



Human and Animal Mental Causation

Causación mental en humanos y animales

Angela Bischof

Northern Arizona University, Arizona, USA

angela.bischof@nau.edu

<https://orcid.org/0009-0000-5497-5370>

DOI <https://doi.org/10.48204/2805-1815.11500>

INFORMACIÓN DEL ARTÍCULO	ABSTRACT/RESUMEN
Recibido el: 15/01/2026 Aceptado el: 12/09/2026 Keywords: Nonhuman animals, automata, mental causation, epiphenomenalism, interventionism Palabras clave: Animal no humano, autómatas, causación mental, epifenomenalismo, intervencionalismo	Abstract: Debates about whether conscious mental states can cause any physical or mental events have traditionally been about human beings; non-human animals (hereafter ‘animals’) are rarely the focus. This is understandable as studying animal minds has proven difficult. I cannot experience an animal’s phenomenology, I cannot ask an animal what beliefs and intentions she has, and I cannot know why an animal chose to behave in one way rather than another. Despite the epistemic limitations of examining animal minds, I will argue that some human mental states are causally irrelevant if and only if this is also true of some animals, and mental causation occurs in some humans if and only if it also occurs in some animals. Resumen: Los debates sobre si los estados mentales concientes pueden causar estados físicos o mentales han sido tradicionalmente acerca de los seres humanos; raramente se han enfocado en animales no-humanos (de aquí en adelante, “animales”). Esto es comprensible, ya que es difícil estudiar la mente animal. No podemos experimentar la fenomenología de un animal, no podemos preguntarle qué creencias e intenciones tiene y no podemos saber por qué elige comportarse de una manera en lugar de otra. A pesar de las limitaciones epistémicas para estudiar la mente de los animales, argumentaré que algunos estados mentales humanos son causalmente irrelevantes si y solo si esto es correcto de algunos animales; asimismo, la causación mental ocurre en algunos humanos si y solo si ocurre en animales.

Introduction

Debates about whether conscious mental states can cause any physical or mental events have traditionally been about human beings; non-human animals (hereafter ‘animals’) are rarely the focus. Since animals have not been the focus of past discussions about mental causation, this paper will concentrate on mammalian species. I hope to show that this argument is sound with regards to mammals who share many physical and behavioural characteristics with humans. If I accomplish this, I believe many of these arguments can be extended to other animals (e.g. avian species). However, it is beyond the scope of this paper to discuss the implications this argument has for all animal species.

This is understandable as studying animal minds has proven difficult. I cannot experience an animal’s phenomenology, I cannot ask an animal what beliefs and intentions she has, and I cannot know why an animal chose to behave in one way rather than another. Despite epistemic limitations when examining animal minds, I will argue that some human mental states are causally irrelevant if and only if this is also true of some animals, and mental causation occurs in some humans if and only if it also occurs in some animals. Here, the “if and only if” expresses material equivalence rather than modal or a metaphysically necessary connection. The claim is about the actual extension of mental causation rather than about what must be true in every possible world. The qualification “some” is also deliberately conservative. My argument does not require that mental causation occur in all animals. Indeed, some animals may lack conscious mental states altogether and therefore cannot exhibit conscious mental causation.

Theories of Mental Causation

The problem of mental causation concerns whether mental states can make a causal difference to what happens in the physical world. I use ‘mental causation’ to refer to the causal relevance of a mental state to a subsequent event, particularly a behavioural event. More specifically, a mental state is causally relevant to a behaviour when a difference in the mental state would, under appropriate conditions, produce a corresponding difference in the behaviour. This formulation is intentionally compatible with the interventionist framework developed below.

For example, suppose I form the intention to reach for an apple and subsequently reach for it with my hand. It seems reasonable to believe that my intention caused my bodily movement. Yet the apparent causal role of mental states raises a difficult philosophical problem. If every mental state is accompanied by a physical or neural state, and the neural state is already sufficient to cause the relevant behaviour, what causal work remains for the mental state to do? The problem becomes particularly pressing if mental states are understood as something over and above physical states. If the physical state is sufficient to produce the behaviour, mental causation appears redundant or even impossible.

