

## Knowledge as a mental state

Forthcoming in *Oxford Studies in Epistemology*

**ABSTRACT:** In the philosophical literature on mental states, the paradigmatic examples of mental states are beliefs, desires, intentions, and phenomenal states such as being in pain. The corresponding list in the psychological literature on mental state attribution includes one further member: the state of knowledge. This article examines the reasons why developmental, comparative and social psychologists have classified knowledge as a mental state, while most recent philosophers—with the notable exception of Timothy Williamson—have not. The disagreement is traced back to a difference in how each side understands the relationship between the concepts of knowledge and belief, concepts which are understood in both disciplines to be closely linked. Psychologists and philosophers other than Williamson have generally have disagreed about which of the pair is prior and which is derivative. The rival claims of priority are examined both in the light of philosophical arguments by Williamson and others, and in the light of empirical work on mental state attribution.

One striking feature of research into mental state ascription or ‘mindreading’ is the extent to which it has involved cooperation between psychology and philosophy. Contemporary empirical work on mindreading is often traced back to a 1978 target article in comparative psychology, an article which raised the question of whether chimpanzees attribute mental states to themselves and others (Premack & Woodruff, 1978). Three of the most influential responses published with that article were written by philosophers, each of whom drew attention to the importance of probing the capacity to represent states of ignorance and false belief (Bennett, 1978; Dennett, 1978; Harman, 1978). In the decades that followed, philosophers made significant contributions to the theoretical frameworks guiding empirical research into mental state ascription (e.g. Goldman, 2006; Nichols & Stich, 2003), while psychologists made significant contributions to philosophical debates about the nature of our access to our own mental states and those of others (e.g. Gopnik & Wellman, 1992; Perner, 1991). Major progress has also been made in work co-authored by philosophers and psychologists (e. g. Apperly & Butterfill, 2009; Gallese & Goldman, 1998). In general, disciplinary boundaries between philosophers and psychologists working on mindreading do not seem to put these two groups in opposition to each other. However, there is one question on which most philosophers and psychologists appear to be divided, and it is arguably a significant question in the theory of mental state ascription. This is the question of whether knowledge is a mental state.

On the psychology side of the line, the common answer to that question is positive. Developmental, social, cognitive, and comparative psychologists explicitly and without hesitation classify knowledge as a mental state, alongside states of belief, desire, intention, perception and feeling (Apperly, 2011; Baron-Cohen et al., 1994; Call & Tomasello, 2008; De Villiers, 2007; Epley & Waytz, 2009; Heyes, 1998; Keysar, 2007; Lin, Keysar, & Epley, 2010; Premack & Woodruff, 1978;

Saxe, 2005; Sodian, Thoermer, & Dietrich, 2006; Wellman & Liu, 2004). On the philosophy side of the line, the answer is almost unanimously negative. Knowledge is conspicuously absent from the lists of mental states presented by many prominent contemporary philosophers of mind (Davies & Stone, 2001; Goldman, 2006; Gordon, 1986; Stich & Ravenscroft, 1994). It is not simply by oversight that philosophers have failed to include knowledge in their lists; there are systematic reasons for the omission, to be discussed in what follows. In addition, philosophers have presented various explicit arguments aimed at establishing that knowledge is not a mental state (Fricker, 2009; Magnus & Cohen, 2003; Molyneux, 2007).

Although the current climate in philosophy is largely hostile to the idea that knowledge is a mental state, there is no historical incompatibility between this idea and a philosophical mindset: philosophers as different as Plato and Locke classified knowledge this way. In present day philosophy, there are some philosophers of mind who continue to list knowledge among the mental states (e.g. Carruthers, 2009), and there is one sustained defense of the recognition of knowledge as a mental state, in the work of Timothy Williamson (1995; 2000; 2009). This position has met considerable resistance from other philosophers. In responding to Williamson's identification of knowledge as mental state, Anthony Brueckner is no doubt reporting the common sentiment among philosophers when he observes that 'to many this is a highly counterintuitive claim' (Brueckner, 2002, 197).

When disciplines seem to disagree, it is natural to wonder whether the dispute is merely terminological. In this case one might suspect that the two sides currently mean something different by 'mental state' or by 'knowledge'. Certainly there are a number of words whose standard technical meanings are quite different in philosophy and psychology, sometimes for theoretical reasons and sometimes just because a different sense of the word is more commonly in play in a given discipline (e.g. 'normative'; 'modal'). Section 1 of this paper argues that, notwithstanding certain relatively superficial differences in usage, the relevant branches of the two disciplines generally have shared targets in their sights both when they speak of mental states and when they speak of knowledge. If the debate is not merely verbal but substantive, at least one side is mistaken: this paper argues that the mainstream philosophers are the ones who have gone wrong.

Section 2 discusses philosophical arguments on the matter and section 3 discusses empirical work. The difference between mainstream philosophers and psychologists is diagnosed as arising from a difference in the way the two sides see the relationship between the concepts of knowledge and belief. All parties to the dispute see these concepts as closely linked, and agree that

belief is a mental state. Where they disagree is on the question of whether knowledge or belief is a more basic concept: mainstream contemporary philosophers take the concept of belief as their starting point, and see the concept of knowledge as something more complex, generated by adding various non-mental conditions to the uncontroversially mental state of belief. Williamson joins the psychologists in taking the concept of knowledge to be more basic: like most psychologists of mental state ascription, he sees the concept of belief as something derived from the concept of knowledge, where knowledge is from the start seen as a mental state in its own right.

A central motivation for the mainstream philosophical approach appears to be the thought that our understanding of action is fundamentally dependent on belief attribution; in this view, the belief component of knowledge is what really matters to our natural understanding of what others do, and it is uneconomical to admit knowledge-based explanations of action alongside belief-based ones. Advocates of the mainstream approach also appeal to a metaphysical intuition that mental states must be a function of strictly local conditions in the agent; observing that the truth of a known proposition is not typically something contained within the agent, they conclude that it would be metaphysically odd to characterize knowledge as a mental state. Against the first motivation for the mainstream view, it is not assumed in psychology that our understanding of other agents is ultimately belief-based; indeed the capacity to attribute mere belief to others seems to be considerably harder to master and deploy than the capacity to attribute knowledge. This paper examines a range of work in developmental and comparative psychology supporting the conclusion that an ability to track what others would know seems to be the precondition, rather than the product, of an ability to track what they would believe. Mature social cognition is also relevant. By the time they reach maturity, human beings are able to recognize both knowledge and belief intuitively, but it seems the availability of belief ascription does not make knowledge ascription redundant in our understanding of the actions of others.

On the metaphysical point, there is reason to resist the intuitively alluring thought that mental states must be localized within the agent: this sense of localization is arguably a natural illusion, like the parallel illusion in 'naïve physics' which makes us feel that the motions of inanimate objects are determined by their own inner 'impetus' and local interactions with what they contact, even though our intuitive calculations of those motions themselves make use of information about non-local factors such as gravitation. The intuitive sense that some human agent knows something is produced by intuitive mechanisms that make calculations about the target person and his relationship to the environment; for example, mechanisms that automatically track the target's

direction of gaze and note which objects lie in his visual field (Samson, Apperly, Braithwaite, Andrews, & Scott, 2010). Because these calculations are intuitive, we do not have explicit awareness of how they are made, and can have the intuitive sense that the target's resulting mental state is 'internal' to him, like the intuitively felt 'internal impetus' of a falling object. Our actual use of the attributed mental state of knowledge in predicting and explaining the target's subsequent actions does not however depend on that illusory impression: just as our intelligent interactions with falling objects incorporate our real capacity to take gravitation into account, so also our intelligent interactions with other persons incorporate our real capacity to take their relationships with the environment into account. The fact that the state of knowing incorporates a relationship with the environment does not disqualify it from counting as a state which is fundamental to our intuitive understanding of other intelligent beings.

Evidence from social, developmental and comparative psychology seems to support the view that knowledge is naturally seen as a mental state in its own right, and not as a composite of belief and non-mental factors. However, one might grant that knowledge is naturally seen this way and still wonder whether knowledge really *is* a mental state. The final section of the paper discusses the relationship between our evidence about the way we are naturally inclined to see knowledge and our conclusions about the kind of state knowledge really is.

### **1. What the two disciplines mean by 'mental states' and 'knowledge'**

One thing philosophy and psychology have in common is that neither discipline has produced a comprehensive and settled account of the nature of mental states or of knowledge. Ongoing controversy over these topics within each discipline makes it difficult to establish conclusively that philosophers and psychologists generally aim to speak of the same thing when they use the expression 'mental state' (or the word 'knowledge'). As each discipline produces and criticizes various successive 'mental state' theories, for example, it is a live question whether there is a single phenomenon that researchers on both sides of the disciplinary divide are trying to explain, or whether we and our colleagues across campus are stalking different prey. Because active controversies could mask subtle but genuine differences between the disciplines, this section will not hope to demonstrate that philosophers and psychologists all have precisely the same thing in mind for each of our key expressions, but will instead try to show extensive similarities in the way 'mental state' and 'know-

ledge' are understood in the two disciplines, to shift the burden of proof onto those who claim that the general interdisciplinary disagreement over knowledge being a mental state is merely verbal.

To give some rough indication of what they mean by 'mental state', both psychologists and philosophers often provide lists of examples. Setting aside the inclusion or omission of knowledge, these lists are strikingly similar across the disciplines. Psychologists and philosophers are widely agreed that the following are naturally classified as mental states: beliefs, desires, intentions, emotions, perceptions, and sensations such as pain and hunger. There is also wide agreement in both disciplines about the relationship between mental states and agency: we take<sup>1</sup> a person's actions (as opposed to mere reflexes or automatic functions such as digestion) to be dependent on<sup>2</sup> his or her epistemic and motivational mental states. There are well established systematic relationships between these mental states: for example, people generally intend to do what they believe will satisfy their desires. Given the universality of these relationships among mental states, a capacity to register the mental states of others—for example, what they believe, desire and intend—enables us to explain and predict what they will do. We do not simply expect old patterns of overt behavior to be repeated; because we can grasp the underlying causes of those patterns, we can anticipate how an agent will act and interact with us in novel circumstances. This social understanding may be imperfect in various ways, but it is strong enough to support successful communication, competition and cooperation across a wide range of social situations. Of course there are a great many theories about the nature of mental states both in psychology and in philosophy, and some of these theories may have discipline-specific or idiosyncratic aims, but in broad brushstrokes recent philosophical and psychological theories about mental states generally seem to aim at capturing similar phenomena.

---

<sup>1</sup> According to some philosophers, our natural inclinations to see human actions as caused by mental states are fundamentally mistaken: eliminativists like Churchland argue that the basic terms of folk psychology do not refer to anything with real causal powers (Churchland, 1981). As this is now very much a minority position within philosophy, and even Churchland himself has come to admit the reality of some mental states, eliminativism will be set aside in what follows, and the reality of mental states such as belief and desire will be assumed as a starting point. A similar initial commitment to the reality of mental states is common in psychological work (see e.g. Apperly, 2011, p.2). One can of course admit the fallibility of folk psychology, and agree that the nature of those states cannot be read straight off our natural patterns of mental state ascription. Furthermore, even those who are still inclined towards an eliminativist reading of much mental state discourse may still find some interest in the question of whether the (apparent) state of knowledge is a part of that domain, or whether knowledge ascription is a form of mental state ascription.

<sup>2</sup> There are various views about the character of this dependence: many regard it as straightforwardly causal, but on some views (e.g. Dancy, 2000) mental states are characterized as necessary background conditions rather than causes of action. For ease of exposition I will describe the dependence as causal in what follows, but I believe that much of what I say could be reworded to fit views such as Dancy's.

