



Zoomorphism

Bence Nanay^{1,2}

Received: 20 December 2017 / Accepted: 8 December 2018
© The Author(s) 2018

Abstract

Anthropomorphism is the methodology of attributing human-like mental states to animals. Zoomorphism is the converse of this: it is the attribution of animal-like mental states to humans. Zoomorphism proceeds by first understanding what kind of mental states animals have and then attributing these mental states to humans. Zoomorphism has been widely used as scientific methodology especially in cognitive neuroscience. But it has not been taken seriously as a philosophical explanatory paradigm: as a way of explaining the building blocks of the human mind. The philosophical explanatory paradigm of zoomorphism may not explain all aspects of human behavior (although it may explain surprisingly many), but if we accept the zoomorphic way of thinking about the human mind, we should only posit new, different kinds of mental states if the zoomorphic attribution of animal mental states fails to explain our behavior.

1 Anthropomorphism and Zoomorphism

Anthropomorphism is the methodology of attributing human-like mental states to animals: when it comes to explaining the minds of non-human animals, we can and should make reference to mental states that we are familiar with from our understanding of the human mind. The merits and demerits of anthropomorphism have been heavily contested in recent times, but I want to bracket these debates for the purposes of this paper.

My aim here is to argue for the converse methodology and explanatory scheme: zoomorphism. Zoomorphism is the methodology of attributing mental states we know from the study of nonhuman animals to humans. My claim is that zoomorphism is very successful as scientific methodology and it should be also taken

✉ Bence Nanay
bence.nanay@uantwerpen.be
<https://uahost.uantwerpen.be/bence.nanay>

¹ Centre for Philosophical Psychology, University of Antwerp, D 413, Grote Kauwenberg 18, 2000 Antwerp, Belgium

² Peterhouse, University of Cambridge, Cambridge CB2 1RD, UK

seriously as philosophical explanatory paradigm: it should guide the way we think about the human mind.

The plan of the paper is the following. In Sect. 2, I distinguish zoomorphism as scientific methodology and zoomorphism as philosophical explanatory paradigm and point out that zoomorphism has long been a thriving and successful piece of scientific methodology (Sect. 3). This should make us take zoomorphism as philosophical explanatory paradigm seriously. I argue that, understood properly, zoomorphism is not a radical explanatory paradigm at all (Sect. 4) and it is preferable to other philosophical explanatory paradigms about the human mind (Sect. 5). Finally, in Sect. 6, I consider what mental states we can expect to attribute to humans if we accept zoomorphism as philosophical explanatory paradigm.

2 Scientific Methodology Versus Philosophical Explanatory Paradigm

I want to make a distinction between zoomorphism as scientific methodology and zoomorphism as philosophical explanatory paradigm. A similar distinction can be made about anthropomorphism as well, so I will introduce this distinction in this more familiar context. Anthropomorphism, the attribution of human-like mental states to animals can be used as a piece of scientific methodology: when explaining some specific behavior of a certain kind of animals, we should attribute certain mental states to them that we know from human psychology.

But anthropomorphism is not merely a piece of scientific methodology, it is also a philosophical explanatory paradigm, according to which we arrive at the correct way of understanding animal minds if we start by postulating that these animals have human-like mental states.

Anthropomorphism as scientific methodology aims to describe the specific mental states of a certain species. Anthropomorphism as philosophical explanatory paradigm, in contrast, is a very general theoretical assumption about how animal minds should and could be explained.

Here is an example: We know (somehow) that we humans have beliefs. So it is permissible to describe the behavior of animals in terms of beliefs. This is anthropomorphism as philosophical explanatory paradigm. Now, just what beliefs one attributes to an animal is a task for anthropomorphism as scientific methodology.

One can take anthropomorphism to be a useful piece of scientific methodology in some contexts without endorsing anthropomorphism as philosophical explanatory paradigm. And one can endorse anthropomorphism as philosophical explanatory paradigm without having any interest whatsoever in what kinds of mental states one needs to postulate to explain specific animal behavior. But anthropomorphism as philosophical explanatory paradigm is the theoretical foundation and regulative ideal of anthropomorphism as scientific methodology.

Debates about various versions of anthropomorphism ('new anthropomorphism', 'naïve vs. critical anthropomorphism', 'animal-centric vs. anthropocentric anthropomorphism, and so on, see de Waal 1991, 1999, 2009; Burghardt 1985; see also Griffin 1984, 1992; Kennedy 1992; Fisher 1996; Allen and Bekoff 1997; Keeley 2004;

Sober 2012; Wynne 2004a, b; Kennedy 1992; Guthrie 1997; Blumberg and Wasserman 1995; Penn 2011) are not always clear about which sense of anthropomorphism is being discussed: scientific methodology or philosophical explanatory paradigm. Some progress might be made in the anthropomorphism literature if we paid attention to this distinction, but that is not the aim of this paper. It's time to return to zoomorphism.

A similar distinction could be made between zoomorphism as scientific methodology and zoomorphism as philosophical explanatory paradigm. Zoomorphism as philosophical explanatory paradigm is a general theoretical assumption about how we can explain the human mind and its building blocks. It asserts that we should only posit mental state types that we know from the study of animal cognition.

Zoomorphism as scientific methodology, in contrast, is a tool for understanding what specific mental states we, humans, have, on the basis of postulating animal-like mental states. As before, one can use zoomorphism as scientific methodology (maybe because a certain experiment would be too intrusive to undertake in humans, but not in animals) without endorsing zoomorphism as philosophical explanatory paradigm. And one can endorse zoomorphism as philosophical explanatory paradigm without doing so in order to be able to use zoomorphic scientific methodology. Nonetheless, zoomorphism as philosophical explanatory paradigm is the theoretical foundation and regulative ideal of zoomorphism as scientific methodology.