Mental States and Neural States

Before considering whether mental states are causally efficacious, it is important to distinguish mental states from the neural states that realize them. I understand mental states primarily in functional terms. A mental state is characterized not merely by its physical constitution, but by the functional role it plays in cognition and behaviour. Beliefs, desires, intentions, and other mental states are thus individuated, at least in part, by the ways in which they interact with other mental states, perceptual inputs, and behavioural outputs.

I use ‘intention’ to refer to a mental state directed toward performing a particular action. To intend to perform an action is therefore not the same as successfully performing that action, nor does intending an action require believing that the action will succeed. What matters is that the individual’s mental state is functionally organized towards performing the action. This use of ‘intention’ is not to be confused with the usage under which ‘intentional’ means representational or being about something (Brentano, 1874/2014).

Neural states, by contrast, are individuated in terms of their physical and neurobiological properties. For the purposes of this paper, I will adopt a broadly functionalist view according to which a particular mental state may be physically realized by a particular neural state without the mental state being identical to the neural state (Putnam, 1967; Fodor 1968).

This distinction is important for my argument because evidence for neural causation does not by itself settle the question of mental causation. Consider a dog that intends to bury a bone. The neural processes that realize this intention may play a causal role in producing the dog’s subsequent behaviour. Nevertheless, the question of mental causation concerns whether the dog’s intention, understood as a functional mental state, is itself causally

relevant to the behaviour. The relevant mental state is identified by its functional role – for example, by its role in organizing the dog's behaviour towards the goal of burying the bone – while its neural realization explains how that functional state is physically implemented.

Additionally, this distinction is important for the present discussion because the evidence considered below operates at both the behavioural and neural levels. Behavioural evidence provides evidence about the functional roles of an animal's mental states, while neuroscientific evidence helps identify the physical processes through which those states are realized.

This distinction does not require that mental states be causally independent of the neural realizers. Indeed, on the view assumed here, mental states are realized by physical states and cannot occur without an appropriate physical basis. The claim that mental states are causally efficacious is instead the claim that the mental level can be causally relevant to behaviour even when the underlying neural processes are also causally relevant. The subsequent interventionist analysis therefore asks whether appropriately varying a mental state would produce a corresponding change in behaviour, rather than assuming that causal relevance of its neural realization excludes the causal relevance of the mental state.

Interventionism

James Woodward's (2008) interventionist approach to causation provides a useful framework for asking what differences in mental states make a causal difference to behaviour. Interventionist theories of causation generally understand causal relationships in terms of how changes produced by appropriate interventions would affect other variables. Woodward writes: "X causes Y if and only if there are background circumstances B such that if some (single) intervention that changes the value of X (and no other variable) were to occur in B, then Y would change" (Woodward, 2008, p. 222). For example, suppose we wanted to determine whether poking a balloon with a needle causes the balloon to pop. Using Woodward's framework, we need to evaluate the truth value of the counterfactual 'if we poke the balloon with a needle rather than not poke the balloon with a needle, the balloon will pop rather than not pop.' If this counterfactual is true, then there is a causal relation between our poking the balloon with a needle and the balloon popping. If the counterfactual is false, then there is no such causal relationship; in some contexts, there are important differences

between ‘X causes Y’, ‘X partly causes Y’, ‘X causally contributes to Y’ and ‘X is causally relevant to Y’. However, these distinctions will generally not matter here.

If one wants to find out whether a mental state (M1) causes another mental state (M2), one needs to answer this question: “What would happen to M2 if we were to vary M1 by means of an intervention while holding N1 fixed?” (Woodward, 2008, p. 254), where N1 is the neural state that realizes M1. Whenever a mental state occurs, a corresponding neural state also occurs (i.e. N1 occurs when M1 occurs, and N2 occurs when M2 occurs). We cannot find out if M1 is causally relevant without also examining the relationships between the mental states and the neural states. Similar points apply *mutatis mutandis* for the question of whether a mental state causes a behaviour.