The question of whether philosophers and psychologists mean the same thing by 'knowledge' is more difficult. One of the characteristics of knowledge to which philosophers are particularly attentive is the truth condition: it is almost universally accepted in philosophy that knowing, unlike mere believing, is an attitude that one can have only to truths. One strand of evidence for this is linguistic: 'know' is a factive verb, a verb that licenses an inference to its complement clause. From the truth of a proposition of the form 'S knows that  $p$ ' it would follow that  $p$  is the case. Nonfactive verbs such as 'believes' do not license such inferences. If one follows philosophers such as Williamson in extending the terminology of linguistics over to epistemology, one can describe knowing itself as a factive propositional attitude, in contrast to nonfactive propositional attitudes such as thinking, hoping and conjecturing (Williamson, 2000). Unlike the identification of knowledge as a mental state, the identification of knowledge as a factive state is not considered particularly controversial in philosophy: even the rare philosopher who doubts the strength of the linguistic evidence for the factivity of the verb 'to know' does not see this as a reason to challenge the classification of knowledge itself as a factive state (Hazlett, 2010).

Psychologists might at first seem not to take knowledge to be a factive state. For example, in the literature on calibration or the relationship between confidence and accuracy, some papers discuss the 'accuracy of knowledge' in various domains, by which the authors mean the percentage of true out of total answers volunteered (e.g. Perfect, Watson, & Wagstaff, 1993). Epistemologists might observe that the accuracy of *knowledge*, strictly speaking, would always have to be 100%; any false answer volunteered would be mere belief at best, and fail to count as knowledge in the first place. But the charitable reader can readily make sense of the calibration literature by penciling in scare quotes where appropriate, for example by reading various calibration studies as probing something more like the accuracy of what the research subjects "know" (or perhaps take themselves to know). When the line between knowledge and what is taken for knowledge is particularly important and there is a real risk of confusion, psychologists do take more care in marking the distinction. In calibration research as elsewhere it is understood, and sometimes even underscored, that strictly speaking what is known must be true (e.g. Koriat, 1995).

In the branches of psychology most directly engaged with mental state attribution, there is greater self-consciousness about the relationship between truth and knowledge, both because those who are examining the linguistic expression of mental state attribution take the factivity of the mental state verb 'know' to be a core component of its meaning (e.g. Abbeduto & Rosenberg, 1985), and because much attention is paid to our competence for recognizing false beliefs, whether

this competence is expressed linguistically or in other forms of behavior. Indeed, the threat of an interdisciplinary difference sometimes appears to loom larger in the treatment of falsity than factivity: in articles about the relative difficulty of attributing knowledge and false belief, the terms 'belief' and 'false belief' occasionally seem to be used interchangeably (e.g. in Kaminski, Call, & Tomasello, 2008). But this pattern of usage is also better explained by its expediency than by any interdisciplinary disagreement about the possibility of true belief. In experimental studies of the capacity to discriminate knowledge from mere belief, the easiest way to generate a situation involving a belief falling short of knowledge is usually to devise a situation in which the observed agent's belief is false. Because some of the clearest tests of the capacity to represent belief (as contrasted to knowledge) are tests involving false belief, psychologists of mental state ascription sometimes gloss over the distinction between belief recognition and false belief recognition.

This occasional elision in usage should not be taken to suggest any positive commitment to the thought that beliefs as such must be false, or to the thought that all true beliefs would have to count as knowledge. The relatively subtle capacity to distinguish knowledge from mere true belief (for example, belief generated by lucky guessing) emerges later than the capacity to distinguish knowledge from false belief: true belief that falls short of knowledge is surprisingly hard to ascribe. After they are capable of recognizing false belief, children pass through a stage in which they tend to describe a target as knowing if his action leads to successful performance—say selecting the box with the treat in it— even if the target had no evidential basis for his choice, for example because he was absent when the treat was placed in the box (Miscione, Marvin, O'Brien, & Greenberg, 1978). The capacity to recognize the difference between knowledge and accidentally true belief is often ignored in studies of young children and nonhuman primates who fail to make even the more rudimentary distinction between knowledge and false belief.<sup>3</sup> However, the more advanced capacity has

---

<sup>3</sup> One might complain that it is still controversial for psychologists to describe the more rudimentary capacity as distinguishing between knowledge and false belief, as opposed to true belief and false belief; perhaps the cases of accidentally true beliefs have special peculiarities that make it harder to handle them appropriately. But it is not in general true that false beliefs are harder to attribute than true ones; for example, children who are capable of recognizing false belief pass through a stage in which they explicitly represent others who in fact simply lack knowledge as holding false, rather than true, beliefs about propositions known to the child (Ruffman, 1996). Furthermore, the prototypical cases of knowing that the child recognizes (in particular, *seeing* that something is the case), are the same as those we recognize as adults. Although the adult has a deeper understanding of knowledge, insofar as the child's early use of 'know' is typically a response to what we also recognize as knowledge, she should charitably be seen as referring to knowledge, just as the ancient Romans are appropriately seen as having referred to gold with their word "aurum", notwithstanding their failure to recognize it as the element with atomic number 79, and notwithstanding their relatively limited accuracy in sorting gold from its look-alikes. Thanks to Hagit Benbaji, Tamar Gendler and Michael Thompson for pressing me on this point.

also been studied heavily enough that there is no reason to think that psychologists of mental state ascription are generally oblivious to the distinction between knowledge and mere true belief (e.g. Johnson & Wellman, 1980; Moore, Bryant, & Furrow, 1989; Pillow, Hill, Boyce, & Stein, 2000; Sodian & Wimmer, 1987).

A last reason to worry that psychologists may have something different in mind when they speak of 'knowledge' is that they sometimes offer what seem to be the sort of analyses of knowledge that would meet considerable resistance in philosophy departments. For example, Bartsch and Wellman offer the following observation: 'Adults use *know* to refer to a belief that is felt to be justified, assumed to be true, or that enjoys markedly higher conviction than one described by *think*' (1995, 40). Meanwhile, according to Miller, Hardin and Montgomery, 'Young children may appreciate some of the criteria for "know" (a true belief for which there is evidence) without fully understanding all of the criteria (a belief that is not only true but held with certainty)' (2003, 350). Some psychologists give even more explicit endorsement of theories of knowledge rejected in philosophy since Gettier (1963); for example, by claiming directly that 'knowledge means justified true belief' (Reber & Unkelbach, 2010, 570). Others amend the classical analysis, but not in a manner that contemporary philosophers are inclined to accept as fully satisfactory, for example by simply adding a causal condition to the conditions of truth, justification and belief (Dienes & Perner, 1999, 739; Perner, 1991, 304).

Because the analysis-of-knowledge project commanded the attention of philosophers for decades after Gettier, counterexamples to all of these analyses will spring readily to mind for contemporary epistemologists. If there were evidence that psychologists wanted to insist on their analyses in the face of these counterexamples, one might worry about a real difference in what is meant by 'knowledge' in the two disciplines. But an examination of the psychologists' grounds for their general assertions about knowledge does not reveal any unorthodox responses to the currently standard epistemological counterexamples; rather, the relevant peculiar cases from the epistemology literature (cases involving deviant causal chains and the like) have not been the targets of theoretical attention in psychology, no doubt in part because such cases are ecologically rare. And, despite some perhaps overconfident phrasing—like the claim to have captured 'all of the criteria' for knowledge—the general claims in the last paragraph are not typically treated even by the psychologists who advance them as definitive. In practice psychologists do not for example insist that a mental state pass some test for 'markedly higher conviction' or a special feeling of certainty before they are willing to classify it as knowledge. Everyday cases of perceptual



observation, for example, are in standard practice taken to issue in knowledge without any prerequisite that they produce a particularly high level of confidence. So although the just-quoted general claims about knowledge may appear to philosophers as questionable analyses, they figure in their original contexts as rough empirical generalizations, typically aiming to deliver a quick outline of what it is that paradigmatic instances of knowing have in common.

Reassuringly enough, lists of paradigm cases of knowledge are the same in the psychology of mental state ascription as they are in contemporary epistemology: perceptual observation in good conditions, sound inference, memory, and appropriate cases of testimony.<sup>4</sup> If psychologists sometimes offer a questionable general recipe for what unites these cases, they are doing something that is often done in philosophy as well. When a philosopher advances what he describes as an analysis of knowledge, and this analysis is confronted with what we would ordinarily call an effective counterexample, it is more charitable to see the philosopher as having said something inaccurate about knowledge, than to see him as having attempted to introduce a new sense of the word 'knowledge'. The same charity should be extended to psychologists: except in cases where for example it is clear that they mean to speak of 'what is commonly taken to be knowledge' rather than knowledge itself, it is only fair to take their talk of knowledge at face value, as anchored in the same paradigmatic examples that drive philosophical work on knowledge, and as referring to a shared topic of pre-theoretical interest.

If there are any underlying terminological differences between philosophy and psychology in their ordinary use of 'knowledge' and 'mental state', these possible differences do not seem to be large enough to explain the starkly opposing positions generally taken in the two disciplines on the question of whether knowledge is a mental state. The next two sections examine the substance of what seems to be a substantive disagreement.

**2. Philosophical arguments about whether knowledge is a mental state.** Most contemporary philosophers deny that knowledge is itself a mental state, but are happy to grant that knowledge

---

<sup>4</sup> Philosophical skeptics have of course raised worries about whether knowledge is ever actually instantiated in any of these ways, but even they would not deny that these are the types of cases that people commonly take to be instances of knowing (Unger, 1971). Most contemporary epistemologists are not skeptics, and this paper will not address the challenge of skepticism. However, even skeptics who doubt the propriety of our ordinary ascriptions of knowledge may still have some interest in the question of whether what we ordinarily take to be knowledge is or is not naturally seen as a type of mental state.

incorporates a mental state, namely the state of belief.<sup>5</sup> On the currently standard philosophical view, in order to know a proposition *p*, an agent must not only have the mental state of believing that *p*, but various further independent conditions must also be met: *p* must be true, the agent's belief in *p* must be justified or well-founded, and so forth. At least some of these further conditions—most importantly, the truth of *p*, where *p* is some proposition about the external world<sup>6</sup>—would clearly be non-mental conditions.

States and events that are composites of the mental and non-mental are not naturally classified as mental: homicide, for example, is not a mental event. Although every homicide incorporates the mental condition of an intention to kill (indeed this mental condition, *mens rea*, is the part of the essence of homicide that distinguishes it from involuntary manslaughter), an event cannot be a homicide without also incorporating the non-mental condition of someone's actual death. The inclusion of this non-mental conjunct disqualifies the event considered as a whole from counting as purely mental. Similarly, on the currently orthodox philosophical view, because it is thought to include the non-mental component of the truth of propositions about external reality alongside the mental component of belief, the state of knowledge is considered composite rather than purely mental.

Enormous effort has been directed towards developing an account of the composition of this composite state. Epistemologists since Gettier have struggled hard to produce a closed list of factors that would need to be added to true belief (or relationships in which truth and belief would need to figure) in order to constitute the state of knowledge, evidently without success (for surveys, see Plantinga, 1993; Shope, 1983; Steup, 2008). Increasingly complex analyses have been crafted to sidestep intuitive counterexamples to earlier analyses, but new counterexamples have steadily emerged to confront the new analyses (and perhaps not only steadily but inevitably, as argued in Zagzebski, 1994). Other, simpler efforts at the analysis of knowledge have collapsed into circularity, managing to analyze knowledge only by adding to true belief a condition which

---

<sup>5</sup> Some philosophers add further mental conditions, maintaining that the mental state or psychological component of knowledge is not simply belief, but particularly confident belief, or perhaps justified belief, where 'justified' is understood in mental terms; these elaborations of the standard view will be discussed in due course. A few have argued that the state of knowledge positively excludes the state of belief; this type of position is now rarely held, and will not be discussed here (for criticism of such views, see Williamson, 2000, pp. 42-44).

<sup>6</sup> Some of the propositions we might be said to know or believe are themselves about the domain of the mental; strictly for ease of exposition I am setting those cases aside in what follows. Of course the truth of a proposition about the mental is not itself a mental condition, but it is easier to see that the composite view of knowledge involves an element that is entirely independent of the mental realm if we focus on knowledge of propositions about the external world.

ultimately incorporates knowledge itself. Certain efforts to explain knowledge in terms of ‘well-foundedness’, in particular, look problematic in this regard, as do appeals to conditions such as epistemic adroitness or competence (e.g. Sosa, 2007; Turri, 2011), insofar as it is hard to see how to understand the relevant concepts of adroitness or competence without appeal to the notion of knowing how to do something (cf. Stanley, 2011; Stanley & Williamson, 2001).

