Peter Carruthers

How Mindreading Might Mislead Cognitive Science

Abstract: *This article explores three ways in which a cognitively entrenched mindreading (or 'theory of mind') system may bias our thinking as cognitive scientists. One issues in a form of tacit dualism, impacting scientific debates about phenomenal consciousness. Another leads us to think that our own minds are easier to know than they really are, influencing debates about self-knowledge, and about mindreading itself. And the third results in a bias in favour of empiricist over nativist accounts of cognitive development. The discussion throughout is tentative and speculative, and can be regarded as an appeal for caution, as well as a call for further research.*

Keywords: consciousness; dualism; empiricism; innateness; self-knowledge; tacit theory; theory of mind.

1. Introduction

It is a truism that humans are deeply social — indeed, hypersocial — beings. We live together in large groups; we engage in joint actions with others; we cooperate not only with friends but with strangers; we learn most of what we know from other people; and we spend large portions of each day interacting with others, both verbally and non-verbally. It isn't controversial to claim that one of the main capacities that undergirds and enables our hypersociality (in addition to language and various forms of social motivation) is our theory-of-mind ability — or *mindreading*, as I prefer to call it. Everyone can agree that, whatever their developmental origins, mindreading capacities are both deeply ingrained and ubiquitously employed. Indeed, our social world

Correspondence:
Email: pcarruth@umd.edu

is awash with representations of mentality. This is why those with autism spectrum disorder, who have difficulties with spontaneous mindreading, find the social world so difficult to navigate. Thus it isn't a matter of dispute that mindreading capacities are a central component of normal human cognitive functioning.

Our focus here, however, will be on the potential downsides of mindreading's centrality, rather than its benefits for human social life. For the central role of mindreading in human cognition may serve to bias our thinking across a number of different domains of enquiry. Some of these biases are familiar, and have been much discussed; some are less so. In what follows I propose to concentrate on biases that mindreading may introduce into debates in recent philosophy and cognitive science, in particular; focusing especially on the question whether the same or distinct properties of the mindreading system would explain them.

It has long been recognized that mindreading might bias us in some areas of cognitive science, of course. Specifically, the propensity of the mindreading system to respond automatically to cues of agency with attributions of mental states can lead to anthropomorphism (Morgan, 1894; Galef, 1996; Penn, Holyoak and Povinelli, 2008). Comparative psychologists are well aware of this bias, and are thus careful to seek independent evidence before attributing mental states and mental processes of any given sort to a non-human animal. I shall be arguing that something similar may also need to be guarded against in other domains of cognitive science.

One biasing effect that won't be discussed here (except briefly now, in this introduction) is quite narrow in scope. It derives from what is arguably the mistaken way that the mindreading system represents search-motivating mental states like curiosity. It seems that the system treats curiosity as a complex state, built out of simpler components, such as WANT and KNOW. This may be why almost all philosophers and cognitive scientists who have written on the topic of curiosity have addressed it in metacognitive terms — as involving a desire for knowledge or true belief, or as an intrinsic motivation to learn, or something of the sort.[1] Even Loewenstein's (1994) well-known 'information gap' theory of curiosity, which *sounds* as if it might not

[1] See Foley (1987), Goldman (1999), and Williamson (2000), among philosophers; and see Litman (2005), Gruber, Gelman and Ranganath (2014), Blanchard, Hayden and Bromberg-Martin (2015), and Kidd and Hayden (2015), among psychologists.

require metacognition, is actually framed in metacognitive terms. Curiosity is said to arise from 'a discrepancy between what one knows and what one *wishes to know*' (*ibid.*, p. 93, emphasis added).

As a result, when cognitive scientists find evidence of flexibly controlled information seeking in primates (Rosati and Santos, 2016) or infants (Goupil, Romand-Monnier and Kouider, 2016), they claim to have found evidence of metacognition. Building on the work of Whitcomb (2010) and Friedman (2013), in contrast, I have argued that curiosity itself (as opposed to our folk concept thereof) is best understood as one of a number of first-order *questioning attitudes* (Carruthers, 2018; 2019a). Curiosity is a motivational, affective, state that is caused in part by ignorance (without representing ignorance) and embeds a question as its content. So a monkey that observes some food being placed in one of three receptacles (but without seeing which) is then motivated by a state with the first-order content, *where the food is*. The monkey is motivated to walk around examining the insides of the receptacles until it gets an answer to that question (or gives up and does something else), but not because it *wants to know*. Or so I argue. If I am right, then cognitive scientists have been misled by the fact that the mindreading system conceptualizes curiosity and related attitudes incorrectly.

In what follows I shall concentrate on three topics of more general significance for science, teasing out the commonalities and differences between them, and asking whether they admit of a common kind of explanation or require distinct types of explanation. (I shall argue that the explanations are likely to be largely distinct.) One topic is the role of the mindreading faculty in helping to make Cartesian dualism seem intuitive, and the influence this has had on recent scientific debates about consciousness. Another concerns the role of mindreading in lending support to the primacy of the first person in our accounts of how we know of the mental states of ourselves and others, as well as in making the very idea of unconscious mentality seem initially counter-intuitive. And the third (which is the least familiar of the three) will be about the support seemingly provided to empiricist (or 'blank slate') conceptions of the human mind.[2]

[2] Empiricism is here contrasted with nativism, which is a belief in innate knowledge and/or innately structured domain-specific learning mechanisms. The notion of innateness is itself a matter of dispute, of course (Ariew, 1996; Griffiths, 2001; Samuels, 2007). For present purposes I shall simply assume that innate properties are *unlearned*. Everyone allows that some properties of the mind are innate in this sense. In particular, general-

I should stress that my conclusions will be tentative. In each case more evidence is needed to demonstrate that mindreading really does have the biasing effects suggested. And in each case, too, my proposed explanation of those effects (assuming them to be real) is speculative — albeit plausible, I hope. In fact, my goal is to provide what are, at best, 'how-probably' explanations of the biases in question (Craver, 2006). Overall, the discussion is intended to be more exploratory than definitive, and can be considered as a call for additional empirical research on the topic as well as an appeal for caution in our handling of the issues on which we may be biased.

We should note at the outset, however, that there are general reasons for expecting that mindreading might bias our scientific thinking about the mind. For we know that naïve theories about a whole range of different subject matters can continue to exist alongside scientific ones (at least implicitly), slowing reaction times to questions and issuing in erroneous answers under pressure (Shtulman and Valcarcel, 2012). Indeed, even in the domain of physical laws of motion and inertia, where Newtonian principles have been well-established for centuries, people continue to have Aristotelian-like intuitions about the forces at work — even when those people are themselves physics undergraduates (Clement, 1982; McCloskey, 1983). How much more likely is it, then, that tacit psychological theories might have an influence on scientific thinking about matters that are still contentious (at least to some degree), and that belong to the comparatively young cognitive sciences?[3]

Let me emphasize that what are at stake in this discussion are tacit biases, not determining causes. They result in people giving more credence to certain ideas than the evidence warrants, or than *would* be given to those ideas were it not for the unconscious influence of the

learning mechanisms are. The real debate concerns the nature of learning itself (whether domain-general or domain-specific), and whether some concepts and/or beliefs are wholly unlearned.

