ORIGINAL ARTICLE

Morgan's Quaker gun and the species of belief

Devin Sanchez Curry

West Virginia University

Correspondence
Devin Sanchez Curry, West Virginia
University.
Email: devin.curry@mail.wvu.edu

Abstract

In this article, I explore how researchers' metaphysical commitments can be conducive—or unconducive—to progress in animal cognition research. The methodological dictum known as Morgan's Canon exhorts comparative psychologists to countenance the least mentalistic fair interpretation of animal actions. This exhortation has frequently been misread as a blanket condemnation of mentalistic interpretations of animal behaviors that could be interpreted behavioristically. But Morgan meant to demand only that researchers refrain from accepting default interpretations of (apparent) actions until other fair interpretations have been duly considered. The Canon backfired largely because of Morgan's background metaphysical commitment to a univocal, hierarchical, and anthropocentric account of cognitive architecture. I make the case that, going forward, comparative psychologists would do well to pair judicious use of Morgan's Canon with an openness to the existence of non-humanlike animal minds comprising phenomena belonging to distinct cognitive and folk psychological ontologies. And I argue that this case gives us pragmatic reason to reconcile deep—e.g., psychofunctionalist—and superficial—e.g., dispositionalist—approaches to the metaphysics of belief.

1 | PRELUDE: IS BELIEF DEEP OR SUPERFICIAL—AND WHY SHOULD WE CARE, ANYWAY?

Why should scientists studying cognition care about the metaphysics of mind? Do researchers' metaphysical commitments make any worthwhile difference to methodology or theory construction in cognitive science?

This question has been weighing on empirically oriented philosophers of mind lately due to the renewal of a longstanding battle over the metaphysics of belief. One camp, helmed in recent years by Eric Schwitzgebel, takes a superficial approach to the metaphysics, identifying beliefs with patterns of behavioral, cognitive, and phenomenal dispositions.[1] A second camp, helmed in recent years by Eric Mandelbaum, delves deeper, identifying beliefs with the psychofunctional states underlying the relevant dispositions.[2]

Pace a third, fictionalist camp (Demeter, Parent, and Toon 2022), let's agree about this much at the outset: human beings have beliefs, which can be initially (uncontroversially and rather uninformatively) glossed as stances of taking the world to be some way. (To believe three cats are meowing in the kitchen is to take that room of the house to feature three meowing cats.) I'll merrily assume, without argument, that this isn't a fiction; it is simply true that humans really believe some things and not others—that humans take the world to be some ways and not others. It is simply true, of me, as I write this sentence, that I believe three cats to be meowing in my kitchen.

Here's where the first disagreement among realists about belief comes in: what's involved, exactly, in taking my kitchen to feature three meowing cats? What does that belief consist in? Is believing three cats are meowing in my kitchen a "temporary or habitual posture of the mind" (Schwitzgebel 2013: 76): a comparatively superficial matter of my having tendencies to act, think, and feel in ways consistent with there being three cats meowing over there? Or is it a comparatively deep matter of my mind or brain having a moving part—a "concrete particular with causal powers" (Quilty-Dunn & Mandelbaum 2018)—that produces those tendencies?

To cut a long battle report short, the conflict seems once again to have reached a stalemate. Alongside their respective allies, Schwitzgebel and Mandelbaum have each developed conceptually and empirically well-supported accounts of belief that appear fundamentally incompatible with each other. Schwitzgebel writes that his dispositionalism "directs our attention to what we ought to care about most in thinking about belief: our overall ways of acting in and reacting to the world" (2021: §1). Jake Quilty-Dunn and Mandelbaum write that "a desideratum for a theory of belief is to explain how beliefs [productively] cause behavior" (2018: §3.1), and argue that their psychofunctional representationalism (featuring fragmented belief storage) fulfills this desideratum. For reasons that will emerge over the course of this article, I think Schwitzgebel is right that his dispositionalism fulfills the desideratum he singles out better than psychofunctionalism. At the same time, however, Quilty-Dunn and Mandelbaum are obviously right that their psychofunctionalism fulfills the desideratum they single out better than dispositionalism, which doesn't even try to cast beliefs as causally productive of behavior. Neither camp is budging, but it seems as though one desideratum or the other has to give: belief cannot be simultaneously superficial and deep.

Or it couldn't be if the word "belief" picked out just one kind of phenomenon. Given the independent attractions of Schwitzgebel and Mandelbaum's accounts, we should consider the possibility that their shared use of the term "belief" obscures the fact that they are actually targeting two distinct kinds of phenomenon of interest—a superficial phenomenon and a deep

phenomenon, respectively. In that case, it is possible that Schwitzgebel is giving a perfectly adequate account of the superficial phenomena and Mandelbaum is giving a perfectly adequate account of the deep phenomena, with no necessary tension between them. Maybe everybody can go home in peace.

In previous articles (Curry 2020, 2021a), I have suggested a distinction between two phenomena picked out by the word "belief" that could facilitate this peace. Those articles present a series of arguments for the conclusion that theorists ought not conflate "'attitudes of belief'—the beliefs that lay people attribute to each other (and other animals) in everyday life—and 'cognitive states of belief'—the beliefs that (some) cognitive scientists posit as cogs in cognitive machines" (Curry 2021a: 7889).[3] Explaining how beliefs produce behavior is a core desideratum for a theory of cognitive states of belief, but it needn't be a desideratum for a theory of attitudes of belief: ordinary belief attributors are concerned with believers' holistic patterns of activity, not with the nitty-gritty details of the inner causes of their behavior (Curry 2018). Meanwhile, capturing our overall ways of acting in the world is a core desideratum for a theory of attitudes of belief, but it needn't be a desideratum for a theory of cognitive states of belief: when doing cognitive science, we shouldn't pretheoretically presume any particular relationship between any given (kind of) cognitive state and holistic patterns of organismic activity.

On my view, Schwitzgebel's dispositionalism is a good theory of belief qua superficial folk psychological kind no matter what we make of psychofunctionalism as a theory of belief qua deep cognitive kind. Maybe Mandelbaum is right about cognitive architecture; maybe not. If he's right, then things worthy of the label "belief" exist qua cognitive states. If he's wrong, then maybe cognitive states of belief don't exist. Either way, people have superficial patterns of dispositions that fit folk psychological models of what it takes to believe. The details of the underlying cognitive architecture don't make a difference when it comes to evaluating Schwitzgebel's theory of (superficial) believing. Indeed, I have argued (Curry 2021a) that even if psychofunctionalism is correct in all of its details, we should take superficial attitudes of belief and deep cognitive states of belief to be metaphysically distinct, since they're individuated by playing distinct functional roles: the latter function as cogs in cognitive systems whereas the former function as fodder for folk psychology. If that's right, then the term "belief" is indeed equivocal between a superficial phenomenon and a deep phenomenon.[4]

So, to return to our opening question: what hangs on how we resolve this metaphysical debate? Who cares whether beliefs are really deep, or really superficial, or whether there really are both deep ones and superficial ones? In the end, it's in answering this practical question that Schwitzgebel and Mandelbaum disagree most acutely. According to Schwitzgebel (2021, 2022), pragmatic considerations should push us to reserve the term "belief" for the superficial phenomena—attitudes of belief—because they're the phenomena of preeminent concern to the lay people, philosophers, and social scientists who talk about belief. According to Mandelbaum (2014), a different set of pragmatic considerations should push us to identify beliefs with deep phenomena—cognitive states of belief—because doing so will lead to progress in the causal-explanatory project of cognitive science.

In this article, I'm going to suggest that a third set of pragmatic considerations—which overlaps in significant respects with both Schwitzgebel and Mandelbaum's considerations—should push us to reconcile the two camps by refusing to conflate attitudes of belief with cognitive states of belief. In particular, I'll argue that this reconciliation would be useful insofar as it could help us think through some persistent tangles in the study of animal minds.

2 | INTRODUCTION: METAPHYSICS AND METHODOLOGY IN COMPARATIVE PSYCHOLOGY

Humans aren't the only believers in the world. As of the time of this writing—7:28pm—my increasingly desperately meowing cats believe it is dinner time. (Unfortunately for everyone, their beliefs are false: cat dinner isn't served until nine o'clock.) Philosophers have often puzzled over the nature of animal belief, and some have denied the phenomenon outright. Most notoriously, influential 20th century empiricists from John Dewey (1925: 169–170, 321–324) to Stephen Stich (1979) and Donald Davidson (1982) argued that all believers are language users, and that to attribute beliefs to cats is crass anthropomorphism. The latter invective was also given voice by the majority of ethologists and comparative psychologists working in the 20th century, who opposed the attribution of any humanlike psychological traits to nonhuman animals.

Over the last half-century, researchers have become more willing to stomach empirically well-attested attributions of mental capacities to animals. Meanwhile, most scientific and philosophical attention has turned from the question of whether animals believe to the question of whether animals attribute beliefs to one another. Researchers have produced a large and influential body of evidence of apparent belief attribution by nonhuman primates (Lewis & Krupenye 2022), but this evidence is still regularly met with both conceptual challenges (Penn & Povinelli 2007) and plausible alternative interpretations that deny belief attribution (Heyes 2017).

Philosophers have periodically pointed out that the metaphysics of mind should be able to help researchers clear up conceptual tangles and choose between competing interpretations. Whether animals have beliefs depends on what beliefs are (Lesson, Tinklenberg & Andrews 2020), and whether animals attribute beliefs depends on what attributing beliefs consists in (Buckner 2014). Nevertheless, psychologists sometimes doubt that metaphysical accounts of belief and belief attribution could help settle the relevant empirical debates (Apperly 2011: 5). After all, the point of animal cognition research is to replace metaphysical speculation about what goes on within the erstwhile black boxes of animal minds with careful experimental and observational study. Even if metaphysical accounts informed researchers about the nature of belief and belief attribution, it isn't obvious that (much less how) they would be informative as to whether researchers were justified in (anthropomorphically?) attributing the relevant psychological capacities to nonhuman animals.

My overarching aim in this article is to clarify some ways in which researchers' metaphysical commitments can be conducive—or unconducive—to progress in animal cognition research. On my view, researchers should care about the metaphysics of mind because their metaphysical commitments are, as a matter of course, going to infect their science one way or the other. Given its inevitability, it would be best to ensure that the relationship is symbiotic. I'll end up arguing that refusing to conflate attitudes of belief with cognitive states of belief can help researchers figure out how to resolve empirical and conceptual debates about animal belief and belief attribution capacities. Before I get to that positive argument, though, I'm going to look back at the early history of comparative psychology, in order to show how metaphysics has informed animal cognition research from the very start, and in order to free a useful methodological tool from its original, inhibitive metaphysical trappings.

Early in the 20th century, opposition to attributing psychological capacities to nonhuman animals found its clarion call in Morgan's Canon, which has traditionally been read as forbidding mentalistic interpretations of animal behaviors that could be interpreted behavioristically. Most recent theorists doubt that Morgan's Canon best captures how animal cognition researchers ought

to avoid (the purportedly bad variety of) anthropomorphism. Although some behavioristic interpretations are baroque to the point of being obviously ad hoc (Chomsky 1959; Lurz 2011), every observable behavior *could* be interpreted behavioristically (Buckner 2014). Always adopting available behavioristic interpretations—however baroque—would entail never attributing beliefs or belief attribution capacities to animals, and more generally prevent animal cognition researchers from getting on with the work of researching animal cognition (irrespective of whether animal cognition resembles human cognition).