Although there is no simple formula for the ascent from belief to knowledge, the intuitive recognition of knowledge is supposed to start with the recognition of belief. According to the view that is currently standard in philosophy, in attributing knowledge of a proposition to some target agent, we attribute the mental state of belief, while simultaneously judging the agent’s situation to satisfy various non-mental conditions, in particular, the condition that the proposition believed be true, plus whatever further mental and non-mental features are necessary to distinguish knowledge from mere true belief. Attributions of knowledge, in this view, play no special or ineliminable role in our ability to explain and predict the actions of others.<sup>7</sup> The purely mental factor of believing is what serves as the key input to our understanding of what the agent does, together with some representation of that agent’s desires. The guiding principle of the current orthodoxy is that our ordinary understanding of action is supplied by belief-desire reasoning (Davidson, 1963; Dennett, 1971; Fodor, 1987).

There are certain sets of cases which make it seem perfectly natural to maintain this exclusive focus on belief as the pivotal mental state. When we are explicitly considering the relative advantages of explaining actions in terms of what is believed as opposed to what is known, it is easy to think of pairs of parallel scenarios in which an agent both knows and believes a proposition *p* in the first scenario, but merely believes *p* in the second scenario, and can be predicted to perform exactly the same salient action in both. For example, in (1) Janet knows (and *a fortiori* believes) that her mobile phone is in her handbag, remembering having seen it there a few moments before; in (2) Janet believes that her mobile phone is in her handbag, again remembering having seen it there a few moments before, but does not know that her mobile phone is in her handbag, because an extremely skilled pickpocket has just removed it without her noticing. If we add a certain

---

<sup>7</sup> The idea that attributions of knowledge play no special role in the success of our intuitive understanding of one another is not equivalent with the idea that knowledge plays no significant role in causal explanations. For example, Hilary Kornblith has argued at length for the causal and explanatory value of knowledge while expressing some skepticism about our natural intuitive capacity to register the presence or absence of knowledge, at least as this capacity is activated by thought experiments in philosophy (Kornblith, 2004). My own view is that our pre-theoretical intuitions about epistemological cases are the products of the same mindreading capacity that underpins ordinary mental state attribution (Nagel, forthcoming).

motivational state to these two scenarios—say, a sudden desire to make a phone call—we will expect to see just the same immediate action performed in each (reaching into the bag). Given that her immediate action in both cases would be the same, it can seem not to matter to the explanation of this action whether Janet knows or merely believes. On the currently orthodox philosophical view, the relevant proximal cause of her action in both scenarios is Janet’s belief that her mobile phone is in her bag, which together with her desire to make a call and various background beliefs will determine what she does. On this view it would be inefficient to say that her action in the first scenario was produced by her knowledge of the key proposition while in the second it was produced by mere belief; for maximal generality we should posit the same set of mental causes in the two cases.

At the first moment of reaching into the bag, the two scenarios may look very similar; however, when we extend the time frame and imagine the continuation of (1) and (2) we do expect differences to emerge. In (1) Janet will make her call; in (2) she will rummage in the handbag for a while and then come to the realization that something is amiss. But the advocate of the orthodox belief-desire view will not see this divergence as any reason to allow that knowledge has a special role in explaining action, or that the key mental states explaining action in (1) and (2) were perhaps different from the start. Rather, the difference in longer-range forecasts would be explained in terms of an interaction between the initial mental states (which on the orthodox belief-based view are the same in both scenarios) and environmental conditions (which differ between the scenarios on the point of whether the mobile phone is actually present), where these evolving conditions will come to produce different mental states in the parallel agents, further shaping what each will do. A considerable part of the value of mental state attribution is its capacity to generate predictions across variations in environmental conditions. Positing just a single initial mental state—belief—across both of our cases looks like an economical thing to do.

However, it will be a false economy if the state of knowledge is generally associated with patterns of behavior that are interestingly different from those produced by the state of mere belief. One could concede both that knowing entails believing, and even that in this particular pair of cases the same immediate action would be generated by knowledge or belief, without granting the general point that attributions of knowledge play no special role in our understanding of intentional action. Arguing against the orthodox philosophical view and in favor of the admission of knowledge as a mental state, Williamson has developed some examples of cases in which our expectations

concerning an agent's extended pattern of conduct do seem to be sensitive to whether we see that agent as having knowledge or mere true belief on some point.

A brief overview of his position, as presented in (Williamson, 2000), will facilitate discussion of these examples. Williamson stresses that his view is consistent with the orthodox view that knowledge in fact entails belief, so that any agent who knows that  $p$  could also be credited with a belief that  $p$ ; what he denies is that knowledge is a composite of belief and other factors, or more generally, a hybrid state composed of mental and non-mental conditions. Knowledge, in his view, is a stronger mental state than belief, but it is nevertheless a purely mental state in its own right, so in attributing it to others we would not ordinarily have to pass through intermediate stages of attributing the weaker mental state of belief and also recognizing the believed proposition as true. To describe knowledge as a purely mental state is not to suggest that it is determined entirely by the internal physical or spiritual configuration of the agent, in a manner that is independent of what is going on around her. In Williamson's view mental states like seeing or knowing that something is the case are not seen as, say, inner neurological states whose relationship to the environment is left entirely open, but rather as states in virtue of which an agent is related to an environment in certain ways (the thought that various neurological configurations can support our being in such states is undisputed).

The relevant ways of being related to the environment mark another point of agreement with the orthodox view: Williamson agrees that knowledge is a factive state. Indeed, what distinguishes knowledge from other mental states is that knowledge is the most general factive mental state. Williamson allows that there are various different factive mental states—seeing that  $p$ , remembering that  $p$ , regretting that  $p$ , being acutely conscious that  $p$  and so on—but in his view all of these entail knowing. Factive mental states other than knowing are invariably just more specific ways of knowing. So, classifying knowledge as a state of mind does not mean giving up on the condition that what is known must be true as a matter of objective fact; the point is rather that states of mind can be constituted in terms of relationships between agents and the objective facts of their environment.<sup>8</sup> The status of knowledge as our core factive mental state is what accounts for its special significance, according to Williamson: “factive mental states are important to us as states

---

<sup>8</sup> It is important that these relations bind agents (or subjects—I use those terms interchangeably in this paper) *directly* to the environment. Orthodox analyses can also bind an agent to the environment, but only indirectly, through the mediation of the state of belief. For example, in Goldman's causal theory the state of knowing that  $p$  was thought to consist in having a belief that  $p$ , where this belief was appropriately caused by the fact that  $p$  (Goldman, 1967).

whose essence includes a matching between mind and world, and knowing is important to us as the most general factive stative attitude" (2000, 40).

Knowledge on this view is a purely mental state rather than a composite of the mental state of belief plus non-mental factors. However, Williamson does not oppose the orthodox view about the entailment between knowing and believing: any agent who knows that  $p$  also believes that  $p$ . While the orthodox approach since Gettier has explained this entailment in terms of a composite theory of knowledge that takes belief as one of its basic building blocks, Williamson instead allows the entailment by characterizing belief as a potentially diluted version of knowing. In his view, 'believing  $p$  is, roughly, treating  $p$  as if one knew  $p$ ' (2000, 47). Matching how things are is not part of the essence of belief, although beliefs do sometimes happen to match how things are: we can believe false propositions, and we can believe propositions whose truth is not essential to our believing them. The attribution of knowledge is not taken to start from an attribution of belief, on Williamson's view; rather, the capacity to recognize belief depends on some prior mastery of the concept of knowledge. Nevertheless, the weaker and derivative mental state of belief obtains wherever the stronger mental state of knowledge does.<sup>9</sup>

Williamson aims to support his view of knowledge as a distinctive and pure mental state by identifying some cases in which the simple attribution of knowledge yields a better explanation of action than the composite attribution of belief plus the truth of the proposition believed. One such case invites us to consider the difference between a burglar who knows that there is an especially valuable gemstone in a given house and a burglar who merely has a true belief to the same effect. Other things being equal, Williamson suggests, the first type of burglar would be expected to be more persistent in searching the house, if only because mere knowledge is in general more secure than mere true belief (2000, 62). Mere true belief has various vulnerabilities not shared by knowledge: for example, if a belief falls short of being knowledge because it is only accidentally true, it might be derived from a false premise. So, Williamson notes, the second sort of burglar could believe that the stone is in the house on the basis of a false report that it is hidden under the bed, and because of his reliance on false testimony fail to count as knowing that the stone is in the

---

<sup>9</sup> This is not to say that an attribution of belief will always sound natural when an attribution of knowledge would be appropriate. Because we ordinarily aim to make our conversational contributions as relevantly informative as possible (Grice, 1975), by attributing the weaker condition of believing we would typically generate the conversational implicature that the stronger condition of knowing is not known to hold. So for example if I say "Justin *thinks* that I want him to come to my party," this carries the implication that I do not in fact want him there. But the implicature is cancellable and therefore merely pragmatic; I could say "Justin thinks that I want him to come to my party; indeed, he knows that I do."

house. After a thorough search under the bed, this burglar would be likely to desist where his counterpart who knows the stone is in the house would continue. Because it is essential to states of knowledge that they match how things are, knowledge is more robust than true belief as such. Discovering that a certain burglar engaged in an exceptionally tenacious search of a house—having spent hours on the property in peril of being discovered—we would better explain this behavior by saying that he *knew* the stone was in the house, rather than that he had the true belief that it was.

This type of example has encountered resistance from defenders of the orthodox belief-oriented line. If resolute behavior can be explained by appeal to knowledge, they contend it can be explained even more successfully by appeal to certain kinds of belief, perhaps particularly confident or inflexibly held belief (Magnus & Cohen, 2003; Molyneux, 2007).<sup>10</sup> As Williamson himself stresses, knowledge may be lost when one encounters misleading evidence; mere belief, on the other hand, may be dogmatic enough to shield its holder from reconsidering a given proposition in any circumstances. If we just want to explain exceptional persistence in a dangerous course of action, an attribution of obstinacy may serve us better than an attribution of knowledge.