In the next section, I argue that as scientific methodology, zoomorphism has been widely used and it has been immensely successfully. The focus of the rest of the paper is zoomorphism as philosophical explanatory paradigm.

3 Zoomorphism as Scientific Methodology

The aim of this section is to show how zoomorphism as scientific methodology works. The first thing we should notice is that the general explanatory scheme zoomorphism uses, namely, to import some considerations about animals to explain humans has been very widespread, starting from anatomical studies of animals in order to understand human anatomy (as early as the writings of Galen of Pergamon and even Aristotle). But the focus of this paper is zoomorphism as a way of understanding the human *mind*. And using considerations about the mind of nonhuman animals when explaining the human mind is an equally widespread and equally successful research methodology, at least in some branches of cognitive science, especially cognitive neuroscience.

An important example of how the scientific methodology of zoomorphism can be used to illuminate how the human mind works is the literature on mirror neurons. Mirror neurons are not mental states. But the zoomorphic attribution of this anatomical feature (that we know from animals) to humans nonetheless can help us to understand the human mind better. The mirror neuron system (or, rather, systems) consists of bi-modal neurons that get activated both when the agent performs an action and when she perceives another agent performing this action (di Pellegrino et al. 1992; Gallese et al. 2004; Rizzolatti and Sinigaglia 2008; Sinigaglia 2009). The mirror neurons do not get activated when the perceived agent does not perform

a goal-directed action but exhibits a mere (not goal-directed) bodily movement (Takei et al. 2001; Umiltà et al. 2008). If the other agent is grasping a ball, the mirror neurons fire, if she is making a grasping movement without there being anything to be grasped, they do not.

The first findings about mirror neurons were about the brain of rhesus monkeys and to date the only single cell recordings are from rhesus monkeys (as the procedure would be too intrusive in humans).¹ So we have found a mental apparatus in a nonhuman animal—the mirror neuron system. The question is: should we now describe humans as having mirror neuron systems? Can we make the step of zoomorphism of attributing a mental apparatus we know from the study of animals to humans?

And here the literature on mirror neurons can give us some guidance about how the methodology of zoomorphism can and should be used. Importantly, it does not follow from the methodology of zoomorphism that we should conclude on the basis of having found a mechanism in nonhuman animals that humans also have this very mechanism—it is a dubious inference that it follows from the single cell recordings in the rhesus monkey brain that the human brain works the same way. But the existence of the mirror neuron system in rhesus monkeys provides a good starting hypothesis for the study of the bimodal neurons in the human brain. Evidence other than single cell recordings (behavioral, TMS, fMRI) needs to be used to verify whether and in what respects the human mirror neuron system resembles the mirror neuron system of rhesus monkeys (Gallese et al. 1996; Fabbri-Destro and Rizzolatti 2008; Keysers and Gazzola 2010; Lingnau et al. 2009; Umiltà et al. 2008; see Rizzolatti and Sinigaglia 2008 for a summary). In this case, the attribution of the animal-like mechanism is not the end of the story, but rather the first step to be followed by a close examination of how this mechanism works differently in the human and in the animal mind. In short, some very successful parts of cognitive neuroscience actively and successfully use some rudimentary form of zoomorphism as scientific methodology.

Mirror neurons are not mental states. So the attribution of mirror neuron system (that we have discovered in rhesus monkeys) to humans is not the attribution of an animal-like mental state to humans, because it is not the attribution of any mental state. However, suppose that there are mental states that are implemented by mirror neurons—a very widespread assumption. Whatever these mental states may be (see Gallese et al. 2004; Gallese and Goldman 1998; Rizzolatti and Sinigaglia 2008; Jacob and Jeannerod 2003; Jacob 2008, 2009; Spaulding 2013; Csibra 2007 for various candidates) I would like to leave open. If there are indeed such mental states, then the attribution of mental states of this kind to humans would be a genuine example of zoomorphic scientific methodology.

More generally, attributing a *mental mechanism* we know from non-human animals to humans is very different from attributing a *mental state-type* we know from non-human animals to humans. The former is a special case of the old, well-respected and not at all controversial way of explaining humans—we have seen

¹ But see Mukamel et al. (2010) for some single cell recordings in humans with brain damage.

that its pedigree goes back to Galen and Aristotle. But the zoomorphic research on social cognition in comparative psychology, cognitive ethology and developmental psychology is doing the latter, not the former (see Gallese et al. 2004; Gallese and Goldman 1998; Rizzolatti and Sinigaglia 2008; Jacob and Jeannerod 2003; Jacob 2008, 2009; Spaulding 2013; Csibra 2007). They use the mirror neuron findings to attribute a certain mental state-type (not merely a mental mechanism) to humans.

Moving away from the mirror neuron example, there is no shortage of genuinely zoomorphic ways of attributing mental states to humans. One strikingly clear example of zoomorphism as scientific methodology comes from Edward Tolman. Tolman used his postulation of mental maps to rats on the basis of their behavior to speculate about the presence and nature of mental maps in humans (Tolman 1948). When Tolman attributed mental maps to humans on the basis of findings in rats, he did not attribute a mental mechanism, but rather a mental state type.

Another illuminating case study for the success of zoomorphism as scientific methodology is the literature on decision-making under risk, where considerations about the way animals make decisions under risk (for example, how affective states influence decision-making under risk) feeds into models about human decision-making under risk (see, e.g., Marsh and Kacelnik 2002; McCoy and Platt 2005; Lakshminarayanan et al. 2011). Again, the literature on decision-making under risk is not describing a mechanism, but rather the contentful mental state types that play a role in human decision-making (and they do so partly on the basis of results in decision-making in animals).