By my lights, this criticism is warranted when leveled against the received reading of Morgan's Canon, and even when leveled against the way C. Lloyd Morgan himself used his Canon, as informed by his background metaphysical commitments. I'll nevertheless argue in what follows that, once shed of those metaphysical commitments, Morgan's Canon ought to be used as originally intended. Indeed, the eponymous Morgan's original intention would be better realized if buttressed by two commitments of which he didn't avail himself.

Morgan's account of cognitive architecture left no room for non-anthropomorphic mentalistic phenomena. Moreover, Morgan's introspectionist methodology—and attendant univocal mental ontology—necessitated a conflation of the objects of cognitive psychology (such as cognitive states of belief) with the objects of folk psychology (such as attitudes of belief). Fortunately, neither Morgan's strictly hierarchical cognitive architecture nor his introspectionism are live options in 21st century psychology. Moreover, philosophers of psychology have recently offered persuasive arguments against both "anthropocentrism" (Andrews 2020: 3)—the thesis that there is a strict positive correlation between the mentalistic and the humanlike—and the conflation of the ontologies of cognitive science and folk psychology (Curry 2021a; Dewhurst 2017, 2021; Poslajko 2022).

On my originalist reading, Morgan's dictum isn't a cannon but a Quaker gun: a harmless log posing as a deadly weapon of war. A Quaker gun doesn't blow anybody to smithereens. It does cause advancing foes to think twice. Likewise, used properly, Morgan's Canon doesn't demolish mentalistic hypotheses. Instead, it causes researchers who would advance mentalistic hypotheses to think twice. Historically, the problem with the Canon has been that, paired with Morgan's metaphysical commitments (or their descendants), this impetus to think twice would inevitably lead comparative psychologists in Morgan's mold away from ascribing any sophisticated psychological capacities to animals. But I'll argue that this result stems from wrongheaded commitments to anthropocentrism and a univocal mental ontology, rather than from Morgan's Canon per se. The Canon itself would be nothing but a boon to animal cognition research—an essential part of any methodology that aims to attribute to animals the psychological characteristics that they actually have—if combined with a metaphysical outlook featuring a non-Morganian openness to the existence of non-humanlike animal minds comprising phenomena belong to distinct cognitive and folk psychological ontologies.

If that's right, then reflection on animal cognitive research will provide us with a new pragmatic argument for refusing to conflate attitudes of belief with cognitive states of belief—and thus for resolving the battle over the metaphysics of belief by declaring everybody a winner. Here's the plan: I'll give my originalist interpretation of Morgan's Canon in §3, combine it with a non-anthropocentric, two-tiered metaphysics of belief in §4, and then put the pair to work disentangling the debates about animal belief in §5 and animal belief attribution in §6. In §7, I'll conclude that the two realist camps in the battle introduced in §1 should give up their rivalry and cheerily join forces.

3 | MORGAN'S QUAKER GUN

At the International Congress of Experimental Philosophy in 1892, the ethologist and philosopher C. Lloyd Morgan declared that "in no case may we interpret an action as the outcome of the exercise of a higher psychical faculty, if it can be fairly interpreted as the outcome of the exercise of one which stands lower in the psychological scale" (reported in Dixon 1892: 392; repeated in Morgan 1903: 242). At a first gloss, this dictum, which has come to be known as 'Morgan's Canon', requires researchers speculating about the causes of an animal behavior to consider the least mentalistic fair interpretation.

Putting his Canon into action, Morgan criticized George Romanes for inferring that dogs "have general ideas of 'good-for-eating' and 'not-good-for-eating,' *quite apart from any particular objects of which either of these qualities happens to be characteristic*" (Romanes 1888: 36). Morgan argued that all relevant doggish behavior can be explained by attributing to dogs the capacity to recognize particular objects as good-for-eating. This recognition is itself the exercise of a psychical faculty, of course; as Morgan wrote, "this is a concept in [some uses] of the term, I admit" (Morgan 1891: 349). However, it is the exercise of a psychical faculty that stands lower on the psychical scale— that is, it is a less sophisticated cognitive capacity[5]—than the capacity to consider the concept of 'good-for-eating' in the abstract, apart from any particular edible objects. Since Morgan knew of no independent evidence bearing on the question "whether this quality can be isolated by the dog, and can exist in his mind divorce from the eatables which suggest it" (348), he was left "unable to attribute to the brute" the higher and human-like "power of analysis—the power of isolating qualities of objects" (350). It's possible that dogs have this higher power, but the less anthropomorphic interpretation was fair in light of the available evidence, so Morgan's Canon dictated that it won the day.

By the 1930s, Morgan's Canon was regularly understood as recommending a strict ban on the attribution of psychological capacities to animals; for example, B.F. Skinner wrote that whereas Charles Darwin attributed mental states to animals, "Lloyd Morgan, with his law of parsimony, dispensed with them in a reasonably successful attempt to account for characteristic animal behavior without them" (1938: 4). By the end of the 20th century, Morgan's Canon had become "possibly the most important single sentence in the study of animal behavior" (Galef 1996: 9). The importance of the Canon derived mainly from its being used to justify the extended reign of Skinnerian behaviorism in animal psychology for many years after the cognitive revolution ousted the behaviorist ruling class in human psychology. If behaviorist interpretations were always fair— a seemingly low bar—then Morgan's Canon was interpreted as dictating that they must always prevail.

Modified versions of the Canon have also found support from epistemically cautious researchers who are nonetheless explicitly interested in studying the unobservable mental states and cognitive capacities of animals (Cheney & Seyfarth 1992; Shettleworth 2010). In the 21st century, 'anthropomorphism' is still "almost a dirty word in the scientific study of animal cognition" (Shettleworth 2007: 4),[6] even though more researchers have become comfortable attributing psychological capacities to the nonhuman animals they study.[7] Meanwhile, Morgan's Canon has received increased scrutiny from both philosophers and comparative psychologists, some of whom have begun self-identifying as cognitive scientists—"cognitive ethologists" (Allen & Bekoff 1997)—as opposed to behaviorists. Many commenters have pointed out that Morgan's talk of "higher" and "lower" psychological faculties relies on a misunderstanding of Darwin's theory of evolution by natural selection. This common objection is well-founded (Arnet 2019),

and I'll discuss how Morgan's naïvely hierarchical account of cognitive architecture got him into methodological trouble in §4.

However, the most incisive objections set their sights beyond the wording (and outdated architectural presumptions) of Morgan's Canon, and instead target the core principle behind its standard interpretation: the rejection of anthropomorphism. For example, Shettleworth writes that "evolutionary continuity justifies anthropomorphism as a source of hypotheses" (2007: 4), citing Elliot Sober's warning that "if nonhuman animals really are like us in certain respects, the canon may lead us to miss this fact about nature" (1998: 229). Sober worries that being overly cautious about anthropomorphism will lead researchers to an unwarranted form of its opposite— what Kristin Andrews and Brian Huss (2013) have termed "anthropectomy"—the denial that nonhuman animals have human-like psychological characteristics. Jerry Fodor echoes Sober's worry that Morgan's Canon avoid anthropomorphic bias only by falling prey to anthropectic bias: "Why doesn't Fodor's Pop Gun tip the scales equally in the opposite direction To wit: in no case may we interpret an action as the outcome of the exercise of a lower psychical faculty, if it can be interpreted as the outcome of the exercise of one which stands higher in the psychological scale" (1999: 12).[8] More recently, Mike Dacey (2017: 1159) argues that "even in cases in which intuitive anthropomorphism has led someone to overestimate the intelligence of an animal, Morgan's Canon may not be effective" because demanding that researchers avoid anthropomorphism doesn't actually help—and may even hinder—their attempts to consider anthropectic interpretations.

In assessing the role Morgan's Canon should play in animal cognition research, all of these writers focus on the ways in which the Canon is currently invoked. The most incisive objections to Morgan's Canon get traction against its standard interpretation, but not against its original meaning.[9] If Morgan's Canon is "perhaps the most quoted statement in the history of comparative psychology" (Dewsbury 1984: 1987), it is also "perhaps the most misrepresented statement in the history of comparative psychology" (Thomas 1998: 156). As Roger Thomas and other historians and philosophers of science have argued (Costall 1993; Thomas 1998; Fitzpatrick and Goodrich 2017; Bohnert & Hilbert 2018; Andrews 2020), Morgan himself never intended his Canon as a stark prohibition against anthropomorphism. Instead, on the reading I'll now advance, he intended it as a (then already decreasingly controversial, now almost totally uncontroversial) empiricist check against bias, given the inevitability of some anthropomorphism in any comparative psychology worth its salt.

Morgan's early philosophical writings reveal a clear development in their author's thinking about the possibility of animal cognition research. As a young avid reader of George Berkeley (Morgan 1930a) struck by the power of the epistemological problem of other minds (Morgan 1880), Morgan wrote that

> The results of comparative psychology—the science which has for its object the comparative study of those distorted images of our own mental processes—are incapable of verification ... "Is there a science of comparative psychology?" [I submit] an emphatic negative. (Morgan 1884: 371).

However, Morgan changed his mind within a decade. By the time Morgan devised his Canon, he proudly self-identified as a comparative psychologist, though his about-face didn't keep him from being preoccupied with the epistemological issues pervading his field. Morgan continued to worry about the way most Darwinists were studying animal minds in the 19th century. Nevertheless, this mature worry didn't concern anthropomorphism; it concerned the then-standard practice of

inferring mental capacities from anecdotes told by gentleman naturalists. Morgan insisted that "observation on one occasion only, no matter how careful and exact that interpretation may be, does not suffice for the interpretation of this or that instance of animal behavior" (Morgan 1930b; excerpted by Cook 1999). The problem identified in this passage isn't the attribution of psychological capacities to animals, but the attribution of *any* capacities on the basis of too little evidence. Morgan was obsessed, first and foremost, with empirical rigor.[10]

Empirical rigor is difficult to achieve in comparative psychology, and it was particularly difficult given that Morgan conceived the study of animal minds to necessarily involve the study of the distorted images of human mental processes. According to Morgan's introspectionism, "the first duty of a psychologist is to attain accurate and systematic acquaintance with the working of his own mind, as the cipher in terms of which all other minds must be read" (1894: Ch. 3). An introspectionist comparative psychologist becomes equipped to start studying animal minds only once she has properly understood her own mind via introspection. She then attributes psychological capacities to animals by judging whether "ejections" of her subjective mental processes would make sense of the animals' behaviors (Morgan 1891). For example, Morgan's psychologist might reason as follows when confronted with my cats around dinner time: *the cats are meowing in an agitated tone and pacing around their food bowls. Introspection tells me that when I engage in analogous behaviors—complaining and rooting around in the fridge or pantry—it is because I am hungry. Thus, the cats must be meowing and pacing out of hunger.*

Morgan explicitly developed his Canon as subservient to the first duty of the introspective method. "In no case may we interpret an action as the outcome of the exercise of a higher psychical faculty, if it can be fairly interpreted as the outcome of the exercise of one which stands lower in the psychological scale" (Morgan 1903: 242). In context this dictum urges nothing more radical than that researchers make sure to rule out alternative explanations. The introspectionist methodology entails that the first explanation Morgan's psychologist devises is always the anthropomorphic hypothesis. In effect, Morgan's Canon commands the following: don't immediately accept the first explanation you devise via introspection; before settling for the first explanation you devise, make sure that there aren't other equally reasonable explanations available. Maybe the cats see a bug flying around their food bowls. Even if that is a possibility, however, Morgan's Canon doesn't dictate that the psychologist should accept the least mentalistic fair interpretation. Morgan himself clarified, "—lest the range of the principle be misunderstood—that the canon by no means excludes the interpretation of a particular act as the outcome of the higher mental processes, if we already have independent evidence of their occurrence in the agent" (1920: 270–271). The Canon dictates only that the psychologist refrain from assenting to the most obvious interpretation before ensuring that the less obvious available interpretations are also less fair.