The advocate of knowledge-based explanations of action can point out that in some circumstances we may have good reasons not to attribute sheer obstinacy to a resolute actor, but it is open to the advocate of the belief-based approach to suggest that in those cases there could be further conditions added to belief that would all together explain the situation as well as knowledge. To avoid granting the point about the special explanatory power of knowledge attributions, these further conditions could not themselves be tantamount to knowledge; well-founded true belief will be off-limits, for example, if we are using the technical expression ‘well-founded’ as some kind of abbreviation for ‘amounting to knowledge’. Magnus and Cohen suggest that a certain kind of rational-belief-that-will-tend-to-survive-scrutiny would be a good belief-

---

<sup>10</sup> An anonymous referee suggests that there may be a sense of ‘knowing that *p*’ which means nothing other than ‘being subjectively certain (or highly confident) that *p*’, citing as evidence for this view comments such as ‘I just *knew* the butler was the murderer, until it turned out, in the last chapter, that he wasn’t.’ On this view an attribution of ‘knowledge’ would be a better explanation of the burglar’s persistence than an attribution of mere belief, but not in a way that would appeal to anyone committed to the essential factivity of ‘knows’. However, it is not clear that we do have reason to admit a strictly confidence-driven and nonfactive sense of ‘knows’: there may be a better way of making sense of utterances like the remark about the butler. To begin, the special emphasis on ‘*knew*’ in that remark is plausibly a clue that ‘*knew*’ is not being used in a straightforward fashion. Richard Holton argues that such utterances gain their dramatic flair from ‘protagonist projection’, in which one puts oneself in the position of another (or in this case, one’s past self), deliberately selecting words fitting that perspective rather than our current one (Holton, 1997). Note that one could say to similar effect, “Up until the last chapter, the butler was obviously the murderer, and then it turned out he wasn’t.”

based explanatory substitute for knowledge (2007, 49) but it is not clear we can make intuitive sense of such a concept except insofar as it is a rough characterization of knowledge itself. Magnus and Cohen do not elaborate on the nature of this tendency, but if they think that we intuitively explain action by appeal to such a concept, the burden of proof is arguably on them to establish that we have any intuitive notion of it. Furthermore, without an account of what grounds this special tendency, explanations that appeal to it would have limited explanatory value: it is not particularly informative to explain rationally tenacious action as being the kind of action that is based on beliefs that are, for some unexplained reason, rationally tenacious.<sup>11</sup>

If mere true belief can fall short of knowledge in an open-ended variety of ways, as the frustrations of the analysis-of-knowledge project would suggest, there is no quick formula to summarize the difference made by attributions of knowledge as opposed to true belief, and it is not immediately obvious which side has a better line on the explanation of action. Several points may still seem to favor belief-based over knowledge-based approaches to mental state attribution. There are costs associated with allowing the richer state of knowledge as an explainer: most notably, we are no longer in a position to give an exactly uniform explanation of what Janet initially does in our two mobile phone cases if her action is explained by knowledge in the first and mere belief in the second. The belief-based composite approach brings with it the main attraction of compositionality, its promise of greater fertility and generality. Molyneux suggests that knowledge-based explanations are too finely tailored to their circumstances to have useful explanatory strength. Noting the similarity of a pair of cases like our Janet cases, he presses the concern that ‘when an explanation type fails to preserve its force across extremely similar scenarios, it provides us with evidence that it does not bear up well through contingent variation of any sort’ (Molyneux, 2007, 270).

A somewhat related line of objection to Williamson would focus on the sorts of cases that he takes to illustrate the special value of knowledge-based explanation of action, and point out that these cases might be ecologically quite rare. If we take the usual non-skeptical stance that ordinary cases of perceptually, inferentially and testimonially grounded true belief are knowledge, then true belief that falls short of knowledge is arguably going to be uncommon, and situations in which it figures are likely to be greatly outnumbered by situations involving ordinary false belief. If we need to evaluate the truth of the key proposition anyway in order to make an attribution of knowledge, it may seem largely redundant to admit knowledge as an independent explainer of action alongside

---

<sup>11</sup> This last point is owed to Tim Williamson, who notes ‘a salient similarity to dormitive virtue explanations’ here (Williamson, p.c.).



mere belief.<sup>12</sup> Situations in which we encounter beliefs that are true overlap so heavily with situations in which we encounter knowledge that it is hard to see how we would be ill-served by an approach that promised to make uniform sense of true belief and false belief situations, even if it did not make perfect sense of those rare cases in which an agent has accidentally true belief. Even if Williamson is right that knowers conduct themselves in subtly different ways from mere believers of the truth, the cases he is able to find to illustrate this point are exotic enough that we might hope to leave them on the sidelines as a range of curious exceptions to our general account of mental state attribution: perhaps the differences between knowledge and mere true belief could be handled by some codicil to a belief-based account of intuitive mental state ascription.

However, if natural social environments provide heavy overlap between observed true belief situations and observed knowledge situations, this is not necessarily bad news for a defender of the knowledge-based approach. If knowledge were a status only rarely attained by true belief, then the composite approach's capacity to generate a smooth compositional story about true belief and false belief would be more attractive than an approach giving a special primary role to the ecologically rare state of knowledge. If the most common prototypical examples of judgment that we witness are in fact moments of knowing, on the other hand, then it becomes more plausible to maintain that recognition of knowledge is prior to recognition of mere belief (whether true or false). Furthermore, it is not obvious that we must separately evaluate the truth of the proposition believed in the course of attributing knowledge. On some views of the testimonial transmission of knowledge, we can reasonably see another person as knowing a proposition without conducting any independent check of that proposition.<sup>13</sup> A defender of the composite approach could of course resist such views of testimony, or object that what the composite approach really needs is not that the mental state ascriber should consult independent evidence on any proposition another is taken to know, but just that the mental state ascriber needs to go through the work of representing that proposition as being true, in addition to being believed by the target.

However, the advocate of the knowledge-based account might observe here that in some cases of attributing knowledge to another, we do not even have to specify the proposition believed, let alone represent it as true: for example, we can see others as *knowing whether q*, where the truth

---

<sup>12</sup> Or perhaps alongside highly confident mere belief, in the event that one's composite concept of knowledge incorporates the requirement of passing a certain threshold of confidence. Another version of this objection would press a worry that knowledge explanations will be redundant with highly confident true belief explanations.

<sup>13</sup> Thanks to Tim Williamson for this point.

value of  $q$  is unknown to us. Not all knowledge attribution has the outward form “S knows that  $p$ ”. More broadly, attributions of knowledge can take a range of interrogative complements: the mental state ascriber can see another as knowing when the party is supposed to start, where the drinks are, or who is invited, without any particular answer to these embedded questions being endorsed by the ascriber.<sup>14</sup> Indeed, our ability to see others as knowledgeable on questions whose answer is unknown to us is arguably a core feature of our intuitive social understanding of others. The advocate of the composite belief-based approach could of course develop a composite treatment of this aspect of our social intelligence: perhaps what we really recognize in others is their reliability, or something of that nature, and perhaps attributions of *knowing-wh* are rather different in kind from ordinary attributions of *knowing that p*. But if we take them at face value and attempt to treat them in a uniform manner, attributions of *knowing-wh* do seem to count against the original suggestion that knowledge attributions always require the composite representation of a known proposition as true and as believed.

For many philosophers, the deepest problem with classifying knowledge as a mental state is quite different in character from the problems canvassed so far. It is a metaphysical problem about the localization of mental states. Because the essence of knowledge involves matching how things are in the world, a distal change in the environment can effect a change of mental state in a subject even if nothing in her internal consciousness or neurology changes at that moment. Early in April of 1865, an ordinary well-informed American citizen knows that Abraham Lincoln is President; at some point on the evening of the 15<sup>th</sup> her mental state switches from knowledge to mere belief.<sup>15</sup> As soon as the President is assassinated and the proposition about his being President thereby becomes false, our citizen’s mental state will have undergone this change, even if she has not heard what has happened in the faraway theatre, and even if she would say and do just the same things immediately following this change as before it. In the same vein, we can think of Janet’s mental state before and after her mobile phone is stolen; when she first puts the phone in her bag she knows that it is there, and when it is stolen (or even just at very immanent risk of being stolen) she no longer has this knowledge of its location, but there need be no significant difference in either her internal neurological or conscious state between those two moments (assuming our pickpocket is stealthy enough to do nothing Janet consciously experiences). In short, the distal vulnerability of factive states may make them seem unfit to serve in causal explanations of behavior: critics of the

---

<sup>14</sup> Thanks to Tim Williamson for this point, and to Jane Friedman for discussion of the issues here.

<sup>15</sup> This example is again Williamson’s (2000, 23).

knowledge-first approach have complained that positing knowledge as a mental state appears to violate strictures on action at a distance. Because nonfactive mental states like *believing that Lincoln is President* or *believing the mobile phone is in the bag* can persist through a change in the truth value of their contents, it can seem that they are better candidates for the causes of our actions. Pointing to intuitive theories according to which causal powers must always be in the same location as their effects, they object that ‘attributions of causal efficacy to knowledge work, if at all, by piggybacking on the causal efficacy of belief’ (Magnus & Cohen, 2003, 40).

In response, Williamson claims that the best current accounts of the propositional attitude of belief do not localize it within the neurological or conscious states of the agent either; perhaps beliefs are also entangled with distal conditions in making reference to external particulars (2000, ch. 2).<sup>16</sup> This paper will not attempt to evaluate the arguments of Williamson and others on the issue of whether we can find some way of making sense of belief as a local mental state; serious treatment of these arguments would require engagement with a large and unwieldy body of work on the nature of mental content and mental states (for some discussion, see Fricker, 2009; Jackson, 2007; McDowell, 1986; Putnam, 1975; Williamson, 2009). What will instead be pursued in the next two sections is a complementary strategy of challenging the claim that good explanations must be anchored in strictly local conditions.<sup>17</sup>

If we consider a discipline that gives uncontroversially good explanations of its target domain—the science of physics as applied to the interactions of mid-sized inanimate observable objects—we find reasons to resist the condition that all real explanation is local. Our intuitive understanding of these interactions may however feel more local than it really is. The ‘naïve physics’ which gives us an intuitive ability to predict the dynamic interactions among inanimate bodies is generated by a complex set of implicit computations involving the relationships among those bodies and so-called ‘environmental invariants’ such as gravity and friction (McCloskey & Kohl, 1983). Interestingly enough, we have a natural tendency to ‘localize’ the results of these computations, for example seeing the trajectory of an object as determined by its own private

---

<sup>16</sup> On a related note, knowledge is sometimes thought to differ from belief in having ‘normative’ status: the person who is described as knowing is being positively evaluated, where the person who merely believes may be doing something wrong. Without disputing the claim that knowledge is normatively superior to mere belief, one might observe that belief itself has normative status: for example, as Plato observes in the *Theaetetus*, in order to have a belief about anything one must be appropriately acquainted with it, where it is at the very least far from obvious how one could cash out ‘appropriately’ in a non-normative fashion (Plato, 1990).

<sup>17</sup> Williamson remarks congenially on the possibility of attempting this strategy, but does not develop it himself (2000, 61).

impetus or 'representational momentum', a strictly local power that is seen as vested in it as the result of its inner nature and contact with other things. However, this localization is a fiction, both because the principles of impetus physics are not entirely accurate, and because in computing it we draw essentially on the relationships between objects, relationships involving "action-at-a-distance" factors such as gravitation (Hubbard, 1995). We can recognize on reflection that an inanimate object's motions do not depend strictly on local interactions between its inner condition and the objects it physically contacts, but this explicit recognition does not extinguish our natural intuitive tendency to see ordinary outer objects as driven in this way (Kozhevnikov & Hegarty, 2001).

As a parallel,<sup>18</sup> it is possible that we do in some sense intuitively represent the mental state of knowledge as localized within an agent, even if we actually intuitively compute the presence of knowledge on the basis of our grasp of relations between that agent and her environment, and even if we are able to recognize on reflection that this state must reflect how things are in the agent's environment. In intuitive judgment, we are aware of the final product or outcome of a process of computation, and not of the processing that gives rise to it (Sloman, 1996); while reflective and deliberate attributions of knowledge may make us conscious of the role played by distal factors in knowledge attributions, it is entirely possible for intuitive attributions of knowledge to incorporate distal factors without alerting us to this fact. If the epistemic state underpinning action is something that subjectively seems local, our intuitive representation of knowledge may well provide this subjective impression: we could at some level feel that an agent's knowledge is localized within her, just as we feel that an object's motion is sustained by something within it.<sup>19</sup> As far as objective reality is concerned, the example from intuitive physics should give us second

---

<sup>18</sup> It is plausible enough that there are parallels between naïve physics and mindreading, given the overall structural similarities in the function of these capacities. A psychological capacity for naïve physics underpins our ordinary experience of the motions of inanimate objects, enabling us to anticipate (and sometimes manipulate) their motions; mindreading underpins our ordinary experience of the actions of animate agents, enabling us to anticipate (and sometimes manipulate) what they do. Both capacities require very rapid assimilation of a broad range of complex information about our environment, much of it below the threshold of conscious or explicit reasoning. Psychologists do sometimes discuss the natural strengths and limitations of one domain in terms of an analogy with the other (e.g. Saxe, 2005).

<sup>19</sup> Tamar Gendler observes (p.c.) that there may be a significant relationship between this tendency and the Fundamental Attribution Error, a bias which inclines us to assign unduly heavy weight to inner personality characteristics and insufficient weight to situational factors in our explanations of the actions of others (Jones & Harris, 1967; Ross, 1977). This is an intriguing suggestion, detailed exploration of which cannot be attempted here, if only because it would lead us into some very complex issues involving self/other asymmetries in the explanation of action, and into problems concerning the natural extent and character of misrepresentation in ordinary action explanations.

thoughts about the wisdom of a general principle ruling out appeal to non-local factors in our explanations of what is really going on, and even in our explanations of what is really going on in the generation of our intuitive assessments.