[3] Talk of 'tacit psychological theories' here needn't be interpreted as a strong commitment to so-called theory-theory as against a simulation theory of mindreading abilities (Carruthers and Smith, 1996). In fact, debates between theory-theory and simulationist accounts of the origins of those abilities are largely irrelevant at this point. In part this is because most people in the field now think that both things play a role, albeit at different phases of development and with differing emphases (Nichols and Stich, 2003; Goldman, 2006; Carruthers, 2013). But it is also because even theorists like Goldman (2006), who think that simulation is basic, acknowledge that development rapidly issues in a body of implicit theory-like generalizations about the mind.

mindreading faculty. Note that this can happen even when the idea in question is explicitly rejected. This means that it can be challenging to detect when a bias has, or has not, been operative. (Hence the need for more evidence. Some possible tests will be mentioned in the Conclusion.) What I hope to show is that it is *plausible* that the mindreading system biases our thinking as cognitive scientists in certain general respects. As a result, we should be careful to control for potential bias in our work on the topics in question, just as comparative psychologists routinely do.

2. Cartesian Dualism and the Problem of Consciousness

Until very recently, with the advent of modern science, all humans in all cultures have believed in the ontological separation of mind and body (Boyer, 2001; Cohen *et al*., 2011; Roazzi, Nyhof and Johnson, 2013). They have thought that mental properties can be tokened independently of physical ones, and most have believed that the self is a distinct thing, independent of the body. As a result, beliefs in some kind of afterlife have been rife — whether merely spiritual in form, or involving resurrection of one's original body, or through reincarnation into a distinct body. The question is: where do these beliefs come from, and why are they a human universal?[4]

One possibility is motivated reasoning in the service of terror management. Humans are aware of their own mortality, and find the prospect of their own non-existence terrifying. This fear can be significantly reduced if one believes that bodily death does not — or at any rate, *need* not — mean the end of one's existence as a mind or self. As a result, belief in the separation of mind and body, once culturally introduced, may quickly spread and fix itself in the population. This suggestion is similar to claims that some theorists have made to explain the universality of language, for example (Tomasello, 2008; Heyes, 2018). Once introduced into a culture, belief in an afterlife (like language) proves too useful ever to be lost. And those benefits mean that belief in an afterlife (like language) spreads quickly

[4] As with other human universals (Brown, 1991), this doesn't mean that every individual human has believed in the separation of mind and body and the possibility of life after death. Rather, these beliefs are universal in the same sense that mindreading itself and sensitivity to rhythm are universal: they are possessed by almost all individuals across all or almost all cultures.

among cultures whenever there is cultural contact (Barlev and Shtulman, under review).

There may be some element of truth in the terror-management account. But it seems unlikely to be the whole story, or even the most important part of the story. This is because there is evidence that young children believe in the separation of mind and body. They think it might be possible for people to switch bodies, as well as accepting the possibility of mental existence after the death of the body (Bering and Bjorklund, 2004; Chudek *et al*., 2018). But this is at ages when it simply isn't plausible to believe that they have become gripped by the prospect of their own deaths. It might be replied, of course, that these beliefs could still have been acquired from the surrounding culture. It may be that the terror-management explanation is a distal one (applying only to the initial origins of these beliefs) rather than proximal. Young children might believe that minds are separate from bodies because all the adults around them do too.

There are a number of reasons for thinking that this alternative cultural-learning explanation is incorrect, however (or is insufficient by itself to explain the data). One is that children's belief in the possibility of afterlife (at least in Western cultures) gets *weaker* with age, not stronger (Bering, 2006). This is the reverse of what one might expect if children were simply acquiring that belief from the people around them. Rather, it suggests that belief in the separation of mind and body is some sort of default among humans, which gets *undermined* through cultural learning in scientifically influenced cultures — at least at the level of explicit verbally-expressed belief, while continuing to exist implicitly alongside those beliefs (Riekki, Lindemann and Lipsanen, 2013; Willard and Norenzayan, 2013; Forstmann and Burgmer, 2015).

A better explanation of the universality of ontological dualism is that it derives from a clash between two bodies of 'core knowledge' (Spelke and Kinzler, 2007), which are either innate or innately channelled, or at least are learned by all normal children early in development independently of cultural input (Bloom, 2004). Both bodies of knowledge seem to emerge quite early in development (Baillargeon *et al*., 2012; Baillargeon, Scott and Bian, 2016). This suggests that, even if they are acquired using general-learning mechanisms rather than being innate or innately channelled, these forms of core knowledge are largely independent of cultural input and cultural learning. One is a set of common-sense physical principles (one object can't pass through another; an object can't move from one

place to another without moving through any intervening places; and so on). The other is a system for representing the mental states of oneself and others, which makes no commitments regarding the physical nature of those states.[5]

What *sort* of clash between mental and physical knowledge are we talking about, however? How, in more detail, are the differences between them supposed to issue in a believed ontological separation between the two? Each operates in accordance with its own set of generalizations and explanatory principles, for sure. But why should this lead us to think that their domains are so different that either one can exist in the absence of the other? For naïve physics and biology, too, utilize distinct sets of generalizations, but this doesn't lead to an ontological separation. No doubt this is because living things are *also* physical things — they can't pass through one another, for example, and can't move from one place to another discontinuously. But that just emphasizes the puzzle: since all the agents we know of are also physical things, why don't we see agency as necessarily tied to the physical?

Agency is at least *causally* tied to the physical, of course. We know that we see things by opening our eyes, that hearing depends on the impact of sound on our ears, and that felt touch requires physical contact with our bodies. Moreover, we know that we can effect change in the physical world by deciding to move our limbs. But interactions among our mental states themselves are *sui generis*, conforming to none of our familiar models of physical causality. Paradigm cases of everyday physical causation are mechanical, involving pushing, pulling, and motion in space. But nothing remotely like that holds for mental-to-mental causal relations. When a perceptual state sparks a memory, or a thought makes one sad, these causal relations strike us as *immediate*, and certainly not as involving anything like mechanical contact.

Even more importantly, our intuitive psychology seems not to require that mental states occupy positions in space, which is the defining feature of physical objects and physical processes. Indeed, it comes close to entailing that mental states *don't* occupy spatial

5 Indeed, there is evidence that infants at seven months of age think that agents aren't subject to ordinary physical principles (Kuhlemeier, Bloom and Wynn, 2004). They seem to think that an agent can move from one place to another without traversing through the spaces in between, whereas an ordinary physical object can't.

positions (McGinn, 1991). For instance, statements such as 'My thought about my mother is two inches behind my right eye' seem hardly even to make sense. The physical subject or bearer of mental states is believed to be spatially located, of course. My thought about my mother is located wherever *I* am located. And most people in scientific cultures know enough about the brain to realize that it has a special role to play. But even this is apt to be expressed in causal rather than constitutive terms ('Something happening in my brain caused me to think it' rather than 'An activity of a specific physical network in my brain *is* me thinking it').