So long as the psychologist is following the first duty of introspectionism (and presuming Morgan's hierarchical account of cognitive architecture, with paradigmatic human cognition perched on top of the hierarchy), the less obvious interpretations are always *therefore* less mentalistic. Thus, as Morgan himself stressed, the duty to always consider anthropectic hypotheses wasn't originally intended as a brief against anthropomorphism; it was just good empiricism in the face of the anthropomorphic default hypotheses delivered by Morgan's other methodological and theoretical—at their core, epistemological and metaphysical—commitments.

Few 21st century scientists consider introspection a reliable—much less necessary—first step in the methodology of animal cognition research. Nevertheless, cognitivist comparative psychologists do regularly begin their inquiry by considering the attribution of human-like psychological capacities to animals (though not necessarily capacities shared by the psychologist herself). They are right to do so: animal cognition research must assume it has an object of inquiry. As Andrews

argues, "it is necessary to permit scientists to propose and define functional categories, including categories in terms of psychological properties that some believe or suspect may be uniquely human" (2020: 6). Morgan was right that the inevitable presence of anthropomorphic default hypotheses will always confront the study of animal minds with looming confounds due to anthropomorphic bias. Even in the 21st century, a suitably contextualized version of Morgan's Canon serves as a useful corrective to this bias.

Contra Sober and Fodor, there is an asymmetry between anthropomorphic bias and anthropectic bias, since humans are systematically predisposed to the former (Dacey 2017). Anecdotally, the evidence that humans are inclined towards indulging in anthropomorphism is overwhelming. Everyone with a dog will tell you, at least in their less reflective moments, that Rover wants a biscuit, enjoys playing fetch, despises taking baths, or tries to stymy others by burying his bone. Consider also the urge to attribute mental states to the gliders that emerge in Conway's Game of Life (Dennett 1998), or to the triangles in Heider and Simmell's famous animation (Heider & Simmell 1944). When faced with living animals rather than mere pixels, this urge only intensifies. A large and diverse body of experimental evidence abets these anecdotes.[11] Perhaps most relevantly, Deborah Kelemen and her lab have demonstrated that "although extended education appears to produce an overall reduction in inaccurate teleological explanation, specialization as a scientist does not, in itself, additionally ameliorate scientifically inaccurate purpose-based theories about the natural world" (2013: 1074). By default, humans, including human scientists, are predisposed to judge animal behavior to be purposeful until presented with evidence to the contrary. Animal cognition researchers ought to take whatever methodological precautions are necessary to counteract this bias toward the anthropomorphic hypotheses that frame their investigations.

Researchers ought not always accept available anthropectic interpretations. But an openness to appropriate anthropomorphism doesn't entail an indictment of Morgan's Canon. Used correctly, Morgan's Canon doesn't demand that if an action can be fairly interpreted as the outcome of a non-anthropomorphic process, then researchers must accept the anthropectic interpretation. It demands only that researchers refrain from accepting the default (de facto, usually anthropomorphic) interpretation before thoroughly considering other interpretations.

Consider cases of sphexishness, where particular animal behaviors appear flexible and intelligent but are actually rote and thoughtless (Dennett 1984; cf. Keijzer 2013). Several species of birds—most notably piping plovers—appear to purposively lure predators away from their nests by feigning broken wings. The predator, seeing that the mother bird appears wounded, chases her instead of finding her nest and eating her eggs or young. When the mother bird has lured the predator far enough away from her nest, she takes flight, showing the broken wing to be a hoax. This phenomenon is ripe for anthropomorphism. To the casual observer, it may seem obvious that the mother bird is cleverly deceiving the predator in order to protect her hatchlings. As Dennett remarks, this explanation is as mentalistic as it is obvious, invoking "not just a goal, but also a *belief* about an *expectation*, and a *hypothesis* about the *rationality* of the predator and a *plan* based on that hypothesis" (2017: 91). However, close observation reveals that piping plovers automatically engage in broken wing display behaviors whenever predators get close to their nests, whether or not eggs or young are present. Researchers have concluded that broken wing display behaviors aren't the product of sophisticated, flexible, and intentional mental processes after all (Ristau 1983).[12]

Properly interpreted, Morgan's Canon does nothing more than direct researchers to engage in precisely this kind of careful study before jumping to anthropomorphic conclusions. Morgan's Canon isn't a cannon that blasts anthropomorphizers to shreds; it is a Quaker gun that scares them into considering other approaches.

4 | THE HETEROGENEITY OF THE MENTAL

Or it would be, in the right hands. With Morgan himself playing powder monkey, the Canon operated alongside two commitments that increased its prescriptive firepower until it backfired. The last section dealt with Morgan's epistemological commitment to introspectionism—and attendant anthropocentrism—in some detail; let's now turn to his metaphysical commitment to a strictly hierarchical cognitive architecture. As Hayley Clatterbuck (2016) explains, Morgan posited three levels of behavior stemming from three levels of cognitive ability. On the bottom of the hierarchy, there is (either obviously or sphexishly) automatic, mechanical behavior. In the middle of the hierarchy, there is intelligent behavior which can be shaped by experience. Finally, on top of the hierarchy there is behavior caused by abstract, logical reasoning. Paired with Morgan's Canon and his epistemological commitment to the first duty of introspectionist psychology, this metaphysical commitment directed Morgan to refrain from accepting the default hypothesis that a behavior has been caused by logical reasoning until he had thoroughly considered the possibility that it has merely been shaped by experience, and to refrain from accepting the default hypothesis that a behavior is intelligent until he had thoroughly considered the possibility that it is sphexish.[13]

The particular details of Morgan's account of cognitive architecture were driven by empirical research into animal behavior (which involved curating anecdotes from other naturalists, in addition to his own observational and experimental studies). However, it is difficult to read Morgan's reports of that empirical research—and his interpretation of it as yielding a three-tiered cognitive hierarchy—without getting the impression that the research itself was driven by a prior metaphysical commitment to cognitive hierarchy: to a Berkeleyian conception of the Great Chain of Being (Bradatan 2006) and to a Darwinian version of Aristotle's posit of a three-tiered cognitive hierarchy of vegetative, sensitive, and rational souls. The latter match up neatly with Morgan's mechanical, intelligent, and rational levels of mentation; I don't know of textual confirmation that Morgan was inspired by Aristotle (or the Aristotelian tradition) in this respect, but I would speculate that the convergence with Aristotle isn't merely coincidental. In any case, where Darwin was driven by his background metaphysical commitments to seek continuity between animal and human minds, Morgan seems to have been driven to establish discontinuity (Radick 2007), at times venturing well beyond his own methodological prescriptions in order to do so (Clatterbuck 2016: 12).

If my speculation is right, Morgan's Canon was misinterpreted for over a century—with considerable practical ramifications for the development of animal cognition research—in large measure because Morgan expressed the Canon in terms that were shaped by a hangover from Aristotelian metaphysics. Even putting this speculation aside, Morgan plainly expressed the Canon in terms that were shaped by his account of cognitive architecture, which informed how he interpreted empirical findings (much more than it was informed by them). It is inevitable that (empirically undermotivated) metaphysical commitments shape research in this manner to some degree—speculating ahead of the evidence is part and parcel of the practice of theory construction in science. So it behooves us to ask: which metaphysical commitments ought to shape animal cognition research? Of course, all and only true commitments would be preferable. But in the absence of a metaphysical consensus, pragmatic considerations have to carry some weight. In particular, we might ask which metaphysical outlook is most conducive to devising progressive research programs (Lakatos 1978).

Let's turn, then, to the distinction between attitudes of belief and cognitive states of belief introduced in §1. I've defined 'attitudes of belief' as the beliefs that lay people (veridically) attribute to

each other (and other animals) in everyday life. Attitudes of belief can thus be said to occupy a folk psychological ontology. In contrast, 'cognitive states of belief' are the beliefs that (some) scientists posit as cogs in cognitive machines: as occupants of a cognitive ontology.[14] On my view, an attitude of belief is a superficial pattern of dispositions to act, react, think and feel as if the world is a particular way.[15] To be precise, to believe[attitude] is to have a pattern of dispositions that fits a belief attributor's folk psychological model of what's involved in taking the world to be some way.[16] And I've argued in previous work that this model-theoretic interpretivism is the correct account of attitudes of belief *no matter what is true about the metaphysics of cognitive states of belief*: no matter whether they exist, and, if they exist, no matter whether they are constituted by neural processes, subpersonal psychofunctional roles, or personal level functional states.[17] In the remaining sections of this article, I won't offer new arguments for the claim that there is a metaphysically real distinction between attitudes of belief and cognitive states of belief. Instead, I'll show how letting the distinction shape the scientific investigation of animal minds would be conducive to devising more progressive research paradigms.

In §5, I'll argue that, when used in accordance with my originalist reading but stripped of Morgan's own metaphysical commitments, Morgan's Canon rightly leads to agnosticism about whether animals have cognitive states of belief (pending further scientific investigation), but not about whether (some) animals have (some) attitudes of belief. In §6, I'll argue that the distinction between attitudes and cognitive states also requires a significant reframing of the debate about animal belief attribution—the first question to ask is whether animals attribute attitudes of belief, like lay folks, not whether they attribute cognitive states of belief, like cognitive scientists. But the rejection of anthropocentrism also puts interesting pressure on this first question: animals might attribute attitudes of belief quite differently than humans, or they might attribute other attitudes entirely. The distinction between attitudes and cognitive states reveals how researchers can fruitfully investigate relevant hypotheses without getting mired in the ongoing—and possibly irresolvable—debate about whether nonhuman animals attribute "hidden" cognitive states to one another.

5 | BRUTE BELIEVERS

When people attribute beliefs to nonhuman animals, they're often accused of illicit anthropomorphism. These accusations tend to get their fangs because accusers conflate attitudes of belief with cognitive states of belief.

Morgan's Canon is paradigmatically aimed at scientific hypotheses about the unobservable psychological processes that produce behavior. When attempting to understand why some creature behaved in some way, candidate cognitive states often spring to investigators' minds. Why did the cats meow and pace? Why did PAC-MAN pursue the apple? Because they felt hungry. Morgan's Canon helpfully says: *slow down. Best make sure there isn't another cause that might reasonably be posited as producing those behaviors.*

Meanwhile, when I, a layperson sitting here in my pajamas, attribute the belief that it is dinner time to my cats, a skeptic might unhelpfully say: *slow down. Best make sure there isn't another cause that might reasonably be posited as producing those behaviors.* The disanalogy between my (unproblematic) belief attribution and the scientist's (problematically hasty) belief attribution stems from the fact that I am not (unless I'm acting very unusually) positing my cats' beliefs as inner causes mechanistically producing their dinnertime behaviors. Instead, I am attributing an attitude of belief—a pattern of dispositions that fits my folk psychological model of what's

involved in (a cat) taking it to be dinner time—to my cats, in order to articulate the reason that contextualizes their behavior, and thereby to better understand them as co-inhabitants of my social environment (Curry 2018).[18]

The aforementioned skeptic about my attribution of beliefs to cats rightly assumes (a) that she knows that humans have attitudes of belief, and (b) that she has a decent epistemic handle on those attitudes. She also rightly assumes (c) that it's an open question whether nonhuman animals have cognitive states of belief. So far, the skeptic and I have no disagreement. But the skeptic goes on to assume (d) the conflation of attitudes of belief with cognitive states of belief.[19] Combining this false assumption with her true ones, she makes inferences to a few increasingly dubious conclusions. The skeptic infers that humans have cognitive states of belief, and that she has a decent epistemic handle on these cognitive states. She also infers that it is an open question whether nonhuman animals have attitudes of belief.