There are several versions of version interventionism, and the argument developed here is not intended to depend on every detail of any particular. Interventionist approaches to causation have been developed in several forms, including the manipulability account of Peter Menzies and Huw Price (1993) and the causal-model approaches of Judea Pearl (2009) and Christopher Hitchcock (2001). Although these approaches differ in important respects, they share the general idea that causal relationships can be understood in terms of changes produced under appropriate interventions. The present argument can be extended to these similar views. Menzies and Price (1993) differ from Woodward because they believe that human agency plays a significant role in the analysis of causation. This introduces a problem as it risks making causation overly dependent on what human beings can actually manipulate, but we clearly want to say that things cause other things even when humans cannot manipulate them. Woodward’s account provides a much cleaner solution since it forgoes questions about whether a human agent could actually manipulate the variable in question. Pearl’s (2009) framework can handle unobserved variables, which is particularly relevant to the animal mind problem. I focus on Woodward rather than Pearl because Pearl’s framework is much more formal and mathematical, whereas Woodward is well suited to answer philosophical questions concerning the causal relevance of mental states.

I work primarily with Woodward’s interventionist account because it provides one of the most fully developed versions of this approach, it is widely accepted, and it is attractive. Most importantly, Woodward’s account offers a particularly clear framework for analyzing mental causation in terms of changes in behaviour. This feature is especially important when

considering animal minds. Because we have limited epistemic access to animal cognitive states, we necessarily rely to a significant extent on behavioural evidence when attempting to determine what mental states they possess and whether those states influence their behaviour. An interventionist framework allows us to connect these observable behavioural changes to questions about the causal relevance of mental states. Rather than requiring direct access to animal mental states, we can ask whether a change in the animal's mental state would, under appropriate conditions, produce a corresponding change in behaviour. The relevant intervention need not be a manipulation that investigators literally perform on an animal's mental state. The question is whether there is an intervention variable that changes the mental state while holding fixed the relevant background conditions, such that the animal's behaviour would consequently change. Thus, the interventionist analysis is being used as a criterion for causal relevance rather than as a claim that researchers can directly manipulate animal intentions.

It is important to acknowledge that interventionist accounts of causation have been subject to a number of objections concerning their general adequacy, including questions about the nature of interventions, their relationship to counterfactuals, and whether interventionism can account for causal relationships that cannot be directly manipulated. I do not attempt to resolve these broader objections here. The purpose of adopting an interventionist framework is instead to examine what follows about mental causation if such a framework is accepted. These objections concern the adequacy of interventionism as a general theory of causation rather than its application to the particular question of human and animal mental causation. Accordingly, the argument that follows should be understood as conditional – if an interventionist account of causation is accepted, then there is not principled reason to treat human and animal mental causation differently.

For this reason, I also do not attempt to defend interventionism against every competing theory of mental causation, and I am not claiming that interventionism is the correct theory of causation. Instead, I argue that the central claims of this paper do not depend exclusively on an interventionist account – other theories of causation can, in principle, accommodate the arguments developed here.

Although it is beyond the scope of this paper to fully reframe my arguments about animal mental causation to all theories of causation, it is important to note that my view is

compatible with other theories. David Lewis' (1973) counterfactual theory of causation works well with my view because many interventionist accounts already use counterfactual language extensively; e.g. Woodward (2008) and Hitchcock (2001). Peter Machamer, Lindly Darden, and Carl Craver's (2000) mechanistic theory of causation aligns well with my position given my arguments concerning neural mechanisms explained later in this paper. They argue that entities cause things as far as they engage in productive activities. Using their account, I could argue that an animal mental state occupies a comparable functional position within a mechanism producing behaviour. The causal-power/dispositional theory developed by Stephen Mumford and Rani Lill Anjum (2011) differs metaphysically from Woodward's as it does not depend on counterfactuals or interventions. According to their view, mental states serve a particular functional role, disposing the individual towards certain behaviours. Applied to my arguments, animal mental states could have certain causal powers that can produce behaviours. I focus on Woodward rather than Mumford and Anjum since Woodward's framework reduces the metaphysical burden of proof given the epistemological problem of accessing animal minds. John Mackie's (1980) regularity theory of causation stipulates that a cause is usually one component of a larger set of conditions that together are sufficient for an effect. He is interested in patterns of conditions that reliably produce effects. Applied to my argument, an animal's mental state may be insufficient by itself to cause a behaviour, but potentially necessary within a particular sufficient configuration. Using his view, I could argue that an animal's mental state is therefore causally relevant to behaviour – animal mental states occupy the same relevant causal role as human mental states within the configurations of producing behaviour. Nonetheless, Mackie's view runs into similar epistemic problems that Woodward's account avoids – Woodward provides a more explicit strategy for distinguishing causal relevance from mere correlation without needing to directly observe the mental state.