It seems, then, that there is a clash between our core knowledge of physics and our core knowledge of psychology, which gets set up in terms of a contrast between a world of space-occupying objects, events, and causes, on the one hand, and a set of apparently *non*-spatial mental states, on the other. This makes it entirely natural to think that there is a deep ontological separation between them.[6] This expectation may be merely tacit initially, but will rapidly transition into explicit dualist beliefs in cultures that articulate them. Such beliefs might be expected to exert a deep 'attractor effect' on cultural evolution, being sustained and transmitted both because of their apparent *naturalness* given the underlying core-knowledge clash, and because of their terror-management roles of the sort discussed earlier.

There are good scientific reasons to reject ontological dualism, of course, no matter whether what is in question is substance dualism or a dualism of properties. For we have every reason to think that the physical world is causally closed. That is to say: every physical event has a sufficient physical cause. The search for physical mechanisms in nature has been the guiding assumption of science for centuries, and seems amply supported by the resulting scientific successes. But it means that mental events can't be non-physical ones if they are to

6 Note that the separation here is ontological, not functional. Barlev and Shtulman (under review) go wrong on just this point in trying to develop an argument against any sort of inherent dualism. Of course, everyone (even little infants) knows that mental and physical states interact with one another causally, and in rich ways, as we noted earlier. The real point is that intuitive psychology doesn't conceptualize causal relations among mental states themselves in physical or mechanical terms, and that it isn't committed to physical locations for those states. Moreover, since all the modes of interaction with the physical world recognized by intuitive psychology involve the body, it is hardly surprising that many beliefs about the spirit-world involve agents with body-like properties, such as capacities to see and to hear (as is also pointed out by Barlev and Shtulman, under review).

have an impact on physical behaviour. If we act as we do because we think and want what we do, this means that our thoughts and desires must somehow be physical. As a result, most cognitive scientists have embraced physicalism about the mind. Nevertheless, we should expect that tacit dualism might continue to exist alongside these explicit physicalist beliefs, and might bias scientific thinking in some other way, or in some other domain.

One such bias arguably concerns the debate over phenomenal consciousness and qualia. Here philosophers have developed a battery of thought experiments in support of the conclusion that phenomenal properties aren't physical ones. These include the conceivability of zombies, colour-deprived Mary, and the so-called 'explanatory gap' (Jackson, 1982; Chalmers, 1996). What is remarkable is that many cognitive scientists take these thought experiments seriously. As a result, most are careful to talk about the neural *correlates* of phenomenal consciousness rather than the neural *nature* or the neural *realizer* of consciousness (Rees, Krieman and Koch, 2002; Tononi and Koch, 2008). I know of no other domain in which scientists allow their conclusions to be influenced by philosophical thought experiments in this way. It cries out for explanation. But we now have an explanation ready to hand: it is because of the scientists' own *tacit* dualism, which makes the philosophers' arguments seem more scientifically acceptable than they actually are (Carruthers, 2019b).

It might be replied that the explanation should really run in reverse: it is because people are convinced of the explanatory gap that dualism about the mind seems plausible. There are a number of reasons for thinking this isn't so, however. One is the sheer implausibility of believing that hunter-gatherers and illiterate subsistence farmers throughout history and all over the world have been influenced in their dualist beliefs by consideration of an explanatory gap. Another is the very young age at which dualist beliefs become manifest. And yet another is that both children and adults are more ready to think that attitude states like beliefs and values might survive the death of an agent than they are to think that phenomenal experiences could (Bering and Bjorklund, 2004).[7] This suggests that our tacit dualism

[7] This finding, in particular, is problematic for the argument of Robbins and Jack (2006), who claim that dualism is only intuitively appealing for phenomenal mental states like perceptual experiences and bodily feelings. In their view, both dualism and the problem of consciousness are a product of first-person empathic identification with such states in other people, which makes it hard to conceive of them in physical terms.

about the mind doesn't arise from considerations having to do with phenomenal consciousness specifically.

But if tacit dualism arises for mental states generally, then why is it that cognitive science isn't systematically biased in a dualist direction, but only for conscious experiences specifically? It may well be generally biased; indeed, I suggest that it is. But there are no arguments now remaining in circulation in support of generalized dualism, to be given more credence than they deserve. As noted above, there are powerful scientific reasons for rejecting dualism. And it is only with respect to phenomenally conscious states in particular that philosophers have developed their pro-dualist thought experiments. Moreover, such states, and such thought experiments, are arguably restricted to states that have non-conceptual contents, and hence are broadly experiential (Carruthers and Veillet, 2017). So it is only in this one domain of the mind where there remains any kind of support for explicit dualism that our tacit dualism might tempt us into taking seriously — or seriously enough to talk cautiously about 'neural correlates' rather than 'neural natures'.

I propose, then, that the human mindreading faculty not only issues in a tacit ontological dualism, which manifests as explicit dualism in non-scientific cultures and social groups; but tacit dualism, in turn, may continue to exert an influence on contemporary cognitive science, biasing people to take qualia realism more seriously than they otherwise would (or arguably *should*; Carruthers, 2019b).

3. First-Person-First Accounts of Mindreading

Cartesian dualism came paired with a distinctive set of epistemological beliefs in Descartes' actual work, of course (Descartes, 1641). Mental states were thought to be self-presenting (to have them is to know that one has them) and to be infallibly knowable by their possessors (hence being suitable to provide the foundation for all other forms of knowledge). Similar views have been the default in most of the Western philosophical tradition, from Aristotle (Caston, 2002) to Kant (1781), at least. Indeed, they have generally been taken as obvious. Hence Locke (1690) could claim, without evidence or argument, 'there can be nothing in the mind that the mind itself is unaware of'. In fact, it wasn't until a little over a hundred years ago, through the work of Nietzsche (Leiter, 2019) and Freud (Mannoni, 1971), that the idea of unconscious mentality entered seriously into scientific discourse.

Although close to ubiquitous in Western thought, the question whether Cartesian epistemology is a human universal is harder to answer definitively than is the corresponding claim for Cartesian dualism. For it won't be manifest to ordinary anthropological observation (indeed, in some cultures explicit talk about the mind is actively discouraged; Chudek *et al.*, 2018; McNamara *et al.*, 2019), and it is only likely to be explicitly articulated in cultures that have developed some sort of reflective philosophical tradition. Nevertheless, Carruthers (2011) attempts to make that case. Relying on communications from experts in the fields of Buddhist philosophy, the philosophy of ancient China (prior to the arrival of Buddhism), and the philosophy of ancient Mayan cultures, I tentatively claimed that a view of the mind as transparent to itself is apt to emerge as the default whenever people reflect on the matter. Supposing that this is true, we can then ask what might explain it, and whether a tacit commitment to Cartesian epistemological principles has exerted, or continues to exert, an influence on cognitive science.