And these examples are not isolated (Mitchell 2000, 2002; Waller 2012). The field of cognitive psychology as a whole is moving towards the direction of zoomorphism by stressing the importance of designing experiments in a way that remains functionally equivalent across humans and nonhumans (see Shettleworth 2010 for a summary and Tomasello and Call 2008 for a philosophically sophisticated case study and Timberlake 2007).

The success of zoomorphism as scientific methodology should make us take zoomorphism as philosophical explanatory paradigm seriously. The rest of the paper is about zoomorphism as philosophical explanatory paradigm. For simplicity, I will use the term 'zoomorphism' in this sense from now on.

4 Zoomorphism as a Philosophical Explanatory Paradigm

Zoomorphism (as philosophical explanatory paradigm) is a very general theoretical answer to the question about how we should explain the building blocks of the human mind. How do we know what mental state types humans have? Zoomorphism answers this question by pointing at animal minds. We know some mental state types from the study of animal cognition. We should try to explain human mind in terms of these mental state types. This may not explain all human behavior, but this is where we need to start, only postulating new mental state types when explanations in terms of fully animal mental state types fail.

It is important to emphasize that zoomorphism is not a theory, but a philosophical explanatory paradigm: it makes no claims about just what these mental

states that we should attribute to humans (because we know nonhuman animals have them) in fact are. I will say more about what these mental states may be, given the present state of cognitive science and cognitive neuroscience of the animal mind and brain in Sect. 6. But first a couple of clarifications need to be made about what zoomorphism does and doesn't entail. Zoomorphism may seem to be a very radical explanatory paradigm (especially compared to the more established alternatives—see Sect. 5 below). I hope to show that while it is easy to misinterpret zoomorphism in a way that would indeed have radical consequences, zoomorphism, properly understood, is not a very radical view at all.

4.1 Zoomorphism is Not the Full Story

It is very important to emphasize that zoomorphism is not supposed to give us a full description of the human mind: it is an explanatory paradigm for finding out about the human mind but there is no reason to believe that it is the ultimate or the only tool we have for understanding the human mind.

A comparison with anthropomorphism may be useful to elucidate this point. Some versions of anthropomorphism—especially ‘critical anthropomorphism’ (Burghardt 1985) or ‘animal-centered anthropomorphism’ (de Waal 1999)—emphasize that anthropomorphism is not supposed to do all the work when it comes to understanding the mind of nonhuman animals. The idea, rather, is that the anthropomorphist step of attributing human mental states to nonhuman animals is followed by an animal-specific second step where it is established how these human-like mental states of nonhuman animals are different from our own mental states and also whether we need to postulate some extra mental states or processes, ones that we can't find in the human mind, in order to explain the behavior of nonhuman animals (de Waal 1991, 1999, 2009; Burghardt 1985; see also Griffin 1984, 1992; Kennedy 1992; Fisher 1996; Allen and Bekoff 1997).

Similarly, zoomorphism is not supposed to do all the work when it comes to understanding the human mind either. All that follows from the explanatory paradigm of zoomorphism is that the first step of understanding the human mind is to attribute mental states to humans that we know from studying nonhuman animals. It would be really odd if this gave us a full story about the human mind as the human mind evolved significantly since our latest common ancestors with nonhuman animals. Thus, the second step would need to be to understand how this conceptual apparatus, which comes from the study of nonhuman animals, needs to be refined and altered in order to explain human behavior.

And this may not be the end of the story either: we may need to postulate some further mental states, ones that we don't find in nonhuman animals, in order to give a full account of the human mind. But the important commitment of zoomorphism is that this process of explaining the human mind needs to begin with the description of the human mind in terms that we are familiar with from the study of the mind of nonhuman animals.

4.2 Zoomorphism and Behaviorism

Another important thing to clarify about zoomorphism is that it does not entail behaviorism—another old and influential philosophical explanatory paradigm. Zoomorphism is very much in the business of attributing mental states both to nonhuman animals and to humans—in fact, the main claim is that we should do the latter on the basis of the former. Behaviorism denies that any such attribution of mental states would be justified.

Behaviorism could be considered to be an extreme limiting case of zoomorphism: the idea would be that as we should not attribute any mental state to animals, we should not attribute any mental state to humans either (this would be, for example the view endorsed in Watson 1930).

The comparison between zoomorphism and behaviorism is important inasmuch as it highlights how much the explanatory scheme of zoomorphism is determined by what we take to be those mental states that we are justified to attribute to nonhuman animals. If you are behaviorist, the answer to this would be no mental states. But the aim of cognitive ethology and comparative psychology (see Allen 2004) is to provide plausible candidates for mental states that we are justified to attribute to nonhuman animals. And then we can use these considerations to justify the attribution of these very mental states to humans.

4.3 Zoomorphism and Animal Minds

I argued for a philosophical explanatory paradigm of attributing mental states of animals to humans. But this obviously raises the question how we know what mental states animals have. So one could worry that all zoomorphism achieved was to shift the problem from humans to non-human animals. If we want to substantiate zoomorphism as a philosophical explanatory paradigm, we need to ask what mental states we should attribute to animals (which, in turn, we also attribute to humans).

I will come back to the problem of how zoomorphism as a philosophical explanatory paradigm about the human mind relies on some other philosophical explanatory paradigm about the animal mind in Sect. 5.2. But there is another issue that needs to be addressed. As I formulated it, the philosophical explanatory paradigm of zoomorphism tells us to attribute mental states of animals to humans. But of what kinds of animals? There is huge variation in animal cognition from jellyfish to bonobo—what animal mental states should we use in our zoomorphic explanations? While for pragmatic reasons it is more likely that we can use mental states we know from the study of bonobo cognition than ones we know from the study of jellyfish, this should not be taken to be an endorsement of some kind of homology or ‘closest evolutionary ancestors’ argument (see Sober 2001, 2005, 2012, for example).

