the nature of knowledge attributions, at least in cases where we would already be comfortable inferring the nature of knowledge from the content of knowledge attributions. If then coupled with cognitive and neurocognitive studies of knowledge attribution, this suggests a powerful new approach to theorizing about the structure of knowledge, on which epistemologists might conduct such studies with the specific aim of answering outstanding epistemological questions. While the studies discussed here specifically aimed to address whether knowledge is attributed like a mental or composite state, there is no reason why new studies couldn't target other questions as well. In this way, we might understand epistemological questions about the nature of knowledge to be directly continuous with empirical questions about the neurocognitive

mechanisms of knowledge attribution, with the empirical techniques of cognitive (neuro)science themselves constituting genuine ways of doing epistemology.

However, here again it is important we exercise some moderation. First, far from fulfilling some grand neurophilosophical ambition in the vein of Churchland (e.g. 1987), this method would still heavily rely on both traditional epistemology and standard experimental philosophy. Not only does the application of neurocognitive parity require that we can be confident in the evidential potential of judgement content, a task that will inevitably require more established epistemological methods, but the cognitive neuroscience of knowledge attribution itself requires significant contributions from both traditional and experimental epistemology. Conceptualizing knowledge states is a task for traditional epistemology (see especially Bricker 2020, §1.1; Phillips et al. 2020, §2), and ensuring that participants are reliable knowledge attributors falls directly under the purview of attribution reports and experimental epistemology (see Bricker 2020, §4.5). Moreover, beyond its reliance on more conventional methods, we have a second reason to temper any excess enthusiasm about neurocognitive parity fueling the next big advancement in philosophical methodology: Inferring the structure of knowledge from the mechanics of knowledge attribution only works because, as discussed in section two, knowledge attribution relies on specific neural mechanisms that can be isolated experimentally. However, this likely isn't the case for many philosophically interesting concepts outside or even within epistemology. While this is of course ultimately an empirical question, and future developments could very well prove my caution unnecessary, it is important that we recognize that anything like an immediate, full-scale generalization of the method suggested here is unlikely to be successful.

Finally, I want to briefly mention that I'm not confident that this argument currently gives us reason to reject the entailment thesis that knowing that *p* entails believing that *p*. While the available neurocognitive evidence indicates that belief attribution is not a component of (at least certain) knowledge attributions, this evidence seems to come from cases that depict knowledge only when they also depict belief (for more on this, see Bricker 2020, §4.5). In contrast with other evidence from standard experimental epistemology (e.g. Murry et al. 2013; Myers-Schulz and Schwitzgebel 2013), we presently have no direct evidence relating to the mechanics of knowledge attribution in cases where participants might withhold belief attribution.

Lest we end things too dourly, however, I do want to reiterate that there are multiple quite significant implications to be drawn from the argument I've presented here. Not only might we confidently conclude that knowledge is at least sometimes a mental state, contra an attitude widely held in philosophy, but we might do so via an innovative new approach that pairs traditional methods with epistemology-driven cognitive neuroscience. While the exact limits of this approach are not yet apparent, I am confident that this is by no means the end of this sort of partnership between cognitive neuroscience and epistemology.

6. Conclusion

Here my goal was to advance a new line of support for the thesis that knowledge is mental state. Central to this argument was the principle of neurocognitive parity, on which judgement content reflects the structure of knowledge only if the neural and cognitive mechanics responsible for that content also reflect the structure of knowledge. Taken together with recent findings from cognitive neuroscience, this indicates that there are cases in which being in some mental state is sufficient for knowing that *p*. However, as the available (direct) cognitive and neurocognitive evidence is presently quite limited, I am hesitant to conclude on this basis that being in this particular mental state is also necessary for knowledge.

Nevertheless, it is at this point clear that the widely accepted composite view of knowledge simply doesn't capture the relationship between knowing and the mental. Knowledge is—at least sometimes—a mental state.

References

- Butterfill, S. (2013). What Does Knowledge Explain? Commentary on Jennifer Nagel, "Knowledge as a Mental State." In *Oxford Studies in Epistemology Volume 4*. Oxford: Oxford University Press.
- Bricker, A. (2020). The neural and cognitive mechanisms of knowledge attribution: An EEG study. *Cognition*, 203, 104412–.
- Carter, J., Gordon, E., & Jarvis, B. (eds.) (2017). *Knowledge First: Approaches in Epistemology and Mind*. Oxford: Oxford University Press.
- Churchland, P. (1987). Epistemology in the Age of Neuroscience. *The Journal of Philosophy*, 84(10), 544–553.
- Craig, E. (1990). *Knowledge and the State of Nature: An Essay in Conceptual*. Oxford: Oxford University Press.
- Evans, J. (2008). Dual-Processing Accounts of Reasoning, Judgment, and Social Cognition. *Annual Review of Psychology*, 59(1), 255–278.
- Gendler, T., & Hawthorne, J. (eds.) (2013). *Oxford studies in epistemology Volume 4*. Oxford: Oxford University Press.
- Gettier, E. (1963). Is Justified True Belief Knowledge? *Analysis*, 23(6), 121–123.
- Gerken, M. (2012). On the Cognitive Bases of Knowledge Ascriptions. In *Knowledge Ascriptions*. Oxford: Oxford University Press.
- Gerken, M. (2013). Epistemic Focal Bias. *Australasian Journal of Philosophy*, 91(1), 41–61.
- Gerken, M. (2017). *On Folk Epistemology: How we Think and Talk about Knowledge*. Oxford: Oxford University Press.
- Gigerenzer, G., & Gaissmaier, W. (2011). Heuristic Decision Making. *Annual Review of Psychology*, 62(1), 451–482.
- Goldman, A. (1986). *Epistemology and cognition*. Harvard University Press.
- Greenough, P., Pritchard, D. (eds.) (2009). *Williamson on knowledge*. Oxford: Oxford University Press.
- Hartwright, C. E., Apperly, I. A. & Hansen, P. C. (2012). Multiple roles for executive control in belief–desire reasoning: Distinct neural networks are recruited for self perspective inhibition and complexity of reasoning. *NeuroImage*, 61(4).
- Hartwright, C. E., Apperly, I. A. & Hansen, P. C. (2015). The special case of self-perspective inhibition in mental, but not non-mental, representation. *Neuropsychologia*, 67.
- Hawthorne, J. (2004). *Knowledge and lotteries*. Oxford: Clarendon.