In contrast, rejecting the conflation of attitudes of belief with cognitive states of belief allows us to countenance humans' quotidian practices of attributing beliefs to nonhuman animals as by-and-large veridical. And it allows us to do so without making any unsubstantiated empirical assumptions about their cognitive capacities. Morgan's Canon—a dictum designed to guide the scientific attribution of cognitive states—doesn't warn directly against quotidian non-scientific attributions of attitudes to animals.

With that said, Morgan's Canon does warn, indirectly, against incautious attributions of attitudes. After all, there's always the danger of illicitly anthropomorphizing with respect to the dispositions that make up attitudes of belief. For example, I'll wrongly attribute a belief to my thermostat if I interpret it as being disposed (among other things) to metacognitively access its own representation of the temperature in my house, or to enthusiastically report the temperature with the (originally intentional) goal of informing fellow sentient beings how cold it is. Being mindful of Morgan's Canon would put a halt to that interpretation. We should be wary of anthropomorphic interpretations of behavior whenever fair anthropectic interpretations are on the table. And, as a matter of fact, my thermostat's report of the temperature doesn't much resemble a human report. An anthropectic interpretation (or at least an interpretation that invokes the intentions of the designer of the thermostat rather than those of the instrument itself) is much more reasonable.[20] Similarly, I may wrongly ascribe to a piping plover the belief that feigning a broken wing distracts predators, if I interpret the plover as being disposed (among other things) to purposively feign a broken wing because it occurrently judges this trickery to be the ideal way to lure a predator away from its babies. Properly interpreted, Morgan's Canon helps us avoid such facile anthropomorphic interpretations of animal behavior, and thus indirectly informs lay belief attribution.

So, do nonhuman animals believe? There are tricky issues concerning whether nonhuman animals have cognitive states of belief. Even if philosophers universally agreed about whether *humans* have cognitive states of belief—which they don't (Churchland 1981; Demeter, Parent, & Toon 2022)—they wouldn't agree about the nature of those cognitive states (Van Leeuwen & Lombrozo 2023). Do cognitive states necessarily involve language, or higher order thought (Davidson 1982)? Full-blown concepts (Stich 1983), or unicepts (Millikan 2017)? The ability to subconsciously perform predicate logic (Bermudez 2003)? If so, some species might boast proto-beliefs as opposed to full-blow human-like cognitive states of belief, or they may lack cognitive states of belief altogether. However, despite their many disputes, all of these philosophers agree that whether particular species have cognitive states of belief will ultimately be decided on the basis of a combination of careful theorizing about the nature of cognitive states of belief and careful experimental research into the representational capacities of the species in question.

For example, consider the question of animal belief through the lens of Mandelbaum's psychofunctional representationalism about cognitive states of belief. On Mandelbaum's view, there are psychofunctional laws governing belief, and whether somebody has a belief depends on whether their cognitive system's relation to the relevant representation is governed by the proper psychofunctional laws. There's a growing body of evidence supporting the hypothesis that many animal cognitive systems feature relations to representations that obey some of the same psychofunctional laws governing human beliefs (Porot & Mandelbaum 2022: 64–66). There's also very good evidence that all nonhuman cognitive systems are lacking in some relevant respects, including most obviously with respect to sophisticated linguistic capacities (Dennett 2017; Moore 2018). So, whether (and how) psychofunctionalists should attribute cognitive states of belief to animals depends on intricacies of both theory and data. In the first place, it depends on how, precisely, belief is defined in a psychofunctional theoretical framework. How much wiggle-room is there? Can a human and a moth share the belief that the light is on, despite the fact that their respective relations to representations play some very different functional roles in their respective cognitive systems? Do owls' night vision and bats' echolocation generate functionally identical beliefs that they sense a rat (Sterelny 1990: 98)? The definition resolving these questions would, among other things, also have to resolve the philosophical debates canvassed in the preceding paragraph. Moreover, whether animals have cognitive states of belief depends on precisely which psychological mechanisms animal cognitive systems are discovered to comprise—that is, per psychofunctionalism, precisely which psychological laws animal cognitive systems are discovered to enact.

These data are by-and-large still awaiting collection, and the theoretical debates are unsettled. Morgan's Canon thus recommends agnosticism. Per my originalist reading, and contrary to the received reading, it certainly doesn't direct us to reject the existence of cognitive states of belief in nonhuman animals. But it does direct us to avoid committing to their existence before we have duly considered other possible cognitive causes that might reasonably be surmised to produce animals' (apparently) flexible and goal-directed behaviors. By my reckoning, animal cognition research is still stuck in the thick of that process of due consideration.

Fortunately, we needn't determine whether species boast cognitive states of belief before assessing whether they boast attitudes of belief. If we avoid conflating the two, then the question of animal attitudes of belief becomes much more tractable. In particular, my model-theoretic interpretivism—the view that to believe[attitude] is to have a pattern of dispositions that fits a belief attributor's folk psychological model of what's involved in taking the world to be some way—does justice to the (common) intuition that many animals believe, without simply discarding as baseless the (not uncommon) intuition that they don't. I'll now discuss these clashing intuitions in turn.

Many people would find it absurd to deny that my cats believe it is dinner time (Spence, Osman, & McElligott 2017). Consider Raimond Gaita's appeal to this intuition.

> I don't *conjecture* whether [a dog] is the kind of creature who is sometimes warm and sometimes cold, who sometimes has pleasures and is sometimes in pain, who sometimes believes one thing and hopes or fears another. Nor do I *assume* it, or *take it as certain*—that is, for practical though not for philosophical or scientific purposes. I am *absolutely certain*; that is to say, I have not the slightest doubt. (Gaita 2002: 44)

Scientists don't know for certain what cognitive mechanisms underlie cats and dogs' attitudes of belief; nor do they know whether the same cognitive mechanisms underlie human and non-

human beliefs. Nevertheless, it seems like an affront to quotidian experience to deny that cats and dogs take the world to be some ways and not others—and, from Gaita's perspective, the lack of a scientific consensus concerning the mechanisms underlying these animal beliefs seems almost totally beside the point when it comes to assessing the veridicality of lay belief attribution practices.

Even in scientific contexts, maintaining skepticism about animal attitudes of belief would require discounting the substantial body of empirical evidence showing that many nonhumans are disposed to represent their environments in flexible and goal-directed manners (Saidel 2009; Lesson, Tinklenberg, & Andrews 2020). My model-theoretic interpretivism makes good on the intuition that animals believe—and takes seriously the evidence backing that intuition up—by allowing that animals of different species can share beliefs even in the face of considerable differences in underlying cognitive architecture. So long as different animals each live out a pattern that sufficiently fits a belief attributor's model, they share a belief. Moreover, on this view, models of nonhuman belief aren't necessarily derivative of models of human belief. As Gaita argues, "our ways of speaking about knowledge and belief have not been first and fully formed just in our lives with human beings and then applied conjecturally with animals" (2002: 72). Instead, our general-purpose models of particular beliefs may well be constructed and refined to cross (some) species boundaries when we grow up with animals in our lives.

One major advantage of a model-theoretic approach to folk psychology—over traditional versions of the theory theory and simulation theory—is that it neatly accounts for belief attributors' dual ability to think abstractly about the general conditions on possession of a given belief and to think practically about what it would take for a particular believer to possess that belief in a particular context. Like scientists, belief attributors have both "an understanding of a general structure or schematic pattern that can have many specific instantiations [and] the ability to construct *specific* hypothetical systems to deal with particular empirical cases" (Godfrey-Smith 2005: 4–5). By wielding this dual ability, "folk psychological attributors can rapidly put together specific, filled-out psychological profiles, to explain and predict the actions of individual agents" (6). By accounting for both general-purpose and believer specific models, model theory uniquely reveals how belief attributors "manage to systematize an extraordinary range of phenomena and understand them as different manifestations of the same general principles" (Maibom 2009: 374–375).

Thus, at the same time as they construct general-purpose models that cross species boundaries, belief attributors construct individualized models of belief for different species of believers. Unlike some other superficial accounts of belief, this view leaves room for animals of different species to believe in different styles (Curry 2022). For example, the model I wield to capture believing-it-is-dinner-time-like-a-cat differs in some clear ways from my model of believing-it-is-dinner-time-like-a-human. I expect the latter to involve dispositions to complain about hunger using words and hand gestures, whereas I expect the former to involve dispositions towards plaintive whines and frantic leg-rubbing.[21] My models of human beliefs generally emphasize metacognitive dispositions. Perhaps because I have been influenced by the scientific literature (Templer 2022), my models of feline beliefs do not. Recognizing styles of belief allows me to understand my cats more thoroughly than if I had solely wielded the blunt tool that is my general-purpose model of cross-species-belief-that-it-is-dinner-time. A model-theoretic interpretivism thus provides flexible tools for accounting for the subtleties of nonhuman animal belief.

It also elucidates why some nagging intuitive skepticism is justified. Many behaviors characteristic of believing differ dramatically from species to species. As such, members of some species (lobsters?) might sufficiently fit intraspecific models of belief but not interspecific

models, and members of other species (coral?) might not sufficiently fit any models of belief. The question of animal attitudes of belief must be taken up on a case-by-case basis, informed by both the propensities of candidate believers and the (various) models wielded by belief attributors.

Further complicating matters, there is still no solution to the epistemological problem of other minds that so troubled Morgan early in his career. Belief attributors don't have direct access to the thoughts and feelings of other animals—much less to their dispositions to think and feel. Contra Gaita, we can never know for sure how well animals fit our models of belief, because we can never know with absolute certainty that they possess many of the relevant dispositions. In the context of belief attribution to nonhuman animals, the problem of other minds isn't merely a radical skeptical hypothesis. Putting radical skepticism aside, Gaita is right that it should be frankly uncontroversial that cats and dogs feel pain. But even while putting radical skepticism aside, the problem of other minds engenders the equally frank assessment that we can't know— at least given the current state of the psychological and neuroscientific evidence—whether dogs are disposed to consciously isolate abstract qualities, or whether cats are disposed to entertain the passage of time itself. Without this knowledge, there will always be lingering questions about many of our attributions of attitudes of belief to nonhuman animals (as well as many of our attributions of attitudes of belief to humans with deficits in linguistic and other cognitive faculties).