The reason why people the world over have thought that the mind is, in a sense, transparent to itself certainly isn't because such beliefs are *true*. On the contrary, as Carruthers (2011) reviews at length, we have ample scientific evidence that people often *confabulate* when ascribing mental states to themselves, claiming to have mental states that really they don't. So knowledge of our own mental states certainly isn't infallible — not even close. Carruthers (2011) argues that our access to our own propositional-attitude states (judgments, decisions, and the rest) is no different in principle from our access to the mental states of other people, requiring interpretation of sensorily accessible cues (included among which are our own visual images and inner speech — I also allow that our access to conscious sensory states is transparent and *not* interpretative, albeit not infallible). Moreover, there is even greater evidence that mental states aren't self-presenting. Indeed, almost the entire field of cognitive science is founded on the idea that there are mental states and processes that are inaccessible to their possessors.

It seems unlikely that tacit acceptance of mental self-transparency should have the same source as tacit ontological dualism. It is quite unclear how a clash between our intuitive physics and our intuitive psychology should lead us to think that mental states are self-presenting to their possessors and infallibly knowable by those who have them. It seems, rather, that the latter beliefs must somehow be

motivated by factors internal to the mindreading faculty itself. At any rate, that is what I propose (2008; 2011).

Carruthers (2011) argues on grounds of reverse-engineering that the following two principles are likely to be built into the operations of the mindreading faculty as tacit inference rules:

(1) One believes one is in mental state M → one is in mental state M.
(2) One believes one isn't in mental state M → one isn't in mental state M.

The first will issue in intuitions of infallible knowledge, and the second in the intuition that mental states are always self-presenting to their possessors. Such principles might be built into the mindreading faculty through some sort of innate channelling, or alternatively through general-purpose learning. For the crucial point is that, even if invalid, one might expect that they would be adopted as simplifying heuristics, greatly speeding up the work of the mindreading faculty, and doing so arguably without any loss in overall reliability (in part because of their greater simplicity).[8] This idea will be elaborated on shortly.

Notice first, however, that statements that directly conflict with principles (1) and (2) can strike one as distinctly strange, even for someone like myself who explicitly rejects them both. Thus consider:

(1*) John thinks he has just decided to go to the party, but really he hasn't.
(2*) John thinks he doesn't intend to go to the party, but really he does.

Reading each of these statements is at least *disfluent*. Initially one is at a loss for how to interpret them. One has to remind oneself of the scientific evidence of confabulation to make sense of (1*), and of the scientific evidence of unconscious mental states to make sense of (2*). One potential explanation is the one I suggest: principles (1) and (2) are default inference-rules deployed by the mindreading faculty itself, thereby initially leading (1*) and (2*) to seem problematic.

As for why (1) and (2) should be useful heuristics to employ, notice that a large part of the work of the mindreading faculty lies in

[8] It is now widely accepted, of course, that simple heuristics in reasoning and decision making can often outperform more complex and information-hungry rules (Gigerenzer *et al.*, 1999).

interpreting the speech of other people — figuring out speaker intent, detecting sarcasm and irony, and so on. Moreover, a lot of talk is *about* mental states. People talk about what they want, what they intend, and what they think; as well as what others want, intend, or think.[9] The process of interpretation is greatly simplified and speeded if principles (1) and (2) are adopted. And their adoption is unlikely to be accompanied by any significant loss of reliability. In part this is because people are pretty good interpreters of themselves, relying not just on perceptions of their overt behaviour, but also on conscious phenomena like their own visual imagery and inner speech. So oftentimes people's reports of their beliefs, judgments, and intentions will be *correct*. Moreover, even in those cases where they aren't, people generally feel constrained to act consistently with what they have *said* they want or think (Frankish, 2004; Zawidzki, 2013). So even if an initial self-attribution is false, people may thereafter ensure that they behave as if it were true. And in addition, of course, more complex and information-hungry attribution principles would introduce their own possibilities for error (as well as being a great deal slower).

If tacit inference rules like (1) and (2) develop within the mindreading faculty under normal circumstances, then this would explain why some form of Cartesian epistemology about the mental would seem to be a human universal. It can also perhaps explain why Freud's initial postulation of unconscious mentality (thereby violating principle #2) should have struck people as *deep* (because deeply counter-intuitive but presented as newly-discovered science), despite its lack of actual scientific credentials (Grünbaum, 1984). But at the same time it can explain why cognitive scientists themselves were initially so resistant to the idea of unconscious perception (Weiskrantz, 1986), a resistance that continues in some quarters (Peters & Lau, 2015; Phillips, 2018). Moreover, it can explain why cognitive scientists took so long to accept the reality of unconscious emotion (Winkielman and Berridge, 2004), leading them to set a high bar for acceptance.

It is possible that some sort of tacit Cartesian epistemology also explains the appeal of first-person-first accounts of our knowledge of

[9] Indeed, some theorists suggest that social talk — otherwise known as gossip — accounts for as much as two thirds of overall human conversation (Dunbar, 2004). And Dunbar (1997) even argues that gossip may have been what language evolved for in the first place (enabling informal punishment and social control).

the mental states of other people (Gallese and Goldman, 1998; Goldman, 2006), despite the strength of the case against them (Carruthers, 2011). And it may likewise explain what leads many comparative psychologists to believe they have found evidence of metacognition (or 'self-awareness') in creatures that lack a capacity for equivalently complex forms of mindreading (Couchman *et al.*, 2009; 2012; Smith, Couchman and Beran, 2014) — again, despite the case that can be made against them (Carruthers, 2011; 2014). The claims just appealed to are highly contentious, of course. And I don't intend to pursue the suggestions here. But it is worth noting that one can accept a role for simulation and so-called 'mirror neurons' in social cognition without thinking that we have direct first-person access to the results. Just as language comprehension seems to function in collaboration with language production to help parse and interpret the incoming linguistic signal (Pickering and Garrod, 2013), so one might think that behaviour interpretation generally calls on the resources of one's own inferential, decision making, and motor systems. But none of this need happen in a way that is accessible to consciousness or introspective report. And likewise, one can explain evidence of so-called 'uncertainty monitoring' in monkeys in terms of risk-based prospective reasoning about alternatives, rather than in terms that require them to have introspective knowledge of their own states of uncertainty.

I have suggested in this section that there may be inference-rules built into the content of the mindreading faculty that involve a tacit commitment to a form of Cartesian epistemology of the mental. Whether or not these tacit principles have *actually* biased cognitive scientists in their work is quite hard to establish, of course; and I have made no strong claims to that effect here. What I do suggest, however, is that this is a live — indeed, plausible — possibility; and it is thus one that may need to be guarded against.