Given my primary epistemic problem is that animal mental states are not directly observable, so I need a causal framework that allows me to reason from observable changes in behaviour to causal claims about mental variables. My argument here attempts to establish that there is no principled reason to treat human and animal mental causation differently. If a human and an animal possess functionally analogous mental states that occupy the same

relevant causal role in producing behaviour, then differences between the species do not, by themselves, provide grounds for assigning those states different causal statuses.

The interventionist framework therefore provides a way of making the question of mental causation precise – do differences in mental states make a difference to behaviour? If they do, then mental states are causally relevant; if they do not, then they are causally inert. The possibility that conscious mental states are causally inert is the central concern of epiphenomenalism. Before turning to the question of whether animals possess causally efficacious mental states, it is therefore useful to consider the epiphenomenalist challenge and, in particular, Thomas Huxley's influential formulation of it.

Epiphenomenalism

Huxley studied humans with brain lesions. He found humans without conscious mental states (due to their brain lesions) occasionally behave as though they do have conscious mental states. He examined a French army sergeant who experienced damage to his left parietal bone. The sergeant led a relatively normal life except he would occasionally lose consciousness. During these episodes, the sergeant behaved as though he were conscious; “he eats, drinks, smokes, walks about, dresses and undresses himself, rises and goes to bed at the accustomed hours” (Huxley, 1886, p. 201). When asked about his thoughts and feelings after he regained consciousness, the sergeant would not recall what had occurred; he had no memories of the episodes.

Huxley, when studying animals, challenged Rene Descartes' arguments that animals, but not humans, are mindless automata; they respond and react to their environments without any conscious states. Huxley believed that animals were just as conscious as humans due to his experiments on animals with similar brain inflictions.

For one experiment, Huxley severed the anterior division of frogs' brains and observed their behaviours. These frogs would swim if placed in water, avoid objects, and they displayed several other behavioural characteristics that imply consciousness (Huxley, 1886, p. 200). Because the anterior division of their brains was removed, their behaviours could not be caused by their mental states. Huxley concluded that epiphenomenalism was true for animals – animals have mental states, but their mental states are not causally relevant to their behaviours.

Huxley's (1886) main point is parity: "the argumentation which applies to brutes holds equally good of men; and, therefore, that all states of consciousness in us, as in them, are immediately caused by molecular changes of the brain substance... in men, as in brutes, there is no proof that any state of consciousness is the cause of the change in the motion of the matter of the organism" (p. 206). What is true of humans is also true of the animals.

Huxley's pathological cases involving damage to a region challenge the claim that conscious mental states are necessary for behaviour, but they do not by themselves establish that conscious mental states never contribute causally to behaviour. If complex behaviour can occur in the apparent absence of consciousness, then at least some behaviour does not require conscious mental states as a cause. Nonetheless, this does not establish that conscious mental states are never causally efficacious.

Some have argued that for some mental state, M1, and for some neural state, N1, M1 and N1 are type identical (Kim, 1966; Polger, 2009; Gozzano, 2023). Proponents of this position argue that the relationship between M1 and N1 is analogous to the relationship between H₂O and water; 'H₂O' and 'water' are simply two different terms that refer to the same substance. Others have argued that even if M1 and N1 are not type identical, there still exists a metaphysical relationship between M1 and N1; they claim that it is metaphysically impossible for M1 to occur without N1 also occurring.