It will be more likely that we find behavioral patterns that are similar to ours in bonobos than in jellyfish. On the other hand, we can do more intrusive experiments on animals that are further away from us in terms of homology. We have a form of trade-off between using findings from animals close to us in homology and animals easier to experiment on. How this trade-off is resolved will vary from case to case

and it is really a question not about zoomorphism as a philosophical explanatory paradigm, but about zoomorphism as scientific methodology.

Zoomorphism as a philosophical explanatory paradigm can be implemented as scientific methodology in a variety of ways. Depending on how we do so, we may get very different answers to the question about which animals we should use as the explanatory basis of human mental states. Nonetheless, an important part of zoomorphism as a philosophical explanatory paradigm (and not as scientific methodology) that as a general rule, it should not be assumed that closeness in terms of homology always wins out (for example, nothing would rule out that the reason for the similarity of behavior is not homology but analogy).

In this section, I was hoping to show that zoomorphism is not a particularly radical explanatory paradigm. Crucially, it does not have any claims to exclusivity when it comes to understanding the human mind: we should use zoomorphism as the first step of understanding the human mind, but a lot needs to be done after this first step is completed.

5 Zoomorphism Versus Other Philosophical Explanatory Paradigms About the Human Mind

Zoomorphism is not the only candidate for a coherent answer to the question about how we should explain the building blocks of the human mind. I consider what I take to be the two main alternatives, to show what kind of explanatory project zoomorphism is supposed to be (a third one is behaviorism, see Sect. 4.2). I argue that zoomorphism as a philosophical explanatory paradigm is superior to these alternatives.

5.1 The First Contrast: Introspection as Philosophical Explanatory Paradigm

I suspect that some readers might find the question this paper raised pointless. What do you mean what mental state types we should attribute to humans? *We are* humans. We should just attribute those mental state types we know we have. Finding these is easy: just turn your attention inwards to check what mental states you feel you are in. And the answer will presumably be some combination of beliefs and desires.

This is a philosophical explanatory paradigm and an old and venerable one. Philosophers have long been assuming that the building blocks of the human mind are beliefs and desires and the reason they give for this is some form of introspective evidence. So the explanatory paradigm would be this: let's start with beliefs and desires and try to explain all human behavior in these terms. This may not explain all human behavior, in which case, we may tweak our theory, maybe postulating some extra assumptions.

Introspection is an old and influential philosophical explanatory paradigm, but it is also a very problematic one. My aim in this section is not to show that introspection as a philosophical explanatory paradigm is hopeless or flawed, but merely that

zoomorphism is preferable to it. And the reason for this is the rich philosophical and psychological literature on the unreliability of introspection (see Nisbett and Wilson 1977 for the *locus classicus* and Schwitzgebel 2008; Spener and Bayne 2010; Spener 2011 for philosophical summaries). This is a vast literature and I can't do justice to all the wrinkles here. But it has been argued that introspection is an unreliable guide to what goes on in our mind. Often, the reasons why we do one thing rather than another are not at all transparent to us.

Here is one of the most famous experiments for the unreliability of introspection: subjects had to choose between four consumer products that were in fact identical. They tended to choose the product that was presented as the first one from the right. But when asked why they chose that particular product, they did not mention its position (and strongly denied that they were influenced by its position when explicitly asked). Rather, they explained their choices in terms of the quality of the product (Nisbett and Wilson 1977). There is a debate in the recent literature about whether this experiment shows that the biases on this choice were unconscious (see Newell and Shanks 2014 for a summary) and we do not need to take sides in this debate. But there is no debate about the existence of the discrepancy between the actual reasons for choosing the product on the right and the reasons the subjects introspectively discerned. People confabulate about what they take to be their reasons on the basis of introspection.

The Nisbett and Wilson experiment is by no means an isolated example. It turns out that our decision-making is subject to a variety of biases: order effect, framing effect, availability bias and more: on whether we have just watched a funny film, on what beverage we are drinking, on whether we hold a teddy bear or on whether we are surrounded by dirty pizza boxes (see Kahneman and Tversky 1973; Tversky and Kahneman 1981; Valdesolo and DeSteno 2006; Schnall et al. 2008; Zhong and Liljenquist 2006; Williams and Bargh 2008; Tai et al. 2011 for just a taste of the vast literature). Whether these biases are in fact unconscious is less important for our purposes; what is crucial, however, is when subjects whose decision is biased in these ways give all kinds of false reasons for their decision-making on the basis of their introspection.

A different set of evidence about the unreliability of introspection comes from action-initiation studies: it turns out that the mental state that initiates our action precedes significantly the mental state we introspectively take to initiate our actions (Wegner 2002; Haggard et al. 2002; Wegner et al. 2004). Some have tried to draw conclusions from these findings for the free will debate, and it is important to emphasize that this is not what I am trying to do here. But regardless of what follows from these findings when it comes to the debate about free will and determinism, it is uncontroversial that our introspection misidentifies the mental state that is responsible for initiating actions.

One might worry that in all these experiments introspection is unreliable with regards to the content of a mental state, but they are not unreliable with regards to the mental states themselves. But there are experiments that show that introspection can fail to identify what kind of mental state we are in. In the Perky experiment, for example, subjects are looking at a white wall and they are asked to visualize objects while keeping their eyes open. Unbeknownst to them, barely visible images

of the visualized objects are projected on the wall. The surprising finding is that the subjects take themselves to be visualizing the objects—while in fact they perceive them (Perky 1910; Segal 1972; Segal and Nathan 1964). The standard interpretation of this experiment is that if perceiving and visualizing could be confused under these circumstances, then they must be phenomenally very similar (but see Hopkins 2012's criticism and Nanay 2012's response, see also Craver-Lemley and Reeves 1992; Reeves and Craver-Lemley 2012).