4. Empiricist Intuitions

We now turn to consider whether mindreading may exert a biasing influence on the debate between empiricists and nativists, and if so, how. It is a common complaint among nativists that empiricism is

treated — unjustifiably — as the default option.[10] But it is another matter to claim that they are right, of course, and it is even harder to demonstrate that the default in question results from a mindreading-induced bias (for example, it might result rather from the enlightenment ideal of an indefinitely improvable human nature, thus making it the outcome of a kind of social-political bias; see Pinker, 2002). All the same, the idea of such a bias seems well worth exploring. Indeed, a plausible case can be made for its existence, as we will see. And in any case there is no need to assume that just *one* bias supports empiricism over nativism in current cognitive science.

Intuitive psychology recognizes three broad categories of the knowledge acquisition process. One is sensory (vision, hearing, touch, and so on); the second is communication from other knowledgeable individuals; and the third is inference from either or both of the first two sources. Each is understood early in development (at least tacitly), and each has a claim to be among the core components of the mindreading system.[11] Note that all of these kinds of knowledge acquisition are empiricist in nature.

Whenever one represents someone as knowing or believing something, then, the mindreading system will automatically create an expectation that the belief in question was acquired through one of its recognized methods. The suggestion that the belief was *not* acquired in any of those ways (but rather via the neural maturation of a genetically shaped system in the brain that evolved for the purpose) will then seem counter-intuitive. And it will seem that way both because it violates one's prior expectations, and because the resulting theory will strike one as unnecessarily complex (because involving an

10 I was involved in a series of workshops and conferences through the early 2000s designed for cognitive scientists who adopt a broadly nativist approach to their work. (The conferences issued in three volumes of essays on the innate mind: Carruthers, Laurence and Stich, 2005; 2006; 2007.) I can report that resentment at a perceived bias in favour of empiricism was quite widespread among the dozens of cognitive scientists who participated. Indeed, the joke in the group (often told with some bitterness) was that empiricism itself is innate. Whether this perception is really warranted isn't easy to establish, however. For remember, all that is really at stake here is a *bias*, not a determining cause. It can be true that empiricism is given more credence than it should be, or than the evidence warrants, even if a majority of cognitive scientists are actually nativists of one stripe or another.

11 For understanding of vision at 15 months, see Song and Baillargeon (2008); for understanding of touch at 18 months, see Knudsen and Lisztowski (2012); for understanding of communication at 17 months, see Southgate, Chevallier and Csibra (2010); and for understanding of inference at 18 months, see Scott *et al*. (2010).

extra postulate in addition to those already recognized). An inference to the best, *simplest*, explanation will seemingly favour empiricism over nativism, other things being equal.

Note that there is a deep contrast here with the circumstances under which nativist beliefs have been espoused in earlier historical eras. When Plato first postulated the existence of innate knowledge in his dialogue *Meno* he combined it with the idea that such knowledge would nevertheless have been learned through a kind of experience, namely acquaintance with the objects of that knowledge ('the forms') while the non-physical soul pre-existed its human embodiment. This enabled the idea to seem intelligible from the perspective of the mind-reading system itself, arguably making it less counter-intuitive. Likewise, when early-modern rationalist philosophers like Descartes and Leibniz postulated innate knowledge, they thought of this knowledge as having been imprinted on the human soul by God. While not exactly resulting from a kind of testimony (which is one of the main belief acquisition mechanisms recognized by the mindreading system, of course), this does at least treat innate knowledge as resulting from a kind of *informative agency*. The hypothesis can thus be formulated using the vocabulary of the mindreading system itself, and is therefore readily incorporated into the latter, implying that it shouldn't seem deeply counter-intuitive from the latter's perspective.

Consistent with this, the kinds of objections raised against innate knowledge by early-modern empiricists like Locke and Hume seem to have had a different source from the one I am proposing now operates. Instead of an argument from simplicity (grounded in mindreading-induced expectations), their basic motivation stemmed from a commitment to scientific *naturalism*, and a concomitant rejection of any appeals to God or supernatural phenomena in explaining the properties of the natural world (Carruthers, 1992). I suggest that innate knowledge wasn't counter-intuitive to them, given the theories of innateness available at the time, but was rather seen as methodologically unacceptable from the standpoint of the emerging sciences. Contemporary forms of nativism, in contrast, take one outside the framework of the mindreading faculty altogether, postulating belief acquisition methods that are biological rather than psychological in nature. They thereby conflict with the tacit expectations created by that faculty and complicate its structure.

How else might a bias in favour of empiricism be explained? Berent (2020) suggests that it results, rather, from a conflict between people's tacit dualism and their tacit *essentialism* about biological creatures in

general (including ourselves). There is extensive evidence in support of the latter idea (Gelman, 2003). Both children and adults assume that the manifest properties of living things result from some sort of inner core or *essence*, from which all those other properties flow. Innate knowledge, then, as *inborn* knowledge, would seemingly have to belong to that essence. But that would make the knowledge in question biological in nature, and hence physical; thereby conflicting with our tacit dualism about mental states in general. The result is an intuitive pressure against accepting the existence of innate ideas.

This is a possible explanation; and Berent and colleagues have collected evidence supporting the various components of the account, at least (Berent, Platt and Sandoboe, 2019; under review). But it has one significant drawback: supposing that the bias against innate ideas is real, Berent's account requires us to think that the cognitive scientists who are thus biased are conflating *causation* with *constitution*. For nearly everyone has always accepted that there are *causal* relations between body and mind, of course (light entering our eyes causes visual experience; contact with our skin causes us to feel; and so on). And it is but a minor extension of these intuitive beliefs to accept that physical events happening in our brains can have consequences for our minds, as Descartes himself postulated. Indeed, this is now universally accepted, even by cognitive scientists who are ontological dualists (Chalmers, 1996). So there should be no problem, from the perspective of tacit dualism, in accepting that facts concerning our biological 'essence' might explain some of our beliefs. For this can be just another instance physical-to-mental causation.

In order for an essentialist construal of innateness to conflict with our tacit dualism, we would have to think that the properties of the biological essence that result in innate ideas don't just *cause* those ideas but *constitute* them. We would have to think that the biological essence is constitutive of this aspect of our minds, meaning that our innate ideas would themselves be physical. It is possible that contemporary empiricists make exactly this confusion. Indeed, the distinction between *cause* and *constitution* is sometimes overlooked, even by philosophers (Carruthers and Veillet, 2011). But I believe it is better to explain the bias against innateness without attributing mere conceptual confusion to the cognitive scientists in question, if we can. Certainly there is no reason to think that a conflation between causation and constitution is built into either tacit dualism or folk essentialism as such. It is an extra assumption that needs to be added to the account, on top of the other two.

Berent, Platt and Sandoboe (under review) find, however, that people are more willing to accept that mental states like desires and emotions are innate than they are to accept the reality of innate ideas (concepts and/or beliefs). They take this to support their view that people's resistance to innateness stems from a conflict between tacit dualism and biological essentialism, on the grounds that affective mental states are more closely linked to the body. If desires and emotions are seen as more biological in nature than are ideas, then the suggestion that they might belong to our biological 'essence' (that is, be innate) will create less of a conflict with tacit dualism. In effect, the suggestion is that people aren't really tacit dualists about desires and emotions (or are only weakly so), whereas they *are* dualists about concepts and beliefs.