There is something jarring about the reflective observation that a subject can switch from one mental state to another (e.g. from knowing to not knowing) through something that happens remotely. However, the attribution of quite different mental states to the agent before and after the change may be natural enough if there is a genuine change in how we intuitively represent them as agents, or if suitably distinct procedures naturally apply to representing knowledge and false belief. If we take belief to be a state that is derivative from the state of knowledge, then the more complex attribution of a false belief that  $p$  could nevertheless in some circumstances enable a prediction of the same immediate action that would be produced by knowledge that  $p$ . This observation could be applied to the explanation of the similar immediate behavior of our two Janets, one of whom knows that her phone is in her bag while the other falsely believes it. Williamson's view does not sever the link between knowledge and belief altogether: if believing that  $p$  is understood as a state that involves 'roughly, treating  $p$  as if one knew  $p$ ', then the predicted outcome of immediate action given the false belief that  $p$  may be systematically similar to the predicted outcome of immediate action given the knowledge that  $p$ , although the former type of prediction may be expected to have some additional level of complexity.

At this point it may be helpful to examine some independent evidence on the nature and relative complexity of our intuitive attributions of knowledge and belief, and on the question of whether the concept of knowledge is in some sense prior to the concept of belief, or whether it is composed from that concept and further conditions. In particular, it may be useful to examine developmental work on the emergence of the concepts of knowledge and belief: arguably, if intuitive representation of knowledge really is a composite involving intuitive representation of belief, the capacity to represent knowledge should not be available unless and until the capacity to represent belief is in place. The next section of this paper examines some recent empirical work which addresses this issue, and the closely related issue of the relative complexity of ordinary intuitive ascriptions of knowledge and belief.

### Section 3: empirical work on the relationship between the concepts of knowledge and belief

The standard view in psychology is that knowledge is a mental state; indeed, knowledge typically features prominently in lists of paradigmatic mental states (e.g. Apperly, 2011; Heyes, 1998; Premack & Woodruff, 1978; Sodian & Kristen, 2010; Wellman & Liu, 2004). Reading the empirical literature on mental state attribution, one sometimes finds an article which omits mention of knowledge as a mental state (e.g. Russell, 2005), but I do not know of any article in this body of literature in which it is explicitly argued or even claimed that knowledge is not a mental state. Marking a further point of difference from the mainstream philosophical view, psychologists of mental state attribution do not seem to assume that the concept of knowledge is an elaboration of the concept of belief; in fact, it is generally agreed that belief is the more sophisticated concept, and is harder to attribute than knowledge.<sup>20</sup> In comparative psychology, there is nothing alarming about entitling an article, “Chimpanzees know what others know, but not what they believe” (Kaminski et al., 2008). In developmental psychology it is widely held that children acquire the concept of knowledge before the concept of belief.<sup>21</sup> After a very brief review of psychological work on the relationship between the concepts of knowledge and belief, this section will discuss the reasoning behind the psychological classification of knowledge as a mental state.

There is a wide range of evidence relevant to the question of whether the concept of knowledge is prior to the concept of belief. One might begin by looking at work on the acquisition of the mental state lexicon. Children use ‘know’ both earlier and more heavily than they use ‘think’ (Bartsch & Wellman, 1995; Shatz, Wellman, & Silber, 1983). Both the earlier age of initial use and the increased frequency of a word for knowing over a word for thinking appear cross-culturally, for example in languages such as Mandarin and Cantonese (Tardif & Wellman, 2000). Drawing on a database of over 200,000 spontaneous utterances by English-speaking children up to the age of six, Bartsch and Wellman found that the verb ‘know’ figured as the main verb in 70% of children’s epistemic claims, with ‘think’ following at 26% (other epistemic mental state verbs, such as ‘believe’, ‘wonder’ and ‘expect’, were rarely used) (Bartsch & Wellman, 1995).<sup>22</sup> Before the age of

---

<sup>20</sup> This observation is of course compatible with the notion that knowledge entails belief. A creature must be a mammal in order to be the conspecific of a primate—being a conspecific of a primate entails being a mammal. Nevertheless many primates have the conceptual capacity to recognize their conspecifics while lacking the conceptual capacity to recognize mammals as such.

<sup>21</sup> Williamson mentions the potential relevance of this fact to his ‘knowledge first’ approach, but does not elaborate (2000, 33).

<sup>22</sup> The dominance of ‘know’ over ‘think’ continues in adult usage but is less pronounced. ‘Know’ and ‘think’ are respectively the 8<sup>th</sup> and 12<sup>th</sup> most common verbs in English, according to the Oxford English Corpus. In

about four, children have trouble with the word 'think', showing some tendency to take it as factive (Abbeduto & Rosenberg, 1985), although children early into their third year do use the word and are sometimes even able to produce the contrastive utterances taken to be the best evidence of an understanding of nonfactive verbs—e.g. "I thought there wasn't any socks, but when I looked, I saw them" (Bartsch & Wellman, 1995, ch.3).

Explaining the initial acquisition of epistemic verbs is no easy matter. In the ordinary course of development, there is a cross-culturally robust and very significant lag between the emergence of action and activity verbs like 'go', 'eat' and 'push', and mental state verbs like 'think' and 'know' (Choi & Gopnik, 1995). Summarizing a few of the relevant difficulties, Papafragou and colleagues note that verbs like 'think', 'remember' and 'know'

do not refer to perceptually transparent properties of the reference world; they are quite insalient as interpretations of the gist of scenes; they appear frequently in maternal speech to babies and yet occur in the child's own speech comparatively late; the concepts that they encode are evidently quite complex or abstract; and they are hard to identify from context even by adults who understand their meanings. (Papafragou, Cassidy, & Gleitman, 2007, 126)

They go on to make the following observation about their approach to the problem:

The idea here is that words that refer to mental states and events lack obvious and stable observational correlates: as a general rule, it is easier to observe that jumpers are jumping than that thinkers are thinking. Assuming that word learning – especially in its initial stages – relies heavily on establishing relations between what is said and what is (observably) happening in the extralinguistic environment, it follows that there will be barriers to acquiring mentalistic terms even where the concepts themselves fall within the learner's repertoire. (Papafragou et al., 2007, 128)

To adults, mental state talk feels so natural that it takes some imaginative work to appreciate the mental state verb acquisition problem from the perspective of the child. Although children hear verbs like 'think' and 'know' heavily enough, it is not easy for them to forge the link between what is heard and what it refers to. Some aspects of the difficulty here can be reproduced even in adults

---

the 425-million word Corpus of Contemporary American English these words occur at a ratio of 53:47. In the context of explicit action explanations there is some evidence that knowledge dominates. In COCA, for example, the ratio between instances of the strings 'because s/he knew' and 'because s/he thought' is 68:32. Including the next most common non-factive ('believe') alongside 'think' drops the ratio to 58:42. This is only superficial evidence of the importance of knowledge ascription in action explanation; a detailed study of this issue would have to examine corpus data for a wider range of factive and nonfactive verbs and constructions, and address questions about the relationship between our ordinary understanding of action and our verbal reports of this understanding. If we tend to remark on people's reasons for acting when their actions are out of line with our expectations, as Gricean norms might predict, then the method of counting occurrences of 'know' and 'think' would presumably underreport reliance on attributions of knowledge.

through an experimental protocol called the Human Simulation Paradigm (HSP). In one version of the HSP, adult subjects watch a series of brief silent videos of mother-infant interaction and try to guess the mystery word spoken by the mother at the moment of an electronic beep; for each mystery word subjects were told in advance whether it was a noun or a verb, and shown a series of six taped interactions in which the word was used. At the end of the series the subjects could reflect on all that they had seen and offer a seventh and final conjecture as to the identity of the target mystery word. None of the chosen words were particularly uncommon: Gillette and colleagues tracked the 24 nouns and the 24 verbs that occurred most frequently in their videotaped samples. Rates of correct identification were strikingly different for different classes of words: on average 44.9% of the nouns were correctly identified by the final try, versus only 15.3% of the verbs. Concrete nouns like 'piggy' (89.3% right), 'ball' (78.6%) and 'plane' (100%) did relatively well, as did some verbs: 'come' (75.0%), 'throw' (85.7%) and 'push' (42.9%). Mental state verbs did very badly: both 'think' and 'know' were identified by zero percent of subjects (Gillette, Gleitman, Gleitman, & Lederer, 1999). Given that adults already have mature mental state concepts, it should be easier for them than it is for infants to pick out adult uses of mental state words; the very poor performance of adults in recognizing mental state words in the HSP experiments suggests that the problem of finding 'observational correlates' of mental state verbs is a serious one.

One reason why the task was difficult was that common mental state verbs are often used for reasons other than referring directly to a salient mental state; for example, in fixed expressions (such as the attention-getting 'know what?') and to perform social functions such as softening a request ('I think it's time to go'). These 'conversational' uses of mental state verbs are relatively frequent (Bartsch & Wellman, 1995) and are agreed to offer very little information to the language learner (Shatz et al., 1983). In their initial HSP experiments, Gillette and colleagues had made no special effort to find scenarios in which mental states would be particularly notable. Papafragou and colleagues experimented with various types of scenario to see whether some circumstances would improve performance on mental state verb recognition. They contrasted scenarios in which someone was fooled or mistaken (FB scenarios) with closely matched counterparts in which nothing went wrong (TB scenarios). In a typical FB scenario, a woman drinking tea while reading a newspaper did not observe her cup being moved by someone wiping the table and a teapot being set in its place; she reaches for the teapot and brings it close to her mouth before noticing her error. In the control TB scenario there is the same sequence of object displacements but the woman is not misled and picks up the cup. Children and adults then had to speculate about what would be said by a mother viewing the scene with her child. As in the earlier HSP experiments, most responses



focused on the concrete and observable aspects of the scenarios. Overall, looking at the TB and FB experiments together, both child and adult responses were skewed towards action verbs (43.9% and 31.6% respectively); among mental state verbs the verbs of desire were dominant (23.5% and 30.9% of child and adult responses referred to motivational mental states). Epistemic mental state verbs were used in 11.5% and 23.5% of the child and adult responses. But strikingly enough, the FB scenarios produced a much higher incidence of epistemic verb use: in comparison to the TB scenarios, the FB scenarios elicited more than triple the rate of epistemic verb use for children, and almost double for adults (Papafragou et al., 2007).

These experiments do not show that we typically fail to register mental states in witnessing the successful actions of others. The discovery that epistemic mental states become particularly salient when actions do not unfold as expected might rather suggest that our background expectation is that agents will act knowledgeably, and we find it remarkable when they do not.

The default expectation of knowledgeable behavior on the part of others is thought to be manifest in a well-known sequence of nonlinguistic mental state ascription tasks. The first of these, passed in some versions even by some nonhuman primates, measures the capacity to distinguish knowledge from ignorance; the second task measures the capacity to attribute false belief, and is passed only by humans, and only later than the knowledge-ignorance task. Some description of the tasks may help to clarify why psychologists of mental state ascription generally take the concept of knowledge to be prior to the concept of belief.