These sets of examples for the unreliability of introspection are by no means isolated cases (see also Carruthers 2011 for a summary). Some further important examples include our implicit bias (Greenwald and Banaji 1995; Dunham et al. 2008) as well as unconscious emotions, actions, learning, attention and perception (e.g., Winkielman and Berridge 2004; Jiang et al. 2006). Do these findings, put together establish that introspection is *never* a reliable guide to what goes on in our mind? Probably not. But this is not the claim I want to defend either. Given the diverse evidence for the unreliability of introspection regarding very different mental processes, if we want to use introspection as the guiding philosophical explanatory paradigm for attributing mental states to humans, we would need some support for the claim that the introspection we use when trying to find out what mental state types we have is reliable (unlike other uses of introspection). Similar worries do not apply in the case of zoomorphism (as a philosophical explanatory paradigm). So other things being equal, zoomorphism is preferable to introspection-based methodology.

5.2 The Second Contrast: Behavioral Evidence as Philosophical Explanatory Paradigm

The second example of philosophical explanatory paradigms I want to consider is a very efficient and widespread way of explaining the building blocks of the human mind. The general explanatory scheme is that we should only posit a mental state type on the basis of behavioral evidence. Suppose that some complex behavior in humans can only be explained if we posit a mental state type *M*. In this case, we have a strong reason to attribute *M* to humans. A weaker version of this methodology is that we have reason to attribute *M* to humans if attributing *M* explains our behavior (much) better than not attributing *M*. So for the purposes of explaining our behavior, attributing *M* may not be the absolutely only option, but not attributing *M* would lead to serious loss of explanatory power.

One potential example comes from Marc Jeannerod's work in the cognitive neuroscience of action (see, e.g., Jeannerod 1997): he argues that various behavioral features of the fine-grained details of our action execution can only be explained if we posit a mental state type he calls 'motor representations'. Another example is the 'theory of mind' literature, where, arguably, the very same explanatory paradigm is used (see Tomasello and Call 2008 for a summary).

This philosophical explanatory paradigm does not rely on introspection and, as a result, is not susceptible to the objection I raised above. But there is another problem. While this is clearly a very important and widely used explanatory scheme, its scope of application is limited. In order to infer that humans must have a certain

specific mental state type, what is required is that the only way we can explain the complexity of our behavior is by positing this mental state type (or that explaining our behavior by positing this mental state type is vastly and obviously superior to *any* alternatives). And this requirement is very rarely (if at all) met.

The reason for this is an analogue of the underdetermination of theory by data argument: there are a vast amount of potential mental state types we might attribute to humans on the basis of behavioral evidence. And the behavioral evidence would need to be very diverse and very multifaceted in order to rule out all but one possible mental state type.

An analogy would be the debate about unobservable entities in the philosophy of science. Mental states are also unobservable: we posit them on the basis of observable phenomena—on the basis of behavior. But we know from the philosophy of science, and especially from philosophy of physics, that the observable phenomena very rarely determine in full specificity the unobservable entity.

Let's go back to Jeannerod's motor representations. He argued that some fine details of action execution cannot be explained without positing a mental state no one had talked about before: a kind of representation that represents the features of the objects the action is performed on in a way that would help the agent perform the action successfully.

But this concept leaves a lot of details unclear—and different philosophers and psychologists give very different interpretation of the same idea of motor representations. For some, motor representations represent the goal of the action (Butterfill and Sinigaglia 2014). For others, they represent the properties of objects in an egocentric manner (Nanay 2013). Another option would be to take motor representations to both represent and prescribe actions (Pacherie 2000). And yet another option would be to take them to represent our own bodily movements (Bach 1978).

Arguably, all of these accounts of what motor representations are would be compatible with the behavioral evidence that made Jeannerod posit motor representations to begin with. The philosophical explanatory paradigm of only positing mental state types on the basis of behavioral evidence is a very prudent and scientifically honest explanatory scheme, but it will only get us so far.

As we have seen, we could weaken this explanatory paradigm so that it only demands that by positing a mental state type, we can explain behavior (much) better. But this raises thorny questions about what counts as (much) better explanation in this context, and these questions are notoriously difficult to tackle. If positing mental state type M can explain a certain range of behavior and positing mental state type N explains a slightly smaller range of behavior but can do so without positing less theoretical generalizations, which one should we prefer? One could argue that lack of clarity about what exactly makes an explanation better than another would be in itself an indication that this version of the behavioral methodology is vague at best.

At this point one might wonder how my zoomorphic account is better off in this respect. After all, the general strategy of zoomorphism is to attribute mental states to humans that we are familiar with from the study of animal mind. But how did we get to attribute this mental state type to animals to begin with? The methodology for that is presumably a form of the behavioral methodology. In this sense, so the

argument would go, zoomorphism relies on the behavioral philosophical explanatory paradigm.

And it indeed does so. But using the behavioral philosophical explanatory paradigm in the case of the human mind is very different from doing so in the case of the animal mind. One important reason why we are in a better position is that we have a more diverse set of observable data when it comes to nonhuman animals—for the simple reasons that we can do nasty experiments on animals we are not allowed to do on humans. As we have seen, one illustration of this point is the research on mirror neurons. The single-cell recordings that provided the most important data-point about the mirror-neuron system could only be done on animals for ethics clearance reasons.

Does this get rid of all the problems plaguing the behavioral explanatory paradigm? Not at all. But it makes it easier to narrow down the kinds of mental states that we could or should attribute to animals. And then we can use the methodology of zoomorphism to attribute these mental states to humans.