- Heleven, E., & Van Overwalle, F. (2018). The neural basis of representing others' inner states. *Current Opinion in Psychology*, 23, 98-103.
- Hogrefe, G.-J., Wimmer, H., & Perner, J. (1986). Ignorance versus False Belief: A Developmental Lag in Attribution of Epistemic States. *Child Development*, 57(3), 567.
- 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/>>.
- Kaminski, J., Call, J., & Tomasello, M. (2008). Chimpanzees know what others know, but not what they believe. *Cognition*, 109(2), 224–234.
- Kornblith, H. (2002). *Knowledge and its place in nature*. Oxford: Clarendon.
- McGlynn, A. (2014). *Knowledge First?* Palgrave Macmillan.
- Murray, D., Sytsma, J., & Livengood, J. (2013). God knows (but does God believe?). *Philosophical Studies*, 166(1), 83–107.
- Myers-Schulz, B., & Schwitzgebel, E. (2013). Knowing That P without Believing That P. *Noûs*, 47(2), 371–384.
- Nado, J. (2014). Philosophical Expertise. *Philosophy Compass*, 9(9), 631–641.
- 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* 4:275-310.
- Ortmann, A., Gigerenzer, G., Borges, B., & Goldstein, D. (2008). Chapter 107 The Recognition Heuristic: A Fast and Frugal Way to Investment Choice? In *Handbook of Experimental Economics Results* (Vol. 1, pp. 993–1003).
- Özdem, C., Brass, M., Schippers, A., Van Der Cruyssen, L., & Van Overwalle, F. (2019). The neural representation of mental beliefs held by two agents. *Cognitive, Affective, & Behavioral Neuroscience*, 19(6).
- Phillips, J., Knobe, J., Strickland, B., Armory, P., & Cushman, F. (2018). Knowledge before Belief: Response-times indicate evaluations of knowledge prior to belief. *Proceedings of the Fortieth Annual Conference of the Cognitive Science Society*.
- Phillips, J., Buckwalter, W., Cushman, F., Friedman, O., Martin, A., Turri, J., Santos, L., and Knobe, J. (2020). Knowledge before Belief. *Behavioral and Brain Sciences*. 1–37.
- Pritchard, D. (2005). *Epistemic Luck*. Oxford: Oxford University Press.
- Pritchard, D. (2012). Anti-Luck Virtue Epistemology. *The Journal of Philosophy*, 109(3), 247–279.
- Roessler, J. (2013). Knowledge, Causal Explanation, and Teleology. In *Oxford Studies in Epistemology Volume 4*. Oxford: Oxford University Press.

- Rysiew, P. (2013). Is Knowledge a Non-Composite Mental State? In *Oxford Studies in Epistemology Volume 4*. Oxford: Oxford University Press.
- Samson, D., Apperly, I. A., Chiavarino, C. & Humphreys, G. W. (2004). Left temporoparietal junction is necessary for representing someone else's belief. *Nature Neuroscience*, 7(5).
- Samson, D., Apperly, I. A., Kathirgamanathan, U. & Humphreys, G. W. (2005). Seeing it my way: A case of a selective deficit in inhibiting self-perspective. *Brain*, 128(5)
- Scheibehenne, B., & Bröder, A. (2007). Predicting Wimbledon 2005 tennis results by mere player name recognition. *International Journal of Forecasting*, 23(3), 415–426.
- Schuwerk, T., Döhnel, K., Sodian, B., Keck, I., Rupprecht, R., & Sommer, M. (2014). Functional activity and effective connectivity of the posterior medial prefrontal cortex during processing of incongruent mental states. *Human Brain Mapping*, 35(7).
- Serwe, S., & Frings, C. (2006). Who will win Wimbledon? The recognition heuristic in predicting sports events. *Journal of Behavioral Decision Making*, 19(4), 321–332.
- Sosa, E. (1999). How to Defeat Opposition to Moore. *Philosophical Perspectives* 13, 141–153.
- Sosa, E. (2007). *A Virtue Epistemology: Apt Belief and Reflective Knowledge*. Oxford: Oxford University Press
- Sveinsdottir, A. (2008). Essentiality Conferred. *Philosophical Studies*, 140(1), 135–148.
- 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).
- Westra, E., & Nagel, J. (2021). “Mindreading in conversation.” *Cognition*, 210, 104618–104618.
- Williamson, T. (2000). *Knowledge and its limits*. Oxford: Oxford University Press.
- Williamson, T. (2005). CONTEXTUALISM, SUBJECT-SENSITIVE INVARIANTISM AND KNOWLEDGE OF KNOWLEDGE. *The Philosophical Quarterly*, 55(219), 213–235.
- Williamson, T. (2011). “PHILOSOPHICAL EXPERTISE AND THE BURDEN OF PROOF.” *Metaphilosophy*, 42(3), 215–229.