In one version of these tasks, pairs of children were given a familiar container with familiar contents to examine (a domino box with picture dominos in it). One child from each pair was then sent out of the room, and in his absence the other witnessed the familiar contents being replaced with a novel item. The second child was then asked (in random order) two questions: (1) Does [name of absent child] know what is in the box now, or does he not know? (2) If we ask [name of absent child] what is in the box, what will he say? Among 3-year-olds, 39% answered question (1) correctly, and only 6% answered question (2) correctly; 4-year olds improved to 81% and 44%, and 5-year olds were right 88% and 76% of the time on the two questions (Hogrefe, Wimmer, & Perner, 1986).<sup>23</sup> If we generally made judgments about the presence or absence of knowledge by attributing belief and then evaluating the truth or falsity of this belief, we would not expect to see such a lag between the capacity to recognize the absence of knowledge and the capacity to attribute

---

<sup>23</sup> Difficulty with the second (false belief) type of question is still encountered if the child simply has to select between two pictures representing the two possible mental contents of the other person (Custer, 1996; Woolfe, Want, & Siegal, 2002).

a false belief.<sup>24</sup>

Young children are not random or indiscriminate in the way they fail the knowledge-ignorance task: they have a general tendency to over-attribute knowledge to others. This tendency has been found across a great variety of different tasks; in general, the trend as children mature is towards the recognition of more and more of the restrictions on knowing that are imposed by the various limits of sensory modalities, perspective, perceptual access and experience (e.g. Apperly & Robinson, 2003; Mossler, Marvin, & Greenberg, 1976; Wimmer & Perner, 1983; Woolley & Bruell, 1996). But studies of immature and evolving mental concept use naturally raise questions about whether young children are really referring to *knowledge* as such when they come to succeed at tasks of the sort just given. In this context, it is useful to examine work in comparative psychology, where there has been a very active debate on the question of just what should count as genuine mental state recognition.

Nonhuman primates can be very responsive to each others' behavior. For example, they track where their conspecifics are looking, following the direction of their gaze (Tomasello, Call, & Hare, 1998) in a way that is sensitive to the presence or absence of opaque barriers blocking the gazer's line of sight (Okamoto-Barth, Call, & Tomasello, 2007). The question of whether they are capable of mental state recognition is however difficult to assess. Early research on mental state reasoning in nonhuman primates seemed to show that they lacked the capacity to distinguish knowledge from ignorance. Chimpanzees will make visible begging gestures equally to human trainers who can see them and to those who have their vision obstructed by a blindfold or bucket, although they do discriminate between trainers who are and are not facing them (Povinelli & Eddy, 2000). In a food-hiding game, chimpanzees also fail to discriminate between the helpful pointing gestures of a trainer they have seen hiding food (behind a barrier, in a location out of the chimpanzee's view) and a trainer who was absent at the time of hiding (Povinelli, Rulf, & Bierschwale, 1994). Negative results such as these seemed to support the view that nonhuman primates track surface behavioral regularities (like 'facing me – likelier to give food') but do not understand the relationship between seeing and knowing, and fail to attribute mental states such as knowledge or ignorance (Heyes, 1998; Povinelli & Vonk, 2003; Tomasello & Call, 1997). On this view, these primates' navigation of

---

<sup>24</sup> This interpretation of the contrast is somewhat controversial, not least because there are various ways of understanding the difference between the knowledge-ignorance task and the false belief attribution task. Some psychologists have raised the concern that the knowledge-ignorance task is inherently easier for various irrelevant reasons, such as having a higher baseline of being answered correctly simply by chance (e.g. Perner, 1995). For a response to such concerns in the context of a large-scale meta-analysis of studies showing knowledge-ignorance to be easier than false belief attribution, see (Wellman & Liu, 2004).

social situations is guided strictly by learned or innate ‘behavioral rules’ linking certain observable situations or patterns of behavior and their typical behavioral consequences. This superficial grasp of behavior patterns would enable nonhuman primates to predict behavior in a fixed set of familiar situations; unlike mindreading abilities that discern the deeper mental causes of action, the ‘behavior-rule’ ability would not generate predictions flexibly, across novel circumstances.

More recent work has reopened the question of whether nonhuman primates might have some grasp of the unobservable mental states underlying behavior. Artificial cooperative games with humans might not be the best setting to test a chimpanzee’s mindreading abilities, and a number of primatologists wondered whether chimpanzees would perform better in a competitive setting involving their conspecifics. Exploiting the fact that subordinate chimpanzees will not challenge dominant chimpanzees for food, Hare and colleagues allowed subordinates to watch food being hidden outside their cages while a caged dominant either was or was not also watching within sight of the subordinate. Both chimpanzees were then released from their cages, with the subordinate getting a brief head start so he could not simply react to the dominant’s behavior. Subordinates were able to distinguish between knowledgeable and ignorant competitors, preferring to go for the food when the dominant competitor was ignorant of its location (Hare, Call, & Tomasello, 2001). A series of control conditions ruled out various potential non-mental explanations of this behavior: for example, chimpanzees did not simply regard the food that had been seen as tainted or dangerous, but kept track of the relationship between the food and the specific individual who had seen it. If a different dominant animal was brought in rather than the one who had seen the food, the subordinate no longer avoided it. Surveying these and an array of other recent experiments, leading comparative psychologists who had earlier concluded that nonhuman primates do not grasp mental states (Tomasello & Call, 1997) came to revise their opinion, arguing that the total current body of evidence supports the conclusion that “chimpanzees, like humans, understand that others see, hear and know things” (Call & Tomasello, 2008, 190).

It is possible to maintain a strict behavior rules interpretation of the recent experimental results, for example, by positing that what the subordinate animal really knows is not that the dominant *knows where the food is*, but that *any dominant will go for food that has recently been in its line of sight* (Povinelli & Vonk, 2003). However, this strategy may put us on a slippery slope: if we are allowed to formulate behavior rules ad hoc, then a behavior rules interpretation could be produced for any pattern of apparent mental state ascription, even one involving advanced social interactions among adult humans (Fletcher & Carruthers, 2012). The fact that a certain interpretative strategy is in principle available—as behaviorism continues to be available, in principle, for

explaining even human social interaction—does not mean that it is advisable, or the that it is the strategy that would yield the best explanation of our data. Considerations of parsimony and fruitfulness also matter. It is not trivial to show that the mindreading interpretation is more parsimonious than its behavioral rules counterpart as an explanation of the data involving nonhuman primates; in particular, the inherent complexity of the ‘mindreading module’ which links knowledge, goals and behavior must itself be taken into account, and cannot be treated as a ‘black box’ which the theorist gets for free (Perner, 2010). However, considerations of parsimony weigh more and more heavily in favor of a mindreading interpretation as creatures exhibit their competence in predicting the behavior of others across a broader range of situations. On the mindreading interpretation, the chimpanzee can systematically combine attributions of *knowing where something is* with attributions of *wanting that thing* to generate predictions about behavior in novel circumstances, where there could be various different outward signals of knowing or wanting; meanwhile, the behavior rules interpretation has to posit a separate rule to connect each type of outward signal with each type of behavioral consequence (Fletcher & Carruthers, 2012).

Interestingly, support for a mindreading interpretation of nonhuman primate abilities arises also from evidence about their limitations in social reasoning. There are some mental state attribution tests that nonhuman primates are apparently unable to pass, although as Fletcher and Carruthers point out in their (2012), it is possible to concoct relatively simple behavior rules that would dictate a successful pattern of reasoning for these tasks. The crucial feature that distinguishes these harder tasks from the ones that the chimpanzees succeed at is that they involve belief rather than knowledge. Nonhuman primates consistently fail false belief tests (Call & Tomasello, 1999; Hare et al., 2001), even in competitive situations and using apparatus that enables them to pass very closely matched knowledge-ignorance tests (Kaminski et al., 2008; Krachun, Carpenter, Call, & Tomasello, 2009). For example, when two chimpanzees or bonobo apes in view of each other have seen a reward placed in one container, and then the target animal is away while the reward is moved to another container, the subject animal does not seem to be able to anticipate that the returning target animal will seek the reward where it was originally placed, even when there would be a competitive advantage to this recognition. If apes were able to pass this test, we could easily write a behavior rule to underwrite their passing it: Fletcher and Carruthers suggest the rule, ‘A potential competitor will approach the previous location of food if the competitor was absent when the food was moved to its present location’. This rule is not inherently any more complex than the behavior rule explaining successful performance on a matched knowledge-ignorance task, so the advocate of the behavior rules approach faces an extra burden in explaining

why such a rule does *not* become part of the ape's repertoire. It is hard to see why only one of these patterns of action can be anticipated by apes without helping ourselves to the idea that the contrast in mental states makes a difference: knowledge attribution is easier than belief attribution.<sup>25</sup>

One way of appreciating the greater simplicity of knowledge attribution is to follow the earlier line according to which knowledge is a state that essentially involves matching how things are, where belief is essentially a state that may or may not match reality. The additional degree of freedom in belief attribution poses an additional computational burden, which matters because a very significant challenge in explaining mature human mindreading is explaining how it is computationally possible, given the open-ended character of the information that might be task-relevant. In contrast to our attributions of a state like jumping, our attributions of mental states are not narrowly constrained by what is immediately observable. Ian Apperly has emphasized this difficulty in connection with the classic test of false belief. Sally sees an object in the round box, and leaves the scene. In her absence, Andrew moves it to the square box, and we are asked where Sally will expect to find it on her return. The 'right' answer—the answer that counts as a 'pass' on false belief tests—is: 'in the round box where she left it.' But Apperly points out that many other mental state attributions are logically consistent with the scenario as given: perhaps Sally will guess correctly that Andrew has moved the object; perhaps she will know that he has moved it, on the basis of what he has told her in the past; perhaps she believes that it has spontaneously teleported to the square box, or to some quite different location (2011, 118). Given the right supplementary evidence about this scenario, any one of those other attributions might be correct.

The default understanding of the scenario makes sense, however, and will ordinarily serve us well: according to Apperly, we ordinarily construct simple and stereotyped models of situations involving other agents, drawing on our background knowledge of interactions like the ones we are

---

<sup>25</sup> By observing that chimpanzees have some capacity to recognize the state of knowledge, one need not thereby credit chimpanzees with any very sophisticated understanding of the nature of knowledge, nor even with the capacity to recognize knowledge across a great range of situations. For example, they may be unable to pass tests requiring the attribution of knowledge gained through hearing rather than vision (Brauer, Call, & Tomasello, 2008). Ian Apperly has expressed concern that a modality-restricted conception of knowledge would be seriously impoverished: commenting on chimpanzees' apparent failure to pass auditory tasks, he notes "This would be significant because a core feature of the concept of "knowledge" is that it provides some unification over the results of a variety of perceptual and inferential processes. If chimpanzees' understanding of "knowledge" is modality-specific then it falls short of providing this conceptual unification" (2011, 53). Nonetheless, Apperly himself stresses the importance of rudimentary mental state concepts (or proto-concepts) in paving the way for our more sophisticated human capacities for explicit mindreading; he also grants the general priority of knowledge recognition over belief recognition. His overall assessment of the comparative literature is that 'there is good evidence that some non-human species "understand something about" seeing or knowing, but do not understand these mental states completely, and have no understanding of other mental states, such as believing' (2011, 109).

observing, in a process he compares to the situation modeling approach to discourse comprehension (e.g. Zwaan & Radvansky, 1998). The default state of knowledge plays an important role in this process of situation modeling. What makes it reasonable to grade ‘Sally thinks the object is in the round box’ as the right answer to the problem (or to attribute this mental state to Sally in a real-world version of the problem) is that this agent would know that the object is in the round box if things had gone for her as they normally do in a person’s momentary absence. If recently having seen something somewhere were not generally a reliable source of knowledge of its location, we would have no reason to attribute that particular belief to Sally.<sup>26</sup> So a prior sense of how knowledge ordinarily arises helps to make belief attribution a tractable problem.

The relative simplicity of factive attitudes also plays a role here. Apperly observes that ‘making a mindreading inference consists of constructing a model of “what is in the head” of the target other person, but this model is both related to *and* part of our larger model of the situation in which that person figures’ (2011, 131). Of the various different relationships that might hold between another person’s mental state and the world, the factive ones are arguably the simplest: where the other person is seen as knowing, her mental state essentially reflects how things are around her. When we have reason to do so, we can posit restrictions on what others are taken to know, and see their mental states as deviating, or even just potentially deviating, from reality, but there are computational costs associated with broadening our horizons to include these more complex relationships.<sup>27</sup> Given these costs, there is a genuine difference in how we intuitively evaluate an action performed by someone who knows some relevant proposition and an action performed by an agent in a parallel scenario who falsely believes a parallel proposition: predicting the second kind of action is a more complex task, requiring the construction of a situation model to represent the agent’s belief, where this embedded model will be at odds with the larger model representing the agent’s situation. It will even increase the complexity of our prediction to see someone as truly believing without knowing; in order to represent a belief that is only accidentally right, we need to model situations in which someone could have had a similar false belief.