6 Zoomorphism and the Human Mind

In this last section, I want to consider, somewhat tentatively, what kind of mental states we are justified to attribute to humans on the basis of the studies about the animal mind and whether this should lead us to reconsider the way we think about our own mind.

Probably the most salient consequence of the explanatory paradigm of zoomorphism is that it underplays the importance of conscious mental processes. If the general explanatory commitment is to only posit a kind of mental state in humans as a last resort if we can't explain human behavior (which of course includes linguistic behavior) with the help of mental states we have evidence for from animal cognition, then at first approximation the human mind should be taken to consist of mental processes that are not necessarily conscious (see Andrews 2015, esp. Chapter 3 for a summary). Whatever mental states cognitive ethology, comparative psychology and cognitive neuroscience is entitled to attribute to nonhuman animals, they are mental states that are not individuated by virtue of their phenomenal character, but rather by their functional role. In other words, mental states that are not necessarily conscious. So it follows from the explanatory paradigm of zoomorphism that as a first step of understanding the human mind, we should attempt to do so in terms of mental processes that are not necessarily conscious.

This is, of course, unlikely to be the end of the story. Unless we want to deny the existence of consciousness, at some point in the explanation of the human mind we need to bring in consciousness. But we shouldn't do so in the first, zoomorphic, step of explaining the human mind.

This way of thinking about the relation between understanding mental processes and understanding consciousness used to be quite mainstream within philosophy: the idea in the 1980s was to first try to understand intentionality without any reference to consciousness and once we have done that we can then go on to try to

understand consciousness (Millikan 2004; Papineau 1987)—this project, at least with respect to consciousness, was a zoomorphist one.

But things have changed. Given the tremendous amount of work done on consciousness since the 1990s, not many people would proceed this way any more when trying to understand the human mind. But maybe the recent work on various unconscious mental processes (unconscious perception, unconscious attention, unconscious action, unconscious emotion, unconscious biases, unconscious learning, implicit bias, see some references in Sect. 5.1) should persuade us to first try to understand these mental processes (perception, attention, emotion, learning, etc.) without talking about consciousness and worry about consciousness later. And this is exactly what the philosophical explanatory paradigm of zoomorphism suggests.

The second important insight about how we should think about our own mind if we take zoomorphism seriously comes from the cognitive neuroscience of action. The big question in the cognitive neuroscience of action is what mental states are needed in order for an agent to perform an action successfully. And the answer is not in terms of beliefs, desires and other folk-psychological categories.

One mental state that seems very important for the successful performance of actions is what I will label ‘motor representation’ here: representation that represents all the parameters of the situations that are necessary for the successful performance of action. Just what these parameters are, as we have seen, is hotly debated: they may include the properties of the objects one acts upon, the properties of one’s own body, one’s bodily movement that is needed to complete the action or maybe the properties of the goal state the action is aimed at (see Jeannerod 1997; Nanay 2013; Poincaré 1905/1958; Bach 1978; Brand 1984; Pacherie 2011; Millikan 2004; Butterfill and Sinigaglia 2014 for very different proposals about this). I want to remain neutral about the content of these motor representations for the purposes of this paper.

Animals perform actions: they run away from predators and approach prey, for example. And they do so successfully most of the time. Thus, we have very strong reason to attribute motor representations to them. But then if we apply the explanatory scheme of zoomorphism, the first step of understanding the human behavior would be to explain it in terms of our motor representations.

This approach, of course, has its limits. While many of our fine-grained bodily movements may be explained by virtue of attributing motor representations, some aspects of human cognition (especially linguistic or rational reasoning) clearly can’t be. But if we want to follow the explanatory paradigm of zoomorphism, then we first need to see how far we can go in our understanding of the human mind by using motor representations and we can only use different mental states when the human behavior is clearly not possible to fully describe only with the help of motor representations.

Acknowledgements This work was supported by the ERC Consolidator grant 726251, the FWO Odysseus grant G.0020.12N and the FWO research grant G0C7416N. I gave a talk on this topic at the Hungarian Academy of Sciences in 2014 and I am grateful for the feedback received.