An alternative interpretation of the data is available, however. This is that intuitive psychology doesn't contain a fixed set of modes of desire acquisition, in the way that it has a fixed set of three modes of belief acquisition (personal experience, testimony, and inference). Indeed, attributions of desire and emotion are much more closely tied to the output side of the mind (that is, to behaviour) than to the input side. While we (and human infants) readily attribute beliefs on the basis of what someone has experienced, been told, or inferred from what they have experienced or been told, we (and human infants) mostly attribute desires on the basis of what someone *does*. Admittedly, we do know that someone who hasn't eaten in a while will be hungry, that someone who is fluid-deprived will want to drink, that someone who hasn't slept all night will want to sleep, and so on. But note that in these and similar cases the mechanism of desire-creation is quite naturally seen as innate. Indeed, it is part of common sense that such desires are biologically caused, since they are found in almost all animals. (Note: this is *not* to say that common sense takes them to be biologically *constituted*. That the causal mechanism is biological doesn't mean that the feelings themselves are.)

Moreover, in ordinary life there aren't really any beliefs that people (or other animals) possess for which we can't spin a story about how those beliefs were acquired through one of folk psychology's three modes. This is why nature/nurture debates about particular cases are so hard to resolve, of course. One can always postulate prior experience or something about cultural input to explain the beliefs in question. In contrast, we often have the experience of finding ourselves with a desire without any idea of why we have it. You taste something for the first time, like it, and want more of it, for example.

Why? Sometimes one can cite the presence of a property one already knows one likes ('It is slightly sweet'). But often one can only say, 'I guess that's just me', or, 'I guess I'm just built that way'. That people are more ready to accept that desires can be innate than that beliefs are can thus be explained within the same intuitive-psychological framework I outlined above: intuitive psychology creates an expectation that beliefs will have been empirically acquired; but it contains no such general expectation for desires.

I suggest, then, that if there is a tacit bias in favour of empiricism created by the mindreading system, it most likely has a distinct origin from either of the other two potential biases discussed earlier. It is distinct from the bias that issues in a sort of Cartesian epistemology, because the latter is claimed to result quite directly from processing principles built into the mindreading system itself, whereas the empiricist bias arises out of reluctance to add an additional belief acquisition mechanism to the seemingly adequate tacit generalizations already employed by that system. And it is distinct from the bias in favour of Cartesian dualism, because the latter results, rather, from the widely differing explanatory frameworks employed by our tacit core psychology and our tacit core physics.

It may be, however, that tacit dualism becomes relevant to nativist/empiricist debates at a second and subsequent stage. For it is common for nativists to pre-empt or reply to an argument from simplicity by pointing out that humans are biological systems, and minds are biological parts of those systems. Biological systems are characteristically multi-component and messily complicated, with their various parts emerging out of complex cascading sets of interactions between genes and environments. Moreover, most components of biological systems are innate (unlearned) — as are fingers and toes, for example — suggesting that the same might be true of components of the mind. There should thus be no presumption in favour of the simplicity of the mind or its mental processes, and hence no general presumption against innateness as one potential source of knowledge acquisition (Carruthers, 2006). Anyone who tacitly thinks that the mind itself isn't a biological system, of course (as opposed to interacting causally with such systems), will find this response by nativists unsatisfying, and will continue to find innate ideas counter-intuitive. Tacit dualism may thus make some contribution to the bias against innateness, doing so by shoring up empiricist appeals to theoretical simplicity, which are arguably inappropriate if minds themselves are biological in nature.

5. Conclusion

I have suggested three ways in which the human mindreading faculty may bias the thinking of cognitive scientists. The first might lead consciousness researchers to take philosophers' thought experiments more seriously than they otherwise would, resulting from the tacit ontological dualism that the mindreading system motivates. A second suggestion is that mindreading may bias cognitive scientists in the direction of first-person-first accounts of mindreading abilities themselves, resulting this time from tacit inference-rules built into those abilities as simplifying heuristics. And the third may issue in a bias in favour of empiricist — or 'blank slate' — accounts of cognitive development, resulting from the fact that all of the modes of belief acquisition recognized by the mindreading system are themselves empirical ones.

It is not easy to see how the ideas sketched here could be subjected to empirical test. How are we to establish that these three biases are real, and really have an impact on debates in cognitive science? And supposing that they are real, how are we to determine whether or not the proffered explanations of their impact are correct? One can imagine at least making progress on these issues by measuring and correlating individual differences. For example, one could develop some sort of qualia realism scale to determine how seriously consciousness researchers take the possibility that consciousness might not be physically reducible. One could then use tacit measures of people's susceptibility to dualist thinking — for example, using the animated 'body swap' scenarios employed by Chudek *et al.* (2018) — predicting an association with the qualia realism scores.

Likewise, if there were an independent test of the extent to which different individuals make chronic use of mindreading in their daily lives, then one might predict that people who score highly on this measure would be more susceptible to the lure of empiricism over nativism. Perhaps the mind-mindedness measures developed by Meins *et al.* (2012) could be adapted for this purpose. Alternatively, one might use measures of autistic spectrum traits of the kind developed by Baron-Cohen *et al.* (2001), predicting that they would be anti-correlated with empiricist leanings.

Another potential avenue to explore would be to see if selective interference with the mindreading system impacts people's empiricist intuitions (either weakening or eliminating them). If transcranial magnetic stimulation (TMS) becomes tightly focused enough for one

to selectively introduce noise into just those regions around the temporoparietal junction that are specialized for mindreading, for example, then one might predict that people so stimulated would be less strongly inclined to agree with a series of novel empiricist-leaning statements than are control subjects.

It is hard to be confident that any such test would be successful, however, even if the ideas sketched in this paper are correct. For it seems unlikely that mindreading is a monolithic faculty, and it is hard to know what specific types of mindreading and mindreading measures should be linked to the postulated biases. A good deal of careful experimental work would need to be done, and likely new measures and methods would need to be developed and independently validated. Nevertheless, I hope to have succeeded in showing that it is at least *plausible* that the central place of mindreading in human cognition may issue in a number of distinct biases that affect humans when they turn their minds to cognitive science. I hope to have shown also that this thesis is plausible enough that we should guard against the impact of those biases as best we can in our work as cognitive scientists.

Acknowledgments

I am grateful to Iris Berent, Evan Westra, and two anonymous referees for their comments on an earlier version of this article.

References

Ariew, A. (1996) Innateness and canalization, *Philosophy of Science*, **63**, pp. S19–S27.

Baillargeon, R., Stavans, M., Wu, D., Gertner, R., Setoh, P., Kittredge, A. & Bernard, A. (2012) Object individuation and physical reasoning in infancy: An integrative account, *Language Learning and Development*, **8**, pp. 4–46.

Baillargeon, R., Scott, R. & Bian, L. (2016) Psychological reasoning in infancy, *Annual Review of Psychology*, **67**, pp. 159–186.