I've explained how animals can be taken to have attitudes of belief even by theorists who are skeptical that they have cognitive states of belief. But it's worth noting at this juncture that it's also possible that animals have cognitive states of belief even if they lack pertinent attitudes of belief (in relation to some models thereof). On certain stringent, anthropocentric models—models people might be well-motivated to wield for some of our social, folk psychological purposes (Curry 2020; McGeer 2021)—the belief that it's dinner time would necessarily involve metacognitive dispositions: the disposition to think about the concept of time, for instance. If my cats lack these metacognitive dispositions, then they might not sufficiently fit the model in question to count as having the attitude of belief in question. Or they might only "sorta" have the attitude of belief in question (Dennett 2017): they might "in-between believe" (Schwitzgebel 2001). Nevertheless, on a psychofunctionalist account of cognitive architecture, cats might have the *very same*—not just *sorta* the same—cognitive state of belief that (in humans) generates that attitude of belief. The cats' inability to go meta could be chalked up to their lack of some other metacognitive capacity that (in humans) works together with the cognitive-state-of-belief-that-it-is-dinner-time to form metacognitive dispositions, including the disposition to think about the concept of time. In this possible scenario, cats would share an underlying cognitive state of belief with humans despite not sharing the sophisticated attitude of belief that is of special interest in some human social cognitive contexts.[22]

Investigators can figure out whether that possible scenario is actual by conducting careful and creative cognitive scientific research guided by Morgan's Canon. In light of the distinction between attitudes and cognitive states—and the lack of a consensus about whether (and which) animals have cognitive states—that ongoing research should treat cognitive states of belief independently from the dispositions that realize attitudes of belief. Nonhuman animals plausibly boast many of the latter—and animal cognition research can unveil the mechanics underlying them—whether or not they boast the former (and, if they do, whether their cognitive states of belief are best understood at a neural, subpersonal (e.g., psychofunctional), or personal level of explanation).

6 | MONKEY MINDREADERS

The distinction between attitudes and cognitive states also sheds light on the question of non-human animal belief attribution. In the decades since this question was first posed about chimpanzees (Premack & Woodruff 1978), psychologists and philosophers have designed a long series of increasingly ingenious experimental paradigms dedicated to ascertaining whether animals (including nonhumans and young and variously disabled humans) *really* attribute beliefs.[23]

For example, inspired by work on human infants (Southgate et al 2007), Christopher Krupenye and colleagues have introduced eye-tracking to animal cognition research. Chimpanzees, bonobos, and orangutans demonstrate the ability to pass two different eye-tracking false belief tasks, thereby (purportedly) also demonstrating that humans aren't the only great ape belief attributors. Ape subjects in both trials watched movies in which agents apparently formed false beliefs about the location of objects. Using an infrared eye-tracker, investigators recorded the parts of the movies to which the apes paid visual attention. These recordings reveal that the apes' gaze regularly anticipated that the agents would look for the objects where they had last seen them, rather than where they actually were.[24] Krupenye and colleagues write that these findings "suggest that apes solved the task by ascribing a false belief to the actor, challenging the view that the ability to attribute reality-incongruent mental states is specific to humans" (Krupenye, Kano et al 2016: 113). The researchers claim that the apes' anticipatory gaze is best explained by the hypothesis that they use the attribution of belief to successfully predict behavior.

Cecilia Heyes has objected that these findings are equally compatible with the hypothesis that apes are (not belief attributors but) mere "submentalizers" (Heyes 2014): creatures with the ability to predict behavior by way of low-level, domain-general psychological processes. On Heyes's interpretation, apes visually anticipate where agents will look for objects by picking up on behavioral cues that help them track agents' dispositions, without ever explicitly attributing the beliefs that tie those dispositions together. For example, Heyes suggests that the apes in the study may have tracked "the appearance and disappearance of the striking green shirt" (2017: 2) worn by the agent, and associated the reappearance of the green shirt with the location of the object when the green shirt was last on the scene.[25] In reply, Krupenye et al (2017) have followed Heyes's methodological suggestions and replicated their 2016 study while controlling for some features of the submentalizing hypothesis.

Although they disagree about whether apes are belief attributors, Heyes and the Krupenye team agree that this disagreement cuts right to the heart of the bigger question of how (indeed, whether) apes understand other minds. Putting this meta-agreement in a nutshell, Michael Tomasello has remarked that being able to pass the false belief task "means understanding that there exists a mental world distinct from the physical world" (quoted in *Duke Today* 2016). According to this mainstream line of thinking, if apes attribute beliefs, then they grasp that there is something going on inside other minds. If apes merely submentalize, then they track dispositions associated with other minds, but don't recognize inner states themselves. Properly controlled false belief tasks are supposed to provide evidence to pull these dichotomous possibilities apart.

However, several outspoken researchers are skeptical of the explanatory power of false belief tasks. According to these skeptics, there is a "logical problem" (Povinelli & Vonk 2003; Hurley & Nudds 2006; Penn & Povinelli 2007) with the methodology of false belief tasks. In Robert Lurz's (2011) terminology, most false belief tasks systematically fail to distinguish the "mindreading" hypothesis that animals attribute beliefs from the "behavior-reading" hypothesis that they track

complex behavioral dispositions (and thus merely submentalize). The skeptics' view is that this failure isn't a mere methodological problem to be solved by further experimental controls, but an experimentally intractable problem that arises from the epistemic logic of false belief tasks. By running false belief tasks, experimenters attempt to infer belief attribution from the observable behaviors of candidate belief attributors. But, logically speaking, all possible observable behaviors could be explained by the alternative hypothesis that subjects are mere submentalizers as opposed to full-blown belief attributors. Thus, false belief tasks can never settle their target question—or even more bleakly, if researchers running false belief tasks follow the letter of Morgan's Canon, they must always deny the capacity to attribute beliefs to animals.

Just as I've argued that charges of anthropomorphism in everyday belief attribution get their fangs from the conflation of attitudes and cognitive states, I'll now argue that the logical problem gets its fangs from the construal of attitudes of belief as unobservable causes productively intervening between observable stimuli and behaviors. If attitudes of belief are conflated with unobservable cognitive states, then false belief tasks will never enable researchers to distinguish belief attribution from submentalizing that tracks observable properties alone.[26]

Model-theoretic interpretivism about attitudes of belief dissolves the logical problem by rejecting the assumption that belief attributors *must* construe beliefs as inner causes, while simultaneously rejecting any conception of submentalizing that precludes the tracking of inner causes. In this framework, the logical problem is ill-posed. Attitudes of belief are patterns of dispositions to act, react, think, and feel. To submentalize is to track some of the behavioral, cognitive, and phenomenal dispositions that might partly compose beliefs. To attribute a belief is to recognize a particular sort of pattern among those dispositions. Heyes's question about whether the apes in the Krupenye study are attributing beliefs or merely submentalizing is a good one, but it shouldn't be cast as the question of whether they are attributing inner producing causes or outer behavioral dispositions. This question would, indeed, be susceptible to the logical problem, just as nearly all questions about inner states are susceptible to the problem of other minds. A better question is whether the apes recognize whole patterns of dispositions as states of taking the world to be some way, or, instead, pick up on individual dispositions in a less systematic manner. In other words, are the apes wielding folk psychological models of beliefs? Or are they merely tracking and responding to dispositions that would go into humans' models of beliefs? There are empirically tractable differences between these hypotheses, having to do with the generality and exportability of the apes' social practices. For example, if apes were able to perform similarly successfully on a variety of eye-tracking false belief tasks—tasks that systematically varied the observable characteristics of the situation while requiring the ape to track the same belief to be successful—then this would provide evidence that they were attributing beliefs rather than tracking individual dispositions.[27]

Whether animals attribute inner thoughts (such as tokens of inner speech) or feelings (such as aches and pangs) to each other is an open question. Animals of certain species might, like humans, do so in a systematic fashion, weaving these inner states into models of their conspecifics' attitudes and traits. Or they might do so by submentalizing, homing in on a specific conspecific's feeling of pain as she howls. Alternatively, nonhuman animals might be ignorant of the existence of phenomenologies other than their own. Fascinating though it is, this isn't a question that researchers need to settle decisively in order to make progress in the debate about animal mindreading. False belief paradigms have provided genuine—if preliminary—evidence that nonhuman great apes attribute belief-like states to one another, regardless of whether they attribute inner states.[28]

Of course, researchers shouldn't be too quick to jump to conclusions. Heyes might be right that submentalizing is equal to any given false-belief task. Moreover, as philosophers have emphasized (Andrews 2017; Halina 2017; Boyle 2019), there is good evidence that apes and monkeys

attribute visual access to others (Hare et al 2000, 2001; Melis et al 2006; Call & Tomasello 2008; Karg et al 2015; Sanchez Amero et al 2020; Aychet et al 2020), as do other animals including crows (Bugnyar et al 2016), goats (Kaminski et al 2006), and some pigs (Byrne et al 2001). Whereas believing is a general state of taking the world to be some way, not tied to any particular sense modality, seeing is the limited state of taking the world to be some way via occurrent vision. Whether seeing is (sufficient for) believing depends on the folk psychological model of belief at play. Importantly, an animal might wield a model of seeing without wielding any models of believing whatsoever. The apes in the Krupenye study might have accomplished their task (not by attributing beliefs or submentalizing but) by attributing visual access—and the memory of visual access—to the agent. To make real strides in understanding whether nonhuman animals attribute beliefs, researchers must figure out which experimental situations demand that the animals under investigation engage in belief attribution, as opposed to either submentalizing or the attribution of mental states other than beliefs. To their credit, some comparative psychologists have started working in this direction (for overviews, see Andrews 2020; Lewis & Krupenye 2022; Phillips et al 2021).

Rejecting the conflation of attitudes of belief with cognitive states of belief would push these researchers to make even more fine-grained distinctions, leading to more nuanced experimental paradigms. Belief—the general (as opposed to sense-modality-specific) state of taking the world to be some way—is a fairly stable category through which humans, across cultures, understand animals (including each other). I have canvassed several contributions a metaphysical commitment to model-theoretic interpretivism about attitudes of belief would make to animal cognition research. Perhaps the foremost contribution among them would be the revelation that different belief attributors—particularly belief attributors who differ enough to belong to different species—are likely to construe beliefs in different ways, and thereby metaphysically constitute beliefs in different ways (Curry 2020, 2021a, 2022).

If that's right, then any research program that takes the question of what makes humans unique as its alpha and omega will obscure as much about nonhuman cognition as it illuminates. Octopus social practices—such as they are—are worthy of study as octopus social practices, not qua potentially proto-human social practices. Model-theoretic interpretivist research on nonhuman animal mindreading would focus on investigating how particular nonhuman animals understand each other, rather than narrowly focusing on whether, like humans, any nonhuman animals understand each other in terms of human-like models of human-like beliefs. Even if nonhuman animals are belief attributors, their models might differ from humans' in surprising ways. If, rather than solely being interested in what makes humans unique, researchers are interested in understanding the psychological capacities of nonhuman animals in their own right, then they should be interested in figuring out the idiosyncratic contours of how particular (species of) animals understand other minded agents. There's only so much to be gleaned through an exclusive focus on figuring out how closely nonhumans' social cognitive practices resemble our own.

I am far from against anthropomorphism. As Morgan recognized, thinking about nonhuman animal minds in terms of human attitudes and cognitive states is often the only effective way to get comparative psychological research up and running. However, I am against anthropocentrism. I am thus also against any presumption in favor of anthropomorphism concerning nonhuman belief attribution. Due consideration of Morgan's Canon (shed of its nascent epistemological and metaphysical trappings) should lead researchers to consider behavioristic interpretations of the evidence about animal behavior, but it should also lead them to consider non-anthropomorphic mentalistic interpretations. Perhaps some nonhuman animals are belief attributors, just like humans. Before settling for that default anthropomorphism, however, researchers should use

Morgan's Canon to explore alternative, non-doxastic ways in which nonhuman animals might model each others' minds. *'Mindreading versus behavior-reading'* is a false and counterproductive dichotomy. Some nonhuman animals are mindreaders—attitude attributors—even if they aren't belief attributors, and perhaps even if they don't attribute any inner states to other animals. For example, many animals attribute states of seeing to their conspecifics, including predominantly asocial animals like octopi, who are unlikely to be belief attributors (Godfrey-Smith 2016: 71).