Open Access This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution,

and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

References

- Allen, C. (2004). Is anyone a cognitive ethologist? *Biology and Philosophy*, *19*, 589–607.
- Allen, C., & Bekoff, M. (1997). *Species of mind: The philosophy and biology of cognitive ethology*. Cambridge: MIT Press.
- Andrews, K. (2015). *The Animal Mind: An Introduction to the Philosophy of Animal Cognition*. New York: Routledge.
- Bach, K. (1978). A representational theory of action. *Philosophical Studies*, *34*, 361–379.
- Blumberg, M. S., & Wasserman, E. A. (1995). Animal mind and the argument from design. *American Psychologist*, *50*, 133–144.
- Brand, M. (1984). *Intending and acting*. Cambridge, MA: The MIT Press.
- Burghardt, G. M. (1985). Animal awareness: Current perceptions and historical perspective. *American Psychologist*, *40*, 905–919.
- Butterfill, S., & Sinigaglia, C. (2014). Intention and motor representation in purposive action. *Philosophy and Phenomenological Research*, *88*, 119–145.
- Carruthers, P. (2011). *The opacity of mind: An integrative theory of self-knowledge*. Oxford: Oxford University Press.
- Craver-Lemley, C., & Reeves, A. (1992). How visual imagery interferes with vision. *Psychological Review*, *99*, 633–649.
- Csibra, G. (2007). Action mirroring and action interpretation: An alternative account. In P. Haggard, Y. Rosetti, & M. Kawato (Eds.), *Sensorimotor Foundations of Higher Cognition. Attention and Performance XXII* (pp. 435–459). Oxford: Oxford University Press.
- De Waal, F. B. M. (1991). Complementary methods and convergent evidence in the study of primate social cognition. *Behaviour*, *118*, 297–320.
- De Waal, F. B. M. (1999). Anthropomorphism and anthropodenial: Consistency in our thinking about humans and other animals. *Philosophical Topics*, *27*, 255–280.
- De Waal, F. B. M. (2009). Darwin's last laugh. *Nature*, *460*, 175.
- di Pellegrino, G., Fadiga, L., Fogassi, L., Gallese, V., & Rizzolatti, G. (1992). Understanding motor events: A neurophysiological study. *Experimental Brain Research*, *91*, 176–180.
- Dunham, Y., et al. (2008). The development of implicit intergroup cognition. *Trends in Cognitive Science*, *12*, 248–253.
- Fabbri-Destro, M., & Rizzolatti, G. (2008). Mirror neurons and mirror systems in monkeys and humans. *Physiology*, *23*(3), 171–179.
- Fisher, J. A. (1996). The myth of anthropomorphism. In M. Bekoff & D. Jamieson (Eds.), *Readings in animal cognition* (pp. 3–16). Cambridge, MA: MIT Press.
- Gallese, V., & Goldman, A. (1998). Mirror neurons and the simulation theory of mind-reading. *Trends in Cognitive Science*, *3*, 493–501.
- Gallese, V., Fadiga, L., Fogassi, L., & Rizzolatti, G. (1996). Action recognition in the premotor cortex. *Brain*, *119*, 593–609.
- Gallese, V., Keysers, C., & Rizzolatti, G. (2004). A unifying view of the basis of social cognition. *Trends in Cognitive Sciences*, *8*, 396–403.
- Greenwald, A. G., & Banaji, M. R. (1995). Implicit social cognition. *Psychological Review*, *102*, 4–27.
- Griffin, D. R. (1984). *Animal thinking*. Cambridge, MA: Harvard University Press.
- Griffin, D. R. (1992). *Animal minds*. Chicago: The University of Chicago Press.
- Guthrie, S. E. (1997). Anthropomorphism: A definition and a theory. In R. Mitchell, N. S. Thompson, & H. L. Miles (Eds.), *Anthropomorphism, anecdotes, and animals* (pp. 50–58). Albany: SUNY Press.
- Haggard, P., Clark, S., & Kalogeras, J. (2002). Voluntary action and conscious awareness. *Nature Neuroscience*, *5*(4), 382–385.
- Hopkins, R. (2012). What Perky did not show. *Analysis*, *72*, 431–439.
- Jacob, P. (2008). What do mirror neurons contribute to human social cognition? *Mind and Language*, *23*, 190–223.

- Jacob, P. (2009). The tuning-fork model of human social cognition: A critique. *Consciousness and Cognition*, *18*, 229–243.
- Jacob, P., & Jeannerod, M. (2003). *Ways of seeing. The scope and limits of visual cognition*. Oxford: Oxford University Press.
- Jeannerod, M. (1997). *The cognitive neuroscience of action*. Oxford: Blackwell.
- Jiang, Y., et al. (2006). A gender- and sexual orientation-dependent spatial attentional effect of invisible images. *PNAS*, *103*, 17048–17052.
- Kahneman, D., & Tversky, A. (1973). Availability: A heuristic for judging frequency and probability. *Cognitive Psychology*, *5*, 207–233.
- Kakei, S., Hoffman, D. S., & Strick, P. L. (2001). Direction of action is represented in the ventral premotor cortex. *Nature Neuroscience*, *4*, 1020–1025.
- Keeley, B. (2004). Anthropomorphism, primatomorphism, mammalomorphism: Understanding cross-species comparisons. *Biology and Philosophy*, *19*, 521–540.
- Kennedy, J. S. (1992). *The new anthropomorphism*. Cambridge: Cambridge University Press.
- Keysers, C., & Gazzola, V. (2010). Social neuroscience: Mirror neurons recorded in humans. *Current Biology*, *20*(8), R353–R354.
- Lakshminarayanan, V. R., Chen, M. K., & Santos, L. R. (2011). The evolution of decision-making under risk: Framing effects in monkey risk preferences. *Journal of Experimental Social Psychology*, *47*, 689–693.
- Lingnau, A., Gesierich, B., & Caramazza, A. (2009). Asymmetric fMRI adaptation reveals no evidence for mirror neurons in humans. *Proceedings of the National Academy of Sciences*, *106*(24), 9925–9930.
- Marsh, B., & Kacelnik, A. (2002). Framing effects and risky decisions in starlings. *Proceedings from the National Academy of Sciences*, *99*, 3352–3355.
- McCoy, A., & Platt, M. (2005). Risk-sensitive neurons in macaque posterior cingulate cortex. *Nature Neuroscience*, *8*(9), 1220.
- Millikan, R. G. (2004). *Varieties of meaning*. Cambridge, MA: The MIT Press.
- Mitchell, R. W. (2000). A proposal for the development of a mental vocabulary, with special reference to pretense and false belief. In P. Mitchell & K. Riggs (Eds.), *Children's reasoning and the mind* (pp. 37–65). Hove: Psychology Press.
- Mitchell, R. W. (2002). Subjectivity and self-recognition in animals. In M. R. Leary & J. Tangney (Eds.), *Handbook of self and identity* (pp. 567–593). New York: Guilford Press.
- Mukamel, R., Ekstrom, A. D., Kaplan, J., Iacoboni, M., & Fried, I. (2010). Single-neuron responses in humans during execution and observation of actions. *Current Biology*, *20*, 750–756.
- Nanay, B. (2012). The philosophical implications of the Perky experiments. *Analysis*, *72*, 439–443.
- Nanay, B. (2013). *Between perception and action*. Oxford: Oxford University Press.
- Newell, B. R., & Shanks, D. R. (2014). Unconscious influences on decision making: A critical review. *Behavioral and Brain Sciences*, *37*, 1–19.
- Nisbett, R. E., & Wilson, T. D. (1977). Telling more than we can know: Verbal reports on mental processes. *Psychological Review*, *84*, 231–259.
- Pacherie, E. (2000). The content of intentions. *Mind and Language*, *15*, 400–432.
- Pacherie, E. (2011). Nonconceptual representations for action and the limits of intentional control. *Social Psychology*, *42*, 67–73.
- Papineau, D. (1987). *Reality and Representation*. Oxford: Blackwell.
- Penn, D. C. (2011). How folk psychology ruined comparative psychology: And how scrub jays can save it. In R. Menzel & J. Fischer (Eds.), *Animal thinking: Contemporary issues in comparative cognition* (pp. 253–266). Cambridge, MA: MIT Press.
- Perky, C. W. (1910). An experimental study of imagination. *American Journal of Psychology*, *21*, 422–452.
- Poincaré, H. (1905/1958). *The value of science*. New York: Dover.
- Reeves, A., & Craver-Lemley, C. (2012). Unmasking the Perky effect: Spatial extent of image interference on visual acuity. *Frontiers in Psychology*, *3*, 296. <https://doi.org/10.3389/fpsyg.2012.00296>.
- Rizzolatti, G., & Sinigaglia, C. (2008). *Mirrors in the Brain. How our minds share actions and emotions*. New York: Oxford University Press.
- Schnall, S., Benton, J., & Harvey, S. (2008). With a clean conscience. Cleanliness reduces the severity of moral judgments. *Psychological Science*, *19*, 1219–1222.
- Schwitzgebel, E. (2008). The unreliability of naive introspection. *Philosophical Review*, *117*, 245–273.