When we think about the relative advantages of explaining action strictly in terms of belief

---

<sup>26</sup>In a similar vein, Fletcher and Carruthers observe that ‘the false-belief data make no sense in the absence of a capacity to attribute goals, perceptions, and knowledge to other agents, since the experiments are all designed around just such an assumption, and since no one has any idea how a set of behavior-rules could fit together with belief understanding to issue in the patterns of behavior we observe.’ (2012, 11)

<sup>27</sup> It may be more accurate to say that there are computational costs involved in admitting false belief when we reach the level of explicit mental state ascription: in Apperly’s interpretation, recent work on implicit false belief recognition in infants (e.g. Onishi & Baillargeon, 2005) suggests that early implicit recognition of false belief may be fully automatic (cf. Apperly, Riggs, Simpson, Chiavarino, & Samson, 2006).

(whether true or false) as it interacts with other mental and environmental conditions, it can seem that such explanations have a cost-free advantage over knowledge-based explanations. Once we are imagining pairs of scenarios in which a key proposition is known in the first, but merely believed in the second, we have already done the work of constructing the more elaborate set of models for the second scenario: for example, we have already imagined Janet as the unwitting victim of a theft, thinking her phone is in her bag when it is not. In this frame of mind, it may seem natural to explain Janet's action in terms of her belief across the pair of cases, but the contemplation of ways in which an agent might be mistaken is not a natural part of the ordinary ascription of knowledge, so our impression of the simpler scenario may already be distorted by our focus on the second.

Apperly's talk of "what is in the head" may seem uncongenial to a treatment of knowledge as a factive mental state, but there is reason to think his scare quotes should be taken seriously here. He unpacks the metaphor elsewhere by saying that mental states themselves are not immediately visible but are known to us through their outward signs, claiming for example that 'we do not have direct access to what other people know, want, intend or believe, but must infer these mental states on the basis of what they do and say' (2011, 1). His more detailed discussion of the manner in which mental states are attributed makes it clear that these attributions draw not only on the target agent's local actions and utterances, but also on environmental conditions around the agent. In attributing knowledge of the external world to an agent, we see that agent as related in a certain way to that world.

#### **4. Conclusion**

This paper might be faulted for attempting to answer the question of whether knowledge actually is a mental state by arguing that it is naturally perceived as one. Even if our natural mindreading 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. Someone might be satisfied that we intuitively attribute knowledge to others as a state which explains their actions, without being satisfied that it really has the role it is intuitively seen to have. If we take mental states to be those states which actually do explain the actions of intelligent agents, we might wonder whether the intuitively recognized state of knowledge merely seems to explain what people do, while the real underlying explanation is substantially different, and perhaps more accurately expressed in terms of the weaker mental state concept of belief, plus

various further conditions.

Indeed, once we are wondering about the relative merits of knowledge- and belief-based explanations of action, we may start to wonder whether there is actually any such thing as being related directly to one's environment in a way that essentially involves getting it right. It may come to seem that when we are being our usual lazy selves, we assume that all is well with ourselves and other agents; the application of a little cognitive effort is all that it takes to recognize that what seems very much like knowledge could always fail to hit the mark. If I find myself intuitively attributing knowledge to someone—Janet just put her phone in her bag, so she knows where it is—it is open to me to rescind that generous ascription, on reflection, in favor of something more limited. Dwelling on counterfactual possibilities in which Janet might have been robbed without her knowing it, or more extreme counterfactual possibilities in which the phone spontaneously dissolves, I might start to feel inclined to attribute to Janet a state of mind that falls short of essentially matching how things are: perhaps she is just lucky that the world she inhabits is not one of the worlds I have just imagined, and in that case should be credited only with true belief.

Insofar as the application of greater cognitive effort often accompanies an increase in the precision of our judgment, we may feel that we are being more accurate in taking this skeptical stance. However, a shift towards skepticism does not necessarily make us better at predicting and explaining how others will act: by adding extra possibilities to our models of their relationships with their environment, we may be weakening rather than strengthening our understanding of what they are doing. The skeptical stance generates an awkwardly large space of possibilities to be contemplated, and our capacity to reason accurately about all those possibilities is limited. If simplicity is generally a virtue in explanations, the greater simplicity of knowledge-based explanations of action should be counted as a point in their favor.

Of course, thinking about possibilities of error sometimes really can increase the accuracy of our mental state ascriptions: the original philosophical critics of Premack and Woodruff were right to observe that the capacity to attribute false belief is a powerful weapon in the mindreader's arsenal. Other things being equal, creatures who are unable to imagine a conflict between belief and reality will not be as accurate in their predictions or explanations of action as creatures who are able to attribute false belief. But this is not to say that predictions and explanations of other agents generally become better the more we think about them, even when we have no positive grounds to suspect an agent is mistaken; serial contemplation of various possibilities of error threatens to leave us with suspension of judgment about the minds of others, where trust in our natural instincts about what they see, want and know would have been more helpful, and would



have enabled us to make better predictions. The fact that we are capable of applying extra caution and retreating from an ascription of knowledge down to an ascription of mere belief does not entail that ascribing mere belief (plus some environmental conditions about the truth of the proposition believed) is always the more accurate thing to do.<sup>28</sup>

Furthermore, the defender of knowledge-based mental state ascriptions can point out that under a more consistently applied skepticism, ascriptions of belief would themselves become problematic. The same forces that naturally generate doubts about whether Janet knows that her phone is in her bag can equally drive us to wonder whether she even believes as much: if we think about the range of possible attitudes she might have on this question, and especially if we do not allow ourselves to assume that her beliefs are a function of what a person in Janet's circumstances would ordinarily know, then the question of what she should be seen as believing becomes very hard to answer. There are many circumstances in which careful reflective re-thinking of our initial intuitive attitudes may well be warranted, but it is probably not a good idea, in our efforts to understand other agents, to aim for pure reflection without reliance on any intuitive input.

This paper has argued that the identification of knowledge as a mental state is one of the central principles of our intuitive mindreading system. Taking a non-skeptical attitude towards that system means agreeing that when we are guided in our understanding of others by an intuitive sense of what they know, we gain significant insight into what they do. If the test of a theory is its capacity to generate good predictions and explanations, then anyone with a generally non-skeptical attitude towards intuitive mindreading should see the thesis that knowledge is a mental state as well confirmed. This is not to say that we have to accept at face value all of our intuitive impressions about the minds of others: like our perceptual systems, our mindreading capacities have certain natural limitations. These limitations—the bias towards egocentrism, for example—are subject to reflective correction, but such corrections can be better executed in light of a clear understanding of the core principles of the natural system that gives us access to other minds.<sup>29</sup>

---

<sup>28</sup> For further discussion of this point, see (Nagel, 2011).

<sup>29</sup> Thanks to Patrick Rysiew, Jonathan Weisberg, and members of the Van Leer Institute Epistemology group for discussion of the issues in this paper. Special thanks to Jane Friedman, Tamar Gendler, Sergio Tenenbaum, Tim Williamson, and an anonymous referee for this journal for helpful comments on an earlier draft.

## References

- Abbeduto, L., & Rosenberg, S. (1985). Children's knowledge of the presuppositions of 'know' and other cognitive verbs. *Journal of Child Language*, 12(03), 621-641.
- Apperly, I. (2011). *Mindreaders: The Cognitive Basis of "Theory of Mind"*. Hove and New York: Psychology Press.
- Apperly, I., & Butterfill, S. (2009). Do humans have two systems to track beliefs and belief-like states? *Psychological Review*, 116(4), 953-970.
- Apperly, I., Riggs, K., Simpson, A., Chiavarino, C., & Samson, D. (2006). Is belief reasoning automatic? *Psychological Science*, 17(10), 841.
- Apperly, I., & Robinson, E. (2003). When can children handle referential opacity? Evidence for systematic variation in 5-and 6-year-old children's reasoning about beliefs and belief reports. *Journal of Experimental Child Psychology*, 85(4), 297-311.
- Baron-Cohen, S., Ring, H., Moriarty, J., Schmitz, B., Costa, D., & Ell, P. (1994). Recognition of mental state terms. Clinical findings in children with autism and a functional neuroimaging study of normal adults. *The British Journal of Psychiatry*, 165(5), 640.
- Bartsch, K., & Wellman, H. (1995). *Children talk about the mind*. New York: Oxford University Press.
- Bennett, J. (1978). Some remarks about concepts. *Behavioral and Brain Sciences*, 1(04), 557-560.
- Brauer, J., Call, J., & Tomasello, M. (2008). Chimpanzees do not take into account what others can hear in a competitive situation. *Animal Cognition*, 11(1), 175-178.
- Brueckner, A. (2002). Williamson on the primeness of knowing. *Analysis*, 62(275), 197-202.
- Call, J., & Tomasello, M. (1999). A nonverbal false belief task: The performance of children and great apes. *Child Development*, 70(2), 381-395.
- Call, J., & Tomasello, M. (2008). Does the chimpanzee have a theory of mind? 30 years later. *Trends in cognitive sciences*, 12(5), 187-192.
- Carruthers, P. (2009). How we know our own minds: the relationship between mindreading and metacognition. *Behavioral and Brain Sciences*, 32(02), 121-138.
- Choi, S., & Gopnik, A. (1995). Early acquisition of verbs in Korean: A cross-linguistic study. *Journal of child language*, 22(03), 497-529.
- Churchland, P. M. (1981). Eliminative materialism and the propositional attitudes. *The Journal of Philosophy*, 78(2), 67-90.
- Custer, W. L. (1996). A comparison of young children's understanding of contradictory representations in pretense, memory, and belief. *Child Development*, 67(2), 678-688.
- Dancy, J. (2000). *Practical reality*: Oxford University Press, USA.
- Davidson, D. (1963). Actions, reasons, and causes. *The Journal of Philosophy*, 685-700.
- Davies, M., & Stone, T. (2001). Mental simulation, tacit theory, and the threat of collapse. *Mental*, 29(1&2), 127-173.
- De Villiers, J. (2007). The interface of language and Theory of Mind. *Lingua*, 117(11), 1858-1878.
- Dennett, D. C. (1971). Intentional systems. *The Journal of Philosophy*, 68(4), 87-106.
- Dennett, D. C. (1978). Beliefs about beliefs. *Behavioral and Brain Sciences*, 1, 568-570.
- Dienes, Z., & Perner, J. (1999). A theory of implicit and explicit knowledge. *Behavioral and Brain Sciences*, 22(05), 735-808.
- Epley, N., & Waytz, A. (2009). Mind perception. In S. T. Fiske & D. T. Gilbert & G. Lindsay (Eds.), *The Handbook of Social Psychology, 5th Edition* (pp. 498 - 541). New York: Wiley.
- Fletcher, L., & Carruthers, P. (2012). Behavior-Reading versus Mentalizing in Animals. In J. Metcalfe & H. Terrace (Eds.), *Agency and Joint Attention*. New York: Oxford.
- Fodor, J. A. (1987). *Psychosemantics: The problem of meaning in the philosophy of mind*: Oxford, England: British Psychological Society; Cambridge, MA, US: The MIT Press.
- Fricker, E. (2009). Is Knowing a State of Mind? The Case Against. In P. Greenough & D. Pritchard (Eds.), *Williamson on Knowledge* (pp. 31-60). New York: Oxford University Press.
- Gallese, V., & Goldman, A. (1998). Mirror neurons and the simulation theory of mind-reading. *Trends in cognitive sciences*, 2(12), 493-501.
- Gettier, E. L. (1963). Is Justified True Belief Knowledge? *Analysis*, 23, 121-123.
- Gillette, J., Gleitman, H., Gleitman, L., & Lederer, A. (1999). Human simulations of vocabulary learning. *Cognition*, 73, 135-176.