Barlev, M. & Shtulman, A. (under review) Minds, bodies, spirits, and gods: Does widespread belief in disembodied being imply that we are inherent dualists?, [Online], https://psyarxiv.com/e9cw4/.

Baron-Cohen, S., Wheelwright, S., Skinner, R., Martin, J. & Clubley, E. (2001) The Autism-Spectrum Quotient (AQ): Evidence from Asperger syndrome/high-functioning autism, males and females, scientists and mathematicians, *Journal of Autism and Developmental Disorders*, **31**, pp. 5–17.

Berent, I. (2020) *The Blind Storyteller: How We Think About Human Nature*, Oxford: Oxford University Press.

Berent, I., Platt, M. & Sandoboe, G. (2019) People's intuitions about innateness, *Open Mind*, **3**, pp. 101–114.

Berent, I., Platt, M. & Sandoboe, G. (under review) How we reason about innateness: The role of dualism and essentialism, [Online], https://psyarxiv.com/vy6j5.
Bering, J. (2006) The cognitive psychology of belief in the supernatural, *American Scientist*, **94**, pp. 142–149.
Bering, J. & Bjorklund, D. (2004) The natural emergence of reasoning about the afterlife as a developmental regularity, *Developmental Psychology*, **40**, pp. 217–233.
Blanchard, T., Hayden, B. & Bromberg-Martin, E. (2015) Orbitofrontal cortex uses distinct codes for different choice attributes in decisions motivated by curiosity, *Neuron*, **85**, pp. 602–614.
Bloom, P. (2004) *Descartes' Baby*, New York: Basic Books.
Boyer, P. (2001) *Religion Explained: The Evolutionary Foundations of Religious Belief*, New York: Basic Books.
Brown, D. (1991) *Human Universals*, New York: McGraw-Hill.
Carruthers, P. (1992) *Human Knowledge and Human Nature*, Oxford: Oxford University Press.
Carruthers, P. (2006) *The Architecture of the Mind*, Oxford: Oxford University Press.
Carruthers, P. (2008) Cartesian epistemology: Is the theory of the self-transparent mind innate?, *Journal of Consciousness Studies*, **15** (4), pp. 28–53.
Carruthers, P. (2011) *The Opacity of Mind*, Oxford: Oxford University Press.
Carruthers, P. (2013) Mindreading in infancy, *Mind & Language*, **28**, pp. 141–172.
Carruthers, P. (2014) Two concepts of metacognition, *Journal of Comparative Psychology*, **128**, pp. 138–139.
Carruthers, P. (2018) Basic questions, *Mind & Language*, **33**, pp. 130–147.
Carruthers, P. (2019a) Questions in development, in Butler, L., Ronfard, S. & Corriveau, K. (eds.) *The Questioning Child*, Cambridge: Cambridge University Press.
Carruthers, P. (2019b) *Human and Animal Minds: The Consciousness Questions Laid to Rest*, Oxford: Oxford University Press.
Carruthers, P. & Smith, P.K. (eds.) (1996) *Theories of Theories of Mind*, Cambridge: Cambridge University Press.
Carruthers, P., Laurence, S. & Stich, S. (eds.) (2005) *The Innate Mind: Structure and Contents*, Oxford: Oxford University Press.
Carruthers, P., Laurence, S. & Stich, S. (eds.) (2006) *The Innate Mind: Vol. 2: Culture and Cognition*, Oxford: Oxford University Press.
Carruthers, P., Laurence, S. & Stich, S. (eds.) (2007) *The Innate Mind: Vol. 3: Foundations and the Future*, Oxford: Oxford University Press.
Carruthers, P. & Veillet, B. (2011) The case against cognitive phenomenology, in Bayne, T. & Montague, M. (eds.) *Cognitive Phenomenology*, pp. 35–56, Oxford: Oxford University Press.
Carruthers, P. & Veillet, B. (2017) Consciousness operationalized, a debate realigned, *Consciousness and Cognition*, **55**, pp. 79–90.
Caston, V. (2002) Aristotle on consciousness, *Mind*, **111**, pp. 751–815.
Chalmers, D.J. (1996) *The Conscious Mind*, New York: Oxford University Press.
Chudek, M., McNamara, R., Birch, S., Bloom, P. & Henrich, J. (2018) Do minds switch bodies? Dualist interpretations across ages and societies, *Religion, Brain & Behavior*, **8**, pp. 354–368.
Clement, J. (1982) Students' preconceptions in introductory mechanics, *American Journal of Physics*, **50**, pp. 66–70.

Cohen, E., Burdett, E., Knight, N. & Barrett, J. (2011) Cross-cultural similarities and differences in person-body reasoning: Experimental evidence from the United Kingdom and Brazilian Amazon, *Cognitive Science*, **35**, pp. 1282–1304.
Couchman, J., Coutinho, M., Beran, M. & Smith, J.D. (2009) Metacognition is prior, *Behavioral and Brain Sciences*, **32**, p. 142.
Couchman, J., Beran, M., Coutinho, M., Boomer, J. & Smith, J.D. (2012) Evidence for animal metaminds, in Beran, M., Brandl, J., Perner, J. & Proust, J. (eds.) *Foundations of Metacognition*, Oxford: Oxford University Press.
Craver, C. (2006) When mechanistic models explain, *Synthese*, **153**, pp. 355–376.
Descartes, R. (1641) *Meditations on First Philosophy*, many editions and translations available.
Dunbar, R. (1997) *Grooming, Gossip, and the Evolution of Language*, Cambridge, MA: Harvard University Press.
Dunbar, R. (2004) Gossip in evolutionary perspective, *Review of General Psychology*, **8**, pp. 100–110.
Foley, R. (1987) *The Theory of Epistemic Rationality*, Cambridge, MA: Harvard University Press.
Forstmann, M. & Burgmer, P. (2015) Adults are intuitive mind–body dualists, *Journal of Experimental Psychology: General*, **144**, pp. 222–235.
Frankish, K. (2004) *Mind and Supermind*, Cambridge: Cambridge University Press.
Friedman, J. (2013) Question-directed attitudes, *Philosophical Perspectives*, **27**, pp. 145–174.
Galef, B. (1996) Historical origins: The making of a science, in Houck, L. & Drickamer, L. (eds.) *Foundations of Animal Behavior*, Chicago, IL: Chicago University Press.
Gallese, V. & Goldman, A. (1998) Mirror neurons and the simulation theory of mind-reading, *Trends in Cognitive Sciences*, **2**, pp. 493–501.
Gelman, S. (2003) *The Essential Child*, Oxford: Oxford University Press.
Gigerenzer, G., Todd, P. and the ABC Research Group (1999) *Simple Heuristics that Make Us Smart*, Oxford: Oxford University Press.
Goldman, A. (1999) *Knowledge in a Social World*, Oxford: Oxford University Press.
Goldman, A. (2006) *Simulating Minds*, Oxford: Oxford University Press.
Goupil, L., Romand-Monnier, M. & Kouider, S. (2016) Infants ask for help when they know they don't know, *Proceedings of the National Academy of Sciences*, **113**, pp. 3492–3496.
Griffiths, P. (2001) What is innateness?, *The Monist*, **85**, pp. 70–85.
Gruber, M., Gelman, B. & Ranganath, C. (2014) States of curiosity modulate hippocampus-dependent learning via the dopaminergic circuit, *Neuron*, **84**, pp. 486–496.
Grünbaum, A. (1984) *The Foundations of Psychoanalysis: A Philosophical Critique*, Oakland, CA: University of California Press.
Heyes, C. (2018) *Cognitive Gadgets: The Cultural Evolution of Thinking*, Cambridge, MA: Harvard University Press.
Jackson, F. (1982) Epiphenomenal qualia, *Philosophical Quarterly*, **32**, pp. 127–136.
Kant, I. (1781) *Critique of Pure Reason*, many translations and editions now available.