- Segal, S. J. (1972). Assimilation of a stimulus in the construction of an image: The Perky effect revisited. In W. Peter (Ed.), *The function and nature of imagery* (pp. 203–230). Sheehan. New York: Academic Press.
- Segal, S. J., & Nathan, S. (1964). The Perky effect: Incorporation of an external stimulus into an imagery experience under placebo and control conditions. *Perceptual and Motor Skills*, *19*, 385–395.
- Shettleworth, S. (2010). *Cognition, evolution, and behavior*. Oxford: Oxford University Press.
- Sinigaglia, C. (2009). Mirror in action. *Journal of Consciousness Studies*, *16*, 309–334.
- Sober, E. (2001). The principle of conservatism in cognitive ethology. In D. Walsh (Ed.), *Naturalism, evolution, and mind*. Cambridge: Cambridge University Press.
- Sober, E. (2005). Comparative psychology meets evolutionary biology: Morgan's Canon and cladistic parsimony. In L. Daston & G. Mitman (Eds.), *Thinking with animals: New perspectives on anthropomorphism* (pp. 85–99). New York: Columbia University Press.
- Sober, E. (2012). Anthropomorphism, parsimony, and common ancestry. *Mind and Language*, *27*(3), 229–238.
- Spaulding, S. (2013). Mirror neurons and social cognition. *Mind and Language*, *28*(2), 233–257.
- Spener, M. (2011). Using first person data about consciousness. *Journal of Consciousness Studies*, *18*, 165–179.
- Spener, M., & Bayne, T. (2010). Introspective humility. *Philosophical Issues*, *20*, 1–22.
- Tai, K., Zheng, X., & Narayanan, J. (2011). Touching a teddy bear mitigates negative effects of social exclusion to increase prosocial behavior. *Social Psychological and Personality Science*, *2*, 618–626.
- Timberlake, W. (2007). Anthropomorphism revisited. *Comparative Cognition and Behavior Reviews*, *2*, 139–144.
- Tolman, E. (1948). Cognitive maps in rats and men. *Psychological Review*, *55*, 187–208.
- Tomasello, M., & Call, J. (2008). Assessing the validity of ape-human comparisons: A reply to Boesch (2007). *Journal of Comparative Psychology*, *122*(4), 449–452.
- Tversky, A., & Kahneman, D. (1981). The framing of decisions and the psychology of choice. *Science*, *211*, 453–458.
- Umlàt, M. A., et al. (2008). How pliers become fingers in the monkey motor system. *Proceedings of the National Academy of Sciences, USA*, *105*, 2209–2213.
- Valdesolo, P., & DeSteno, D. (2006). Manipulations of emotional context shape moral judgment. *Psychological Science*, *17*, 476–477.
- Waller, S. (2012). Science of the monkey mind: Primate penchants and human pursuits. In J. A. Smith & R. W. Mitchell (Eds.), *Experiencing animals: Encounters between human and animal minds* (pp. 79–94). New York: Columbia University Press.
- Watson, J. B. (1930). *Behaviorism*. New York: W.W. Norton & Company Inc.
- Wegner, D. M. (2002). *The illusion of conscious will*. Cambridge: MIT Press.
- Wegner, D. M., Sparrow, B., & Winerman, L. (2004). Vicarious agency: Experiencing control over the movements of others. *Journal of Personality and Social Psychology*, *86*, 838–848.
- Williams, L. E., & Bargh, J. A. (2008). Experiencing physical warmth promotes interpersonal warmth. *Science*, *322*, 606–607.
- Winkielman, P., & Berridge, K. C. (2004). Unconscious emotions. *Current Directions in Psychological Science*, *13*, 120–123.
- Wynne, C. D. L. (2004a). The perils of anthropomorphism. *Nature*, *428*, 606.
- Wynne, C. D. L. (2004b). *Do animals think?*. Princeton: Princeton University Press.
- Zhong, C.-B., & Liljenquist, K. (2006). Washing away your sins: Threatened morality and physical cleansing. *Science*, *313*, 1451–1452.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.