Woodward believes that one of these types of relationships holds between mental and neural states. Because of the relationship between M1 and N1, Woodward argues that it is not possible to intervene on M1 while holding N1 fixed; every time M1 occurs, N1 simultaneously occurs, and every time M1 changes, N1 changes. For this reason, we cannot intervene on M1 without also intervening on N1. Woodward (2008) writes: "a scenario in which we imagine intervening to change M1 while holding the value of N1 fixed... seems inappropriate for assessing the causal efficacy or inertness of M1" (p. 256). For example, suppose we want to find out whether my intention to reach for an apple causes me to reach for an apple. On an interventionist approach, we ask whether, given an appropriate intervention that changes my intention, there would be a corresponding change in my behaviour. Woodward's account provides a particularly useful way of articulating this question. If we intervene to change my intention while leaving other relevant background

conditions fixed, would my behaviour of reaching for the apple also change? If yes, then the intention is causally relevant to the behaviour.

The debate over mental causation consequently does not depend only on whether humans have mental states; it also depends on what counts as a causal relationship in the first place. Someone who accepts epiphenomenalism will deny that human mental states cause behaviour, whereas someone who accepts Woodward's account may find mental causation possible. Importantly, neither position by itself establishes a difference between human and animal mental causation. If mental states are causally inert, then the question is whether they are causally inert in humans and animals alike. If mental states are causally relevant, the further question is whether anything about the relevant causal relationship makes it uniquely human.

Woodward's interventionist account can also be used to assess whether animal mental states are causally relevant. However, finding out whether animal mental states cause their behaviours introduces a different sort of problem. Consider Descartes' view, the view Huxley rejected; animals are not conscious. Descartes believed animals are similar to machines; they respond mechanically to the environmental stimuli and completely lack mental lives. If one shares Descartes' intuitions about animals, there is no reason to discuss whether or not animal mental states are causally relevant since the mental states simply do not exist. Before I can discuss whether or not animal mental states are causally relevant, I need to respond to Descartes and others (e.g. Davidson argues in his 1982 paper "Rational Animals") who share his intuitions. I will need to answer two questions: 1) Do animals have functional mental states? and 2) Do animals act on their functional mental states? I hope to show that behavioural and neuroscientific evidence supports a positive answer to both questions.

Behavioural studies

Koko, a gorilla who has been taught sign language, provides behavioural evidence that animals have mental states. She communicates with her caregivers using sign language and can sign nearly 1,000 words (Tanner et al., 2006). She tells her caregivers what her favourite foods are, what toys she wants to play with, or what she wants to do during the day. If we use language to find out whether other humans have mental states, we can use Koko's language to find out whether she has mental states.

Many have argued Koko's sign language abilities are not the same as a human's ability to use language. They claim Koko has been thoroughly trained and conditioned to behave in this way; she does not truly understand or conceive of the meaning of her signing behaviours; all she really understands is if she uses her hands in certain circumstances, she is rewarded. Whether or not Koko is actually using language in the same way as humans is not relevant to my argument. All I need to show is Koko is behaving in ways that indicate functional mental states.

For example, Koko knows the sign for 'thread' (Tanner et al., 2006). On one occasion, Koko incessantly signed the word 'thread' while picking her teeth. Koko only stopped these behaviours when she was given dental floss. She did not know the sign for the word 'dental floss'; her behaviour implies that she had the desire for dental floss and behaved in a way that allowed her desire to be fulfilled. If her language abilities are a product of conditioning and are not like the language abilities of humans, these types of innovated behaviours still suggest Koko has functional mental states, and she acts on these states.

A possible worry when evaluating behavioural research is anthropomorphism, attributing human mental states to animals when the animals really do not have mental states. Although I find this objection improbable, if one does not agree with the conclusions of the behavioural evidence, neuroscientific evidence should provide better reason to believe animals have and act on their mental states.

Neuroscientific evidence

There are many advances in neuroscience that begin with animal experimentation. The neuroscientific evidence presented here is intended to establish that the neural mechanisms involved in animal intentional behaviour are continuous with human mechanisms. If animal brains did not function similarly to human brains, data collected from animal experimentation could not successfully be applied to human subjects. When it is successful, we have strong reasons to believe human and animal minds are working in similar ways. Here I present an example to illustrate my point.