If nonhuman animals don't attribute beliefs to one another, it isn't necessarily because they are generally cognitively deficient (relative to human beings). Instead, it might be because different animals model minds differently, to suit their different needs in their different social environments. Even if (as I suspect) some nonhuman animals do attribute beliefs to one another, the details of their models of belief doubtlessly differ in interesting respects from species-typical human models. As Morgan aptly remarked in a related context, "given two different minds and the same facts, how different are the products!" (1891: 335). For example, I am confident in speculating that nonhuman animals don't, as a rule, model beliefs as involving dispositions to assent to propositions (Moore 2020), much less as involving dispositions to assent to propositions linguistically. Philosophers of animal minds and comparative psychologists alike would do well to heed the differences in how different animals construe each others' minds, without antecedently setting human belief attribution as the interspecific standard.

7 | AFTERMATH: ATTITUDES OF BELIEF ARE SUPERFICIAL, COGNITIVE STATES OF BELIEF ARE DEEP

Schwitzgebel's superficial approach to attitudes of belief clearly captures the gist of the phenomenon he's interested in: that is, the nature of the beliefs that people veridically attribute to themselves, each other, and other animals—and which nonhuman animals may or may not attribute to themselves or others—in everyday life. Mandelbaum's psychofunctional approach to cognitive states of belief, meanwhile, is at the very least an important contender among theories of the constituent moving parts of (human and nonhuman) cognitive architectures. Despite Schwitzgebel and Mandelbaum each casting their work in terms of its opposition to the other's view, the two camps needn't be in conflict, since they seek to understand the metaphysics of distinct (and perfectly compatible) phenomena. Indeed, the foregoing article has supplemented my previous theoretical arguments with a pragmatic reason to put a stop to the fighting: distinguishing between the camps' phenomena of interest does a great deal to clarify the parameters and stakes of scientific debates about anthropomorphism, animal belief, and animal belief attribution. With luck, those clarifications can in turn illuminate a path towards more fruitful study of the diverse ways in which animals understand their own minds and the minds of others.

Insofar as applying the distinction between attitudes and cogs has similarly salutary consequences for inquiry into other topics of scientific, scholarly, or lay interest, the pragmatic argument in its favor will be further strengthened. As always, there is much work to be done, on many fronts—better to have allies than foes wherever feasible.

were delivered at the 2012 York University Graduate Philosophy Conference in Toronto, the 2012 meeting of the Society for Philosophy and Psychology in Boulder, the 2013 Penn History and Philosophy of Science Workshop in Philadelphia, and the 2013 meeting of the International Society for the History, Philosophy, and Social Studies of Biology in Montpellier.

ENDNOTES

[1] See Schwitzgebel (2001, 2002, 2021, 2022). The superficial camp has previously been led by dispositionalists like Ryle (1949) and Baker (1995), as well as interpretivists like Dennett (1987), Davidson (2001), and Mölder (2010). Elsewhere, I've argued that dispositionalism and interpretivism ought to be understood as two sides of the same coin (Curry 2021b, 2023).

[2] See Mandelbaum (2014); Quilty-Dunn & Mandelbaum (2018); Bendana & Mandelbaum (2021); Porot & Mandelbaum (2022). The deep camp has previously been led by Mandelbaum's fellow psychofunctionalists, including most prominently Fodor (1987), as well as teleofunctionalists like Millikan (1984) and Dretske (2000).

[3] Other proponents of reconciliation include Dewhurst (2021), Poslajko (2022), and (approaching the issue from a different angle) Van Leeuwen & Lombrozo (2023).

[4] It's worth flagging that, unlike Fodorian psychofunctional representationalism, Mandelbaumian psychofunctional representationalism doesn't centrally rely on the assumption that attitudes of belief can be unproblematically conflated with cognitive states of belief. Quilty-Dunn and Mandelbaum do conflate the two, but their conflation is a vestigial assumption rather than a working premise in their arguments. Indeed, insofar as they stress that the "generalizations about belief [established by cognitive science] are counterintuitive" (2017: §3.5), Mandelbaumian representationalism might draw strength from the distinction (whereas Fodor asserted that psychological laws mostly affirm what his Granny already knew). Similarly, in the other camp, Schwitzgebel fails to distinguish between attitudes of belief and cognitive states of belief, and thereby casts his dispositionalism as a rival to psychofunctionalist representationalism. But nothing important in his theory hangs on that move; by my lights, drawing the distinction would only strengthen Schwitzgebel's position by showing how psychofunctionalism needn't be a rival to his theory after all.

[5] Per Morgan's hierarchical theory of cognitive architecture, less sophisticated cognitive capacities are necessary evolutionary building blocks for more sophisticated cognitive capacities (such as the power of analysis). As I'll argue in §4, this theory—which leads Morgan to speak in terms of "higher" and "lower" faculties in the "psychological scale"—is partly responsible for the Canon's historical misfires.

[6] Shettleworth cites Mitchell (2005) and Wynne (2007) to back up this assertion.

[7] See Andrews (2020), Rosati, Machanda, and Slocombe (2022), and Lewis and Krupenye (2022) for up-to-date surveys of some of the 21st century research that proceeds from this comfort.

[8] To foreshadow the discussion to come, notice that Fodor's formulation of Fodor's Pop Gun—which, to be fair, follows the wording in Morgan (1894)—leaves out the word "fairly".

[9] This isn't to deny the interpretive quandaries facing the Canon originalist. Just as a constitutional originalist must be creative in determining how the U.S. founding fathers' intentions apply to modern society, the Canon originalist must be creative in determining how Morgan's intentions apply to modern comparative psychology, with its cognitive scientific (rather than introspectionist) methodology and its rejection of Morgan's account of cognitive architecture in favor of a menu of competing theories which posit varying levels of modularity and hierarchical (or heterarchical) structure.

[10] For an equal parts amusing and disturbing illustration of this obsession, see Morgan's (1883a, 1883b, 1887) reports in *Nature* of his rigorous and inventive experiments falsifying the hypothesis that scorpions commit suicide. (Morgan and other late-19th century naturalists took that hypothesis to directly challenge the theory of evolution by natural selection—if the overarching telos of life is reproduction and survival, then how could an animal have evolved to kill itself?)

[11] Mitchell & Hamm (1997); Herzog & Galvin (1997); Kelemen (1999, 2003); Johnson et al. (2001); Knight et al. (2004); Caslet & Kelemen (2008); Rosset (2008); Kelemen & Rosset (2009); Tao, McCarthy, & Scholl (2010); Maust-Mohl et al. (2012); Howell et al. (2013); Hawkins & Williams (2016).

[12] It doesn't matter for present purposes, but the relevant capacities of piping plovers actually might be a fair bit more subtle than this sketch suggests. Dennett's (2017: Ch. 5) full discussion of plovers' behavioral-flexibility-within-sphexish-limits is enlightening.

[13] My interpretation of Morgan's Canon differs slightly—but significantly—from Clatterbuck's. Clatterbuck argues that the Canon results from the "same vera causa inferential strategy" as Darwin's commitment to the continuity of human and animal minds: "we ought to admit no more causes of natural things, than such as are both true and sufficient to explain their appearances" (Clatterbuck 2016: 1). In my view, Clatterbuck is right that Morgan frequently employed *vera causae* reasoning in conjunction with Morgan's Canon, but wrong to go on to assert that the Canon is "inert" without it. On my reading, Morgan's Canon serves only to exhort researchers to consider alternatives to default hypotheses. In Morgan's hands, this means considering alternatives lower on the psychological scale than the default anthropomorphic hypothesis. But the Canon itself can be severed from Morgan's epistemological and metaphysical commitments that give it this particular meaning. In principle, then, the Canon could also be fruitfully employed alongside a rejection of the *vera causa* inferential strategy. For instance, a cognitive ethologist could put Morgan's Canon to good use—making sure to canvas the full range of fair interpretive possibilities, including behavioristic interpretations—despite being driven (perhaps by independently motivated theoretical commitments) to admit a cause of a behavior that goes beyond those that are sufficient to explain its appearances.

[14] This language of a "folk psychological ontology" and a distinct "cognitive ontology" comes from Dewhurst (2021), who independently argues for (roughly) the same distinction as mine.

[15] My account of attitudes of belief is—explicitly (Curry 2020, 2022, 2023)—a refinement of Schwitzgebel's superficial account of belief (since, in his relativization of beliefs to stereotypes, Schwitzgebel himself already bakes into his dispositionalism a version of interpretivism as well as a model-theoretic approach to folk psychology). My principal disagreement with Schwitzgebel arises from my distinction between attitudes and cognitive states, which Schwitzgebel collapses (arguing on pragmatic grounds that only attitudes of belief deserve the title of 'belief').

[16] This definition marries three theories: interpretivism about belief (Dennett 1998; Davidson 2001; Molder 2010), dispositionalism about belief (Ryle 1949; Baker 1995; Schwitzgebel 2002), and a model-theoretic approach to folk psychology (Maibom 2003, 2009; Godfrey-Smith 2005; Spaulding 2018; Moore 2020). I've argued (Curry 2021b, 2023) that all interpretivisms are dispositionalisms (and vice versa), since interpretivists take interpretive schemes to identify beliefs with patterns of dispositions, and dispositionalists take patterns of dispositions to emerge *as beliefs* relative to interpretive schemes. Andrews (2012: 204–205) first suggested (without endorsing) the marriage of a model-theoretic approach to belief attribution with interpretivism/dispositionalism about belief.

[17] For those arguments, see Curry 2018, 2020, 2021a; for sympathetic lines of argument, see also Dewhurst 2021; Hutto 2022; Poslajko 2022.

[18] This isn't to say that lay belief attributors can't be wrong (Curry 2021b). The cats might be pacing because they see a fly and lack the belief that it is dinner time. In that case, I am wrong about which relevant pattern of dispositions they possess—about which pattern truly contextualizes their behavior (thickly described), relative to my folk psychological models—and thus wrong about what they believe.

[19] To put a face on this skeptic, consider Fodor's Granny (1987: 6), whom he swears has forgotten more about belief than any philosopher will ever know, and who, like Fodor himself, proudly conflates attitudes of belief with cognitive states of belief.

[20] As this example illustrates, my interpretivism about attitudes of belief is perhaps less permissive than Dennett's (1998: 327–331), without being nearly as strict as Davidson's (1982).

[21] The implicit suggestion here—that styles of belief depend largely on how believers are embodied—could be important for the scientific study of animal attitudes. Endless forms most beautiful and most wonderful invite endless variety in the mentalistic interpretations that render animal life intelligible.

[22] Relatedly, models of what philosophers have variously termed "assent" (de Sousa 1971), "opinion" (Dennett 1978), "acceptance" (Cohen 1992), and "superbelief" (Frankish 2004)—that is, patterns of dispositions to affirm linguistically structured propositions—are a particular variety of specialized model of belief which cast humans as specialized believers—assenters—and plausibly exclude all nonhuman (because nonlinguistic) animals.

[23] The philosophers Jonathan Bennett, Daniel Dennett, and Gilbert Harman all proposed false-belief tasks in influential commentaries on Premack and Woodruff's landmark article. See Lewis & Krupenye (2022) for an overview of how the literature on primates (including humans) has evolved from there.