- Goldman, A. (1967). A Causal Theory of Knowing. *The Journal of Philosophy*, 64(12), 357-372.
- Goldman, A. (2006). *Simulating minds: The philosophy, psychology, and neuroscience of mindreading*. New York: Oxford University Press.
- Gopnik, A., & Wellman, H. (1992). Why the child's theory of mind really is a theory. *Mind & Language*, 7(1 2), 145-171.
- Gordon, R. (1986). Folk psychology as simulation. *Mind & Language*, 1(2), 158-171.
- Grice, H. P. (1975). Logic and conversation. *Syntax and Semantics*, 3, 41-58.
- Hare, B., Call, J., & Tomasello, M. (2001). Do chimpanzees know what conspecifics know? *Animal Behaviour*, 61(1), 139-151.
- Harman, G. (1978). Studying the chimpanzee's theory of mind. *Behavioral and Brain Sciences*, 1(04), 576-577.
- Hazlett, A. (2010). The myth of factive verbs. *Philosophy and Phenomenological Research*, 80(3), 497-522.
- Heyes, C. M. (1998). Theory of mind in nonhuman primates. *Behavioral and Brain Sciences*, 21(01), 101-114.
- Hogrefe, G. J., Wimmer, H., & Perner, J. (1986). Ignorance versus false belief: A developmental lag in attribution of epistemic states. *Child Development*, 567-582.
- Holton, R. (1997). Some telling examples: A reply to Tsohatzidis. *Journal of pragmatics*, 28(5), 625-628.
- Hubbard, T. L. (1995). Environmental invariants in the representation of motion: Implied dynamics and representational momentum, gravity, friction, and centripetal force. *Psychonomic Bulletin & Review*, 2(3), 322-338.
- Jackson, F. (2007). Is belief an internal state? *Philosophical Studies*, 132(3), 571-580.
- Johnson, C. N., & Wellman, H. M. (1980). Children's developing understanding of mental verbs: Remember, know, and guess. *Child Development*, 51(4), 1095-1102.
- Jones, E. E., & Harris, V. A. (1967). The attribution of attitudes. *Journal of Experimental Social Psychology*, 3(1), 1-24.
- Kaminski, J., Call, J., & Tomasello, M. (2008). Chimpanzees know what others know, but not what they believe. *Cognition*, 109(2), 224-234.
- Keysar, B. (2007). Communication and miscommunication: The role of egocentric processes. *Intercultural Pragmatics*, 4(1), 71-84.
- Koriat, A. (1995). Dissociating knowing and the feeling of knowing: Further evidence for the accessibility model. *Journal of Experimental Psychology: General*, 124(3), 311-333.
- Kornblith, H. (2004). *Knowledge and its Place in Nature*. Oxford University Press, USA.
- Kozhevnikov, M., & Hegarty, M. (2001). Impetus beliefs as default heuristics: Dissociation between explicit and implicit knowledge about motion. *Psychonomic Bulletin & Review*, 8(3), 439.
- 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.
- Lin, S., Keysar, B., & Epley, N. (2010). Reflexively mindblind: Using theory of mind to interpret behavior requires effortful attention. *Journal of Experimental Social Psychology*, 46(3), 551-556.
- Magnus, P., & Cohen, J. (2003). Williamson on knowledge and psychological explanation. *Philosophical Studies*, 116(1), 37-52.
- McCloskey, M., & Kohl, D. (1983). Naive physics: The curvilinear impetus principle and its role in interactions with moving objects. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 9(1), 146-156.
- McDowell, J. (1986). Singular thought and the extent of inner space. In J. McDowell & P. Pettit (Eds.), *Subject, Thought and Context* (pp. 137-168). Oxford: Clarendon Press.
- Miller, S., Hardin, C., & Montgomery, D. (2003). Young children's understanding of the conditions for knowledge acquisition. *Journal of Cognition and Development*, 4(3), 325-356.
- Miscione, J. L., Marvin, R. S., O'Brien, R. G., & Greenberg, M. T. (1978). A Developmental Study of Preschool Children's Understanding of the Words "Know" and "Guess". *Child Development*, 49(4), 1107-1113.
- Molyneux, B. (2007). Primeness, Internalism and Explanatory Generality. *Philosophical Studies*, 135(2), 255-277.
- Moore, C., Bryant, D., & Furrow, D. (1989). Mental terms and the development of certainty. *Child Development*, 60(1), 167-171.
- Mossler, D. G., Marvin, R. S., & Greenberg, M. T. (1976). Conceptual perspective taking in 2-to 6-year-old children. *Developmental Psychology*, 12(1), 85.
- Nagel, J. (2011). The psychological basis of the Harman-Vogel paradox. *Philosophers' Imprint*, 11(5), 1-28.

- Nagel, J. (forthcoming). Intuitions and Experiments: A defense of the case method. *Philosophy and Phenomenological Research*.
- Nichols, S., & Stich, S. (2003). *Mindreading: An integrated account of pretence, self-awareness, and understanding other minds*: Oxford University Press, USA.
- Okamoto-Barth, S., Call, J., & Tomasello, M. (2007). Great apes' understanding of other individuals' line of sight. *Psychological Science, 18*(5), 462.
- Onishi, K. H., & Baillargeon, R. (2005). Do 15-month-old infants understand false beliefs? *Science, 308*(5719), 255.
- Papafragou, A., Cassidy, K., & Gleitman, L. (2007). When we think about thinking: The acquisition of belief verbs. *Cognition, 105*(1), 125-165.
- Perfect, T. J., Watson, E. L., & Wagstaff, G. F. (1993). Accuracy of confidence ratings associated with general knowledge and eyewitness memory. *Journal of Applied Psychology, 78*(1), 144.
- Perner, J. (1991). *Understanding the Representational Mind*: MIT Press Cambridge, MA.
- Perner, J. (1995). The many faces of belief: Reflections on Fodor's and the child's theory of mind. *Cognition, 57*(3), 241-269.
- Perner, J. (2010). Who took the cog out of cognitive science? Mentalism in an era of anti-cognitivism. In P. French & R. Schwarzer (Eds.), *Cognition and neuropsychology: International perspectives on psychological science* (Vol. 1, pp. 241-261). New York: Psychology Press.
- Pillow, B. H., Hill, V., Boyce, A., & Stein, C. (2000). Understanding inference as a source of knowledge: Children's ability to evaluate the certainty of deduction, perception, and guessing. *Developmental Psychology, 36*(2), 169.
- Plantinga, A. (1993). *Warrant: The current debate*. New York: Oxford University Press.
- Plato. (1990). *The Theaetetus of Plato* (M. J. Levett, Trans.). Indianapolis: Hackett.
- Povinelli, D. J., & Eddy, T. J. (2000). *What young chimpanzees know about seeing*. New York: Wiley-Blackwell.
- Povinelli, D. J., Rulf, A. B., & Bierschwale, D. T. (1994). Absence of knowledge attribution and self-recognition in young chimpanzees. *Journal of Comparative Psychology, 108*(1), 74.
- Povinelli, D. J., & Vonk, J. (2003). Chimpanzee minds: suspiciously human? *Trends in cognitive sciences, 7*(4), 157-160.
- Premack, D., & Woodruff, G. (1978). Does the chimpanzee have a theory of mind? *Behavioral and Brain Sciences, 1*(04), 515-526.
- Putnam, H. (1975). *Mind, language and reality*. Cambridge: Cambridge University Press.
- Reber, R., & Unkelbach, C. (2010). The Epistemic Status of Processing Fluency as Source for Judgments of Truth. *Review of Philosophy and Psychology, 1*-19.
- Ross, L. (1977). The intuitive psychologist and his shortcomings: Distortions in the attribution process. *Advances in experimental social psychology, 10*, 173-220.
- Ruffman, (1996). Do children understand the mind by means of simulation or a theory? Evidence from their understanding of inference, *Mind language* (Vol. 11, pp. 388). Oxford: B. Blackwell.
- Russell, J. (2005). Justifying all the fuss about false belief. *Trends in cognitive sciences, 9*(7), 307-308.
- Samson, D., Apperly, I., Braithwaite, J., Andrews, B., & Scott, S. (2010). Seeing it their way. *Journal of Experimental Psychology-Human Perception and Performance, 36*(5), 1255-1266.
- Saxe, R. (2005). Against simulation: the argument from error. *Trends in cognitive sciences, 9*(4), 174-179.
- Shatz, M., Wellman, H. M., & Silber, S. (1983). The acquisition of mental verbs: A systematic investigation of the first reference to mental state. *Cognition, 14*(3), 301-321.
- Shope, R. K. (1983). *The analysis of knowing: A decade of research*: Princeton University Press.
- Slooman, S. (1996). The empirical case for two systems of reasoning. *Psychological Bulletin, 119*(1), 3-22.
- Sodian, B., & Kristen, S. (2010). Theory of mind. In B. M. Glatzeder & V. Goel & A. von Muller (Eds.), *Towards a Theory of Thinking* (pp. 189-201). Berlin: Springer-Verlag.
- Sodian, B., Thoermer, C., & Dietrich, N. (2006). Two-to four-year-old children's differentiation of knowing and guessing in a non-verbal task. *European Journal of Developmental Psychology, 3*(3), 222-237.
- Sodian, B., & Wimmer, H. (1987). Children's understanding of inference as a source of knowledge. *Child Development, 58*(2), 424-433.
- Sosa, E. (2007). *A virtue epistemology: Apt belief and reflective knowledge* (Vol. 1): Oxford University Press, USA.
- Stanley, J. (2011). *Know How*. New York: Oxford University Press.
- Stanley, J., & Williamson, T. (2001). Knowing how. *The Journal of Philosophy, 98*(8), 411-444.

- Steup, M. (2008). The analysis of knowledge. *Stanford Encyclopedia of Philosophy*, <http://plato.stanford.edu/entries/knowledge-analysis/>.
- Stich, S., & Ravenscroft, I. (1994). What is folk psychology? *Cognition*, 50(1-3), 447-468.
- Tardif, T., & Wellman, H. M. (2000). Acquisition of mental state language in Mandarin-and Cantonese-speaking children. *Developmental Psychology*, 36(1), 25.
- Tomasello, M., & Call, J. (1997). *Primate cognition*. New York: Oxford University Press.
- Tomasello, M., Call, J., & Hare, B. (1998). Five primate species follow the visual gaze of conspecifics. *Animal Behaviour*, 55(4), 1063-1069.
- Turri, J. (2011). Manifest failure: The Gettier problem solved. *Philosophers' Imprint*, 11(8), 1-11.
- Unger, P. (1971). A defense of skepticism. *The Philosophical Review*, 80(2), 198-219.
- Wellman, H. M., & Liu, D. (2004). Scaling of Theory of Mind Tasks. *Child Development*, 75(2), 523-541.
- Williamson, T. (1995). Is knowing a state of mind? *Mind*, 104(415), 533.
- Williamson, T. (2000). *Knowledge and its Limits*. New York: Oxford University Press.
- Williamson, T. (2009). Replies to Critics. In P. Greenough & D. Pritchard (Eds.), *Williamson on Knowledge* (pp. 279-284). New York: Oxford University Press.
- Wimmer, H., & Perner, J. (1983). Beliefs about beliefs: Representation and constraining function of wrong beliefs in young children's understanding of deception. *Cognition*, 13(1), 103-128.
- Woolfe, T., Want, S. C., & Siegal, M. (2002). Signposts to development: Theory of mind in deaf children. *Child Development*, 73(3), 768-778.
- Woolley, J., & Bruell, M. (1996). Young children's awareness of the origins of their mental representations. *Developmental Psychology*, 32(2), 335-346.
- Zagzebski, L. (1994). The inescapability of Gettier problems. *The Philosophical Quarterly*, 44(174), 65-73.
- Zwaan, R. A., & Radvansky, G. A. (1998). Situation models in language comprehension and memory. *Psychological Bulletin*, 123(2), 162.