Kidd, C. & Hayden, B. (2015) The psychology and neuroscience of curiosity, *Neuron*, **88**, pp. 449–460.
Knudsen, B. & Liszkowski, U. (2012) 18-month-olds predict specific action mistakes through attribution of false belief, not ignorance, and intervene accordingly, *Infancy*, **17**, pp. 672–691.
Kuhlmeier, V., Bloom, P. & Wynn, K. (2004) Do 5-month-old infants see humans as material objects?, *Cognition*, **94**, pp. 95–103.
Leiter, B. (2019) *Moral Psychology with Nietzsche*, Oxford: Oxford University Press.
Litman, J. (2005) Curiosity and the pleasures of learning: Wanting and liking new information, *Cognition and Emotion*, **19**, pp. 793–814.
Locke, J. (1690) *An Essay Concerning Human Understanding*, many editions available.
Loewenstein, G. (1994) The psychology of curiosity: A review and reinterpretation, *Psychological Bulletin*, **116**, pp. 75–98.
Mannoni, O. (1971) *Freud: The Theory of the Unconscious*, London: Verso.
McCloskey, M. (1983) Naïve theories of motion, in Gentner, D. & Stevens, A. (eds.) *Mental Models*, Mahwah, NJ: Lawrence Erlbaum.
McGinn, C. (1991) *The Problem of Consciousness*, Oxford: Blackwell.
McNamara, R., Willard, A., Norenzayan, A. & Henrich, J. (2019) Weighing outcome vs. intent across societies: How cultural models of mind shape moral reasoning, *Cognition*, **182**, pp. 95–108.
Meins, E., Fernyhough, C., de Rosnay, M., Arnott, B., Leekam, S. & Turner, M. (2012) Mind-mindedness as a multidimensional construct, *Infancy*, **17**, pp. 394–415.
Morgan, C.L. (1894) *An Introduction to Comparative Psychology*, London: Walter Scott.
Nichols, S. & Stich, S. (2003) *Mindreading*, Oxford: Oxford University Press.
Penn, D., Holyoak, K. & Povinelli, D. (2008) Darwin's mistake: Explaining the discontinuity between human and nonhuman minds, *Behavioral and Brain Sciences*, **31**, pp. 109–130.
Peters, M. & Lau, H. (2015) Human observers have optimal introspective access to perceptual processes even for visually masked stimuli, *eLife*, **4**, e09651.
Phillips, I. (2018) Unconscious perception reconsidered, *Analytic Philosophy*, **59**, pp. 471–514.
Pickering, M. & Garrod, S. (2013) An integrated theory of language production and comprehension, *Behavioral and Brain Sciences*, **36**, pp. 329–347.
Pinker, S. (2002) *The Blank Slate: The Modern Denial of Human Nature*, New York: Viking Press.
Rees, G., Krieman, G. & Koch, C. (2002) Neural correlates of consciousness in humans, *Nature Reviews Neuroscience*, **3**, pp. 261–270.
Riekki, T., Lindeman, M. & Lipsanen, J. (2013) Conceptions about the mind–body problem and their relations to afterlife beliefs, paranormal beliefs, religiosity, and ontological confusions, *Advances in Cognitive Psychology*, **9**, pp. 112–120.
Roazzi, M., Nyhof, M. & Johnson, C. (2013) Mind, soul and spirit: Conceptions of immaterial identity in different cultures, *International Journal for the Psychology of Religion*, **23**, pp. 75–86.
Robbins, P. & Jack, A. (2006) The phenomenal stance, *Philosophical Studies*, **127**, pp. 59–85.

Rosati, A. & Santos, L. (2016) Spontaneous metacognition in Rhesus monkeys, *Psychological Science*, **27**, pp. 1181–1191.
Samuels, R. (2007) Is innateness a confused notion?, in Carruthers, P., Laurence, S. & Stich, S. (eds.) *The Innate Mind: Vol. 3: Foundations and the Future*, Oxford: Oxford University Press.
Scott, R., Baillargeon, R., Song, H. & Leslie, A. (2010) Attributing false beliefs about non-obvious properties at 18 months, *Cognitive Psychology*, **61**, pp. 366–395.
Shtulman, A. & Valcarcel, J. (2012) Scientific knowledge suppresses but does not supplant earlier intuitions, *Cognition*, **124**, pp. 209–215.
Smith, J.D., Couchman, J. & Beran, M. (2014) Animal metacognition: A tale of two comparative psychologies, *Journal of Comparative Psychology*, **128**, pp. 115–131.
Song, H. & Baillargeon, R. (2008) Infants' reasoning about others' false perceptions, *Developmental Psychology*, **44**, pp. 1789–1795.
Southgate, V., Chevallier, C. & Csibra, G. (2010) Seventeen-month-olds appeal to false beliefs to interpret others' referential communication, *Developmental Science*, **13**, pp. 907–912.
Spelke, E. & Kinzler, K. (2007) Core knowledge, *Developmental Science*, **10**, pp. 89–96.
Tomasello, M. (2008) *Origins of Human Communication*, Cambridge, MA: MIT Press.
Tononi, G. & Koch, C. (2008) The neural correlates of consciousness: An update, *Annals of the New York Academy of Sciences*, **1124**, pp. 239–261.
Weiskrantz, L. (1986) *Blindsight*, Oxford: Oxford University Press.
Whitcomb, D. (2010) Curiosity was framed, *Philosophy and Phenomenological Research*, **81**, pp. 664–687.
Willard, A. & Norenzayan, A. (2013) Cognitive biases explain religious belief, paranormal belief, and belief in life's purpose, *Cognition*, **129**, pp. 379–391.
Williamson, T. (2000) *Knowledge and its Limits*, Oxford: Oxford University Press.
Winkielman, P. & Berridge, K. (2004) Unconscious emotion, *Current Directions in Psychological Science*, **13**, pp. 120–123.
Zawidzki, T. (2013) *Mindshaping*, Cambridge, MA: MIT press.

Paper received April 2019; revised July 2019.