[24] Videos available at http://science.sciencemag.org/content/suppl/2016/10/06/354.6308.110.DC1

[25] Andrews (2018), meanwhile, invoking Heider and Simmel (1944), argues that the methodology suggested by Heyes results from flawed theorizing about apes' understanding of agency.

[26] Nor will false belief tasks' more sophisticated cousins do the trick (Buckner 2014).

[27] As supplements to the practical suggestions put forward by Heyes (2015), Butterfill & Apperly (2013) provide an excellent discussion of how to distinguish experimentally between belief attribution and mere submentalizing (or, as they call it, "minimal mindreading").

[28] In addition to the Krupenye et al (2016) study discussed in the main text, see Kano et al (2019); Krachun et al (2009); Buttelmann et al (2017); and Hayashi et al (2020).

REFERENCES

Allen, C. & Bekoff, M. (1997). *Species of Mind: The Philosophy and Biology of Cognitive Ethology*. MIT Press.

Almagro Holgado, M. & Fernandez Castro, V. (2019). "The Social Cover View: A Non-epistemic Approach to Mindreading", *Philosophia* 48 (2): 483–505.

Andrews, K. (2012). *Do Apes Read Minds? Toward a New Folk Psychology*. MIT Press.

Andrews, K. (2017). "Chimpanzee mindreading: Don't stop believing", *Philosophy Compass*, 12(1).

Andrews, K. (2018). "Apes track false beliefs but might not understand them", *Learning & Behavior*, 46: 1, 3–4.

Andrews, K. (2020). *How to Study Animal Minds*. CUP.

Andrews, K & Huss, B. (2013). "Anthropomorphism, Anthropectomy, and the Null Hypothesis." *Biology and Philosophy*, 29(5), 711–729.

Apperly, I. (2011). *Mindreaders: The Cognitive Basis of "Theory of Mind"*. Psychology Press.

Arnet, E. (2019). "Conwy Lloyd Morgan, Methodology, and the Origins of Comparative Psychology." *J Hist Biol*. 52(3): 433–461.

Aychet, J., Pezzino, P., Rossard, A., Bec, P., Blois-Heulin, C., & Lemasson, A. (2020). "Redcapped mangabeys (*Cercocebus torquatus*) adapt their interspecific gestural communication to the recipient's behaviour." *Scientific Reports*, 10(1), 12843.

Baker, L.R. (1995). *Explaining Attitudes: A Practical Approach to the Mind*. MIT Press.

Bendana, J. & Mandelbaum, E. (2021). "The Fragmentation of Belief." In Cristina Borgoni, Dirk Kindermann & Andrea Onofri (eds.), *The Fragmented Mind*. OUP.

Bermúdez, J. (2003). *Thinking Without Words*. OUP.

Boesch, C. (2007). "What makes us human (*Homo sapiens*)? The challenge of cognitive cross-species comparison", *Journal of Comparative Psychology* 121: 227–240.

Böhnert, M., & Hilbert, C. (2018). "Other minds than ours': a controversial discussion on the limits and possibilities of comparative psychology in the light of C. Lloyd Morgan's work." *History and Philosophy of the Life Sciences*, 40(3), 1–28.

Boyle, A. (2019). "Mapping the Minds of Others", *Review of Philosophy and Psychology*, 10: 747–767.

Bradatan, C. (2006). "'One is All, and All is One': The Great Chain of Being in Berkeley's *Siris*." In: Donald, D., O'Gorman, F. (eds) *Ordering the World in the Eighteenth Century. Studies in Modern History*. Palgrave Macmillan, London.

Buckner, C. (2013). "Morgan's Canon, Meet Hume's Dictum: Avoiding Anthropofabulation in Cross–Species Comparisons" *Biology & Philosophy* 28:5, 853–871.

Buckner, C. (2014). "The Semantic Problem(s) with Research on Animal Mindreading", *Mind and Language*, 29(5): 566–589.

Bugnyar, T., Rebar, S., & Buckner, C. (2016). "Ravens Attribute Visual Access to Unseen Competitors", *Nature Communications* (7).

Butterfill, S. & Apperly, I. (2013). "How to Construct a Minimal Theory of Mind" *Mind and Language* 28 (5):606–637.

Buttelmann, D., Buttelmann, F., Carpenter, M., Call, J., & Tomasello, M. (2017). "Great apes distinguish true from false beliefs in an interactive helping task", *PLoS ONE* 12(4): e0173783.

Byrne, R., Mendl, M., Devereux, C., & Held, S. (2001). "Behavior of Domestic Pigs in a Visual Perspective Taking Task", *Behaviour*, 138(11-12), 1337–1354.

Call, J. & Tomasello, M. (2008). "Does the chimpanzee have a theory of mind? 30 years later", *Trends in Cognitive Science*, 12(5):187–192.

Cheney, D. & Seyfarth, R. (1992). *How Monkeys See the World*. Chicago University Press.

Chomsky, N. (1959). "Review of Skinner's *Verbal Behavior*", *Language* 35:26–58.

Clatterbuck, H. (2016). "Darwin, Hume, Morgan, and the *verae causae* of psychology", *Studies in History and Philosophy of Biological and Biomedical Sciences* 60: 1–14.

Cohen, L.J. (1992). *Belief and Acceptance*. OUP.

Costall, A. (1993). "How Lloyd Morgan's Canon backfired", *Journal of the History of the Behavioural Sciences*, 29, 113–122.

Cook, R. (1999). "Morgan's comments on the need to see the process leading to an "intelligent" behavior", http://pigeon.psy.tufts.edu/psych26/morgan.htm

Crockford, C., Wittig, R., Mundry, R., & K. Zuberbuhler (2012). "Wild Chimpanzees Inform Ignorant Group Members of Danger", *Current Biology* 22:2, 142–146.

Curry, D.S. (2018). "Beliefs as inner causes: The (lack of) evidence", *Philosophical Psychology*, 31:6, 850–877.

Curry, D. S. (2020). "Interpretivism and norms", *Philosophical Studies* 177(4), 905–930.

Curry, D. S. (2021a). "How beliefs are like colors", *Synthese* 199: 3–4, 7889–7918.

Curry, D. S. (2021b). "Interpretivism without judgement-dependence", *Philosophia*, 49:2, 611–615.

Curry, D. S. (2022). "Belief in character studies", *American Philosophical Quarterly* 59:1, 27–42.

Curry, D. S. (2023). "Why dispositionalism needs interpretivism: a reply to Poslajko", *Philosophia*, 51: 2139–2145.

Dacey, M. (2017). "Anthropomorphism as Cognitive Bias", *Philosophy of Science*, 84, no. 4: 1152–1164.

Davidson, D. (1982). "Rational Animals" *Dialectica* Vol 36:4, 317–327.

Davidson, D. (2001). *Subjective, Intersubjective, Objective*. OUP.

Demeter, T., Parent, T., & Toon, A. (2022). *Mental Fictionalism*. Routledge.

Dennett, D.C. (1978). *Brainstorms: Philosophical Essays on Mind and Psychology*. MIT Press.

Dennett, D.C. (1984). *Elbow Room: The Varieties of Free Will Worth Wanting*. MIT Press.

Dennett, D.C. (1987). *The Intentional Stance*. MIT Press.

Dennett, D.C. (1998). *Brainchildren: Essays on Designing Minds*. MIT Press.

Dennett, D.C. (2017). *From Bacteria to Bach and Back: The Evolution of Minds*. Norton.

de Sousa, R. (1971). "How to Give a Piece of Your Mind: Or, the Logic of Belief and Assent", *Review of Metaphysics* 25 (1):52–79.

Dewey, J. (1925). *Experience and Nature*. Dover.

Dewhurst, J. (2017). "Folk psychology and the Bayesian brain." In Metzinger & Wiese (Eds.), *Philosophy and Predictive Processing* 9: 1–13.

Dewhurst, J. (2021). "Neurocognitive and Folk Psychological Ontologies." In Calzavarini, F. & Viola, M. (Eds.), *Neural Mechanisms*, 311–334. Springer.

Dewsbury, D.A. (1984). *Comparative Psychology in the Twentieth Century*. Hutchinson Ross.

Dixon, E. (1982). "The Limits of Animal Intelligence", *Nature* 46:1191, 392–393.

Dretske, F. (2000). *Perception, Knowledge, and Belief*. CUP.

DukeToday. (2016). "Apes understand that some things are all in your head", https://today.duke.edu/2016/10/falsebeliefs

Fitzpatrick, S. & Goodrich, G. (2017). "Building a Science of Animal Minds: Lloyd Morgan, Experimentation, and Morgan's Canon" *Journal of the History of Biology* 50, 525–569.

Fodor, J. (1987). *Psychosemantics*. MIT Press.

Fodor, J. (1999). "Not so Clever Hans", *London Review of Books*, February 4, 12–13.

Frankish, K. (2004). *Mind and Supermind*. CUP

Gaita, R. (2002). *The Philosopher's Dog*. Melbourne.

Galef, B. (1996). "Historical origins: the making of a science" In L. Houck and L. Drickamer (eds), *Foundations of Animal Behavior: Classic Papers with Commentaries*. Chicago University Press.

Gao, T., McCarthy, G. & Scholl, B. 2010. "The wolfpack effect: Perception of animacy irresisbly influences interactive behavior." *Psychological Science* 21 (12): 1845–1853.

Godfrey-Smith, P. (2005). "Folk Psychology as a Model" *Philosophers' Imprint* 5:6, 1–16.

Godfrey-Smith, P. (2016). *Other Minds: The Octopus, the Sea, and the Deep Origins of Consciousness*. Farrar, Straus, and Giroux.

Halina, M. (2017). "What apes know about seeing" in (K. Andrews & J. Beck, Eds.) *Handbook of Philosophy of Animal Minds*, pp. 238–246. Routledge.

Hare, B., Call, J., Agnetta & Tomasello, M. (2000). "Chimpanzees know what conspecifics do and do not see." *Animal Behaviour* 59 (4): 771–785.

Hare, B., Call, J. & Tomasello, M. (2001). "Do chimpanzees know what conspecifics know? *Animal Behaviour* 61 (1): 139"151.

Hawkins, R.D. and Williams, J.M. (2016) "Children's Beliefs about Animal Minds (Child-BAM): Associations with Positive and Negative Child–Animal Interactions." *Anthrozoos* 29, 503–519.

Hayashi, T., Akikawa, R., Kawasaki, K., Egawa, J., Minamimoto, T., Kobayashi, K., Kato, S., Hori, Y., Nagai, Y., Iijima, A., Someya, T., & Hasegawa, I. (2020). "Macaques Exhibit Implicit Gaze Bias Anticipating Others' False-Belief-Driven Actions via Medial Prefrontal Cortex." *Cell Reports*, 30(13), 4433–4444.e5.

Heider, F. & Simmel, M. (1944). "An experimental study in apparent behavior". *The American Journal of Psychology*, 57, 243–259.

Herzog, H.A. and Galvin, S. (1997) "Common Sense and the Mental Lives of Animals: An Empirical Approach." In *Anthropomorphism, Anecdotes and Animals* (Mitchell, R. W. et al., eds), pp. 237–253, State University of New York Press

Heyes, C. (2014). "Submentalizing: I Am Not Really Reading Your Mind" *Perspectives on Psychological Science* 9: 131–143.

Heyes, C. (2015). "Animal mindreading: what's the problem?" *Psychonomic Bulletin and Review* 22, 313–327.

Heyes, C. (2017). "Apes submentalize" *TICS* 21: 1–2.