Researchers are attempting to create technology that allows paralyzed humans to move external devices with their minds with the hope that eventually this technology can be used to help paralyzed individuals walk again. Researchers (Chapin et al., 1999) first tested

the technology on rats by training them to move a lever to obtain water. When the rats moved the lever, the lever activated a robotic arm that delivered water to them. During the initial training, the rats received the water reward as a consequence of physically moving the lever. Researchers simultaneously recorded neural activity while the rats performed this behaviour. They then developed an algorithm capable of decoding neural activity associated with the rats' lever movements and connected the decoded neural signals directly to the robotic arm. Once this connection was established, the rats could cause the robotic arm to deliver water through the relevant neural activity without physically moving the lever. As the rats learned to control the robotic arm through their neural activity, their physical lever movements diminished or ceased while they continued to obtain the water reward.

After this successful experiment, the technology was extended to monkeys (Wessberg et al., 2000). When the monkey trials were successful, researchers attempted to implant the technology into human subjects, (Hochberg et al., 2006) and have progressively replicated, refined, and translated the basic phenomenon from rats and monkeys to human subjects, including situations in which the humans' ordinary motor output was unavailable (Rizzoglio et al., 2023; Collinger et al., 2013; Stavisky et al., 2018; Willett et al., 2021; Flesher et al., 2021).

For this research, paralyzed human participants were instructed to imagine particular hand and arm movements. Their motor-cortical neurons exhibited movement-specific patterns of activity, and researchers used these patterns to decode his intended movements. The decoded neural activity was then used to control external devices – the participant voluntarily opened and closed a prosthetic hand and used a robotic arm to grasp and transport an object.

Of course, not every behaviour is intentionally produced. For example, I may flinch when someone feints a punch without consciously intending to do so. Woodward's theory asks whether, when a relevant intention is present, varying that intention – or its neural realization – would produce a corresponding difference in behaviour. My argument therefore requires only that some behaviours be consciously intended and that, in such cases, the relevant intention is causally relevant.

The neuroscientific evidence does not, by itself, establish that animals possess mental states. A neural pattern that reliably precedes or accompanies a behaviour demonstrates a

relationship between neural activity and behaviour, but neural correlation alone does not establish that the neural activity constitutes or realizes a particular mental state. I therefore do not take the neuroscientific evidence as independent proof that animals have intentions or other conscious mental states.

Instead, the behavioural and neuroscientific evidence play different roles in my argument. The behavioural evidence provides the primary basis for attributing functional mental states to animals, while the neuroscientific evidence helps identify the physical mechanisms through which those states are realized, and supports continuity between animal and human neural mechanisms. The two questions – whether animals possess relevant mental states and whether those states are causally efficacious – must therefore be kept conceptually distinct. First, we need sufficient behavioural evidence to attribute a functional mental state to an animal. Second, we need to ask whether that functional state is causally relevant to the animal's behaviour. Neuroscience informs the second question by helping us understand the physical realization and mechanisms underlying the behaviour, but it does not settle the first question by itself.

Operating within the interventionist framework described above, and using Woodward's account as the primary model, we can assess whether animal mental states are causally relevant to their behaviours – if varying the relevant mental state would make a difference to behaviour, is the mental state causally relevant? Suppose we examine the causal relationship between my dog's intention to bury her bone and her act of burying the bone. We can ask whether an appropriate intervention that changes my dog's intention to bury the bone would produce a corresponding change in her behaviour. If yes, then her intention is causally relevant to her behaviour. The same basic question can be asked of human mental states. The relevant issue is therefore not whether the mental states belong to a human or an animal, but whether an appropriate intervention on the mental state would make a difference to the behaviour. If the same type of functional state occupies the relevant causal role in humans and animals, what principled reason is there to say that the human mental state causes the behaviour, but the animal mental state does not cause the behaviour?

Objections

There might be some resistance to grant animal mental states causal relevance. There are differences between the cognitive abilities of humans and animals, and perhaps