Howell, T.J. et al. (2013) "The Perceptions of Dog Intelligence and Cognitive Skills (PoDIaCS)" Survey. *J. Vet. Behav.* 8, 418–424

Hurley, S. & Nudds, M. (2006). *Rational Animals?* OUP.

Hutto, D. (2022). "A Brickhouse Defense of Folk Psychology." In Demeter, T., Parent, T., & Toon, A. (Eds.), *Mental Fictionalism*. Routledge.

Johnson, S.C., Booth, A., & O'Hearn, K. (2001). "Inferring the goals of a nonhuman agent". *Cognitive Development*, 16, 637–656.

Kaminski, J., Call, J., & Tomasello, M. (2006). "Goats' behaviour in a competitive food paradigm: Evidence for perspective taking?" *Behaviour*, 143(11), 1341–1356.

Kano, F., Krupenye, C., Hirata, S., Tomonaga, M., & Call, J. (2019). "Great apes use selfexperience to anticipate an agent's action in a false-belief test." *Proceedings of the National Academy of Sciences*, 116(42), 20904–20909.

Karg, K., Schmelz, M., Call, J., Tomasello, M. (2015). "Chimpanzees strategically manipulate what others can see." *Anim Cogn*. 18:1069–1076.

Karten, H. (2013). "Neocortical Evolution: Neuronal Circuits Arise Independently of Lamination", *Current Biology* 23 1:R12–15.

Keijzer, F. (2013). "The Sphex story: How the cognitive sciences kept repeating an old and questionable anecdote", *Philosophical Psychology* 26(4): 502–509.

Kelemen, D. (1999). "The scope of teleological thinking in preschool children". *Cognition*, 70, 241–272.

Kelemen, D. (2003). "British and American children's preferences for teleological–functional explanations of the natural world". *Cognition*, 8, 201–221.

Kelemen, D., & DiYanni, C. (2005). "Intuitions about origins: Purpose and intelligent design in children's reasoning about nature". *Journal of Cognition and Development*, 6, 3–31.

Kelemen, D. & Rosset, E. (2009). "The Human Function Compunction: Teleological explanation in adults". *Cognition*, 111, 138–143.

Kelemen, D., Rottman, J. & Seston, R. (2013). "Professional physical scientists display tenacious teleological tendencies. Purpose–based reasoning as a cognitive default." *Journal of Experimental Psychology: General*, 142(4), 1074–1083.

Knight, S. et al. (2004). "Attitudes towards animal use and belief in animal mind." *Anthrozoos* 17, 43–62.

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

Krupenye, C., Kano, F., Hirata, S., Call, J., & Tomasello, M. (2016). "Great apes anticipate that other individuals will act according to false beliefs". *Science*, 354(6308), 110–114.

Krupenye, C., Kano, F., Hirata, S., Call, J., and Tomasello, M. (2017). "A test of the submentalizing hypothesis: Apes' performance in a false belief task inanimate control". *Communicative & Integrative Biology*. 10(4): e1343771.

Lakatos, I. (1978). *The Methodology of Scientific Research Programs*. CUP.

Lesson, S., Tinklenberg, B. & Andrews, K. (2020). "Belief and representation in nonhuman animals" in Robins, Symons & Calvo (eds.), *The Routledge Companion to Philosophy of Psychology*, Routledge. Ch. 23.

Lewis, L.S. & Krupenye, C. (2022). "Theory of Mind in Nonhuman Primates", In B.L. Schwartz & M.J. Beran (Eds.) *Primate Cognitive Studies*. CUP, 439–482.

Lombrozo, T. (2010). "Causal–explanatory pluralism: How intentions, functions and mechanisms influence causal ascriptions". *Cognitive Psychology* 61 303–332.

Lurz, R. (2011). *Mindreading Animals: The Debate Over What Animals Know about Other Minds*. MIT Press.

Maibom, H. (2003). "The Mindreader and the Scientist", *Mind & Language* 18:3, 296–315.

Maibom, H. (2009). "In Defence of (Model) Theory Theory", *Journal of Consciousness Studies*, 16, 360.

Malle, B. (2004). *How the Mind Explains Behavior*. MIT Press.

Mandelbaum, E. (2014). "Thinking is Believing", *Inquiry* 57: 1, 55–96.

Maust-Mohl, M. et al. (2012) "Wild Minds: What People Think about Animal Thinking." *Anthrozoos* 25, 133–147.

McGeer, V. (2021) Enculturating Folk Psychologists. *Synthese* 199, 1039–1063.

Millikan, R. (1984). *Language, Thought and Other Biological Categories*. MIT Press.

Millikan, R. (2017). *Beyond Concepts: Unicepts, Language, and Natural Information*. OUP.

Mitchell, S. (2005). "Anthropomorphism and Cross–Species Modelling", in Daston & Mittman (eds.) *Thinking with Animals: New Perspectives on Anthropomorphism (Columbia)*, 100–108,

Mitchell, R.W. and Hamm, M. (1997) "The interpretation of Animal Psychology: Anthropomorphism or Behavior Reading?" *Behaviour* 134, 173–204

Mölder, B. (2010). *Mind Ascribed: An Elaboration and Defense of Interpretivism*. John Benjamins Publishing.

Moore, R. (2018). "Gricean communication, language development, and animal minds." *Philosophy Compass* 13 (12):e12550.

Moore, R. (2020). "The cultural evolution of mind-modelling." *Synthese* 199 (1-2):1751-1776.

Morgan, C.L. (1883a). "Suicide of Scorpions." *Nature* 27, 313–314

Morgan, C.L. (1883b). "Scorpion Suicide." *Nature* 27, 530

Morgan, C.L. (1884). "Instinct." *Nature* 29, 370–374

Morgan, C.L. (1887). "Scorpion Virus." *Nature* 35, 535

Morgan, C.L. (1891). *Animal Life and Intelligence*. Ginn & Company.

Morgan, C.L. (1894). *An Introduction to Comparative Psychology*. Walter Scott Publishing Company.

Morgan, C.L. (1903). *An Introduction to Comparative Psychology*. Second Edition. Walter Scott Publishing Company.

Morgan, C.L. (1930a). "Autobiography of C. Lloyd Morgan" in Murchison, C. (Ed.), *History of Psychology in Autobiography* Vol. 2, Springer, 237–264.

Morgan, C.L. (1930b). *The Animal Mind*. Edward Arnold.

Penn, D. & Povinelli, D. (2007). "On the lack of evidence that non–human animals possess anything remotely resembling a "theory of mind"" *Philos Trans R Soc Lond B Biol Sci*.; 362(1480): 731–744.

Phillips, J., Buckwalter, W., Cushman, F., Friedman, O., Martin, A., Turri, J., Santos, L. & Knobe, J. (2021). "Knowledge before Belief." *Behavioral and Brain Sciences* 44, E140: 1–37.

Porot, N. & Mandelbaum, E. (2022). "The science of belief: A progress report." *WIREs Cognitive Science*. Online First.

Poslajko, K. (2022) "The Lycan-Stich argument and the plasticity of "belief"." *Erkenntnis* 87(3), 1257–1273.

Povinelli, D. & Vonk, J. (2003). "Chimpanzee Minds: Suspiciously Human?" *TICS* 7:4, 157–160.

Premack, D. & Woodruff, G. (1978). "Does the chimpanzee have a theory of mind?" *Behavioral and Brain Sciences* 4 (4):515–629.

Quilty-Dunn, J. & Mandelbaum, E. (2018). "Against dispositionalism: belief in cognitive science." *Philosophical Studies* 175 (9):2353-2372.

Radick, G. (2007). *The Simian Tongue: The Long Debate about Animal Language*. Chicago.

Ristau, C. (1983). "Language, Cognition, and Awareness in Animals?", *Annals of the New York Academy of Sciences* 406(1): 170–186.

Romanes, G. (1888). *Mental Evolution in Man*. Kegan Paul, Trench & Co.

Rosati, A., Machanda, Z, & Slocombe, K. (2022). "Cognition in the Wild: Understanding Animal Thought in its Natural Context", *Current Opinion in Behavioral Sciences*. Online First.

Rosset, E. (2008). "It's no accident: Our bias for intentional explanations" *Cognition* 108, 771–780.

Ryle, G. (1949). *The Concept of Mind*. Chicago University Press.

Saidel, E. (2009). "Animal Thoughts" in Robert Lurz (ed.) *Animal Minds*. CUP.

Sánchez-Amaro, A., Tan, J., Kaufhold, S. P., & Rossano, F. (2020). "Gibbons exploit information about what a competitor can see", *Animal Cognition*, 23(2), 289–299.

Schwitzgebel, E. (2001). "In-Between Believing", *Philosophical Quarterly* 51, 76–82.

Schwitzgebel, E. (2002). "A Phenomenal, Dispositional Account of Belief", *Nous*, 36, 249–275.

Schwitzgebel, E. (2013). "A dispositional approach to the attitudes: thinking outside the belief box." In *New Essays on Belief*, ed. Nikolaj Nottelmann (Palgrave Macmillan), 75–99.

Schwitzgebel, E. (2021). "The Pragmatic Metaphysics of Belief" in D. Kindermann and C. Borgoni, *The Fragmented Mind* (Oxford: 2021), 350–375.

Schwitzgebel, E. (2022). "The Nature of Belief from a Philosophical Perspective, with Theoretical and Methodological Implications for Psychology and Cognitive Science", *Frontiers in Psychology* 13, 1–4.

Shettleworth, S. (2007). *Cognition, Evolution and Behavior*. OUP.

Shettleworth, S. (2010). "Clever animals and killjoy explanations in comparative psychology" *Trends in Cognitive Sciences* 14:11, 477–481.

Skinner, B.F. (1938). *The Behavior of Organisms*. Appleton & Company.

Sober, E. (1998). "Morgan's Canon." In C. Allen and D. Cummins (eds.), *The Evolution of Mind*, OUP, 224–242.

Southgate, V., Senju, A., & Csibra, G. (2007). "Action anticipation through attribution of false belief by 2–year–olds", *Psychological Science*, 18:7, 587–592.

Spaulding, S. (2018). *How We Understand Others: Philosophy and Social Cognition*. Routledge.

Spence, C., Osman, M., & McElligott, A. (2017). "Theory of Animal Mind: Human Nature or Experimental Artefact?", *Trends in Cognitive Sciences* 21:5, 333–343.

Sterelny, K. (1990). *The Representational Theory of Mind*. Blackwell.

Stich, S. (1979). "Do animals have beliefs?" *Australasian Journal of Philosophy* 57:15–28.

Stich, S. (1983). *From Folk Psychology to Cognitive Science*. MIT Press.

Tanney, J. (2013). *Rules, Reason, and Self–Knowledge*. HUP.

Templer, V. (2022). "Metacognition" In B.L. Schwartz & M.J. Beran (Eds.) *Primate Cognitive Studies*. CUP, 314–340.

Thomas, R. (1998). "Lloyd Morgan's Canon". In G. Greenberg and M. Haraway (eds), *Comparative Psychology: A handbook*. Garland.

Thomas, R. (2001). "Lloyd Morgan's Canon: A History of Misrepresentation," *History & Theory of Psychology Eprint Archive*.

Van Leeuwen, N. & Lombrozo, T. (2023). "The Puzzle of Belief," *Cognitive Science* 47, 13245.

Wynne, C. (2007). "What are Animals? Why Anthropomorphism is Still not a Scientific Approach to Behavior", *Comparative Cognition and Reviews*, 2: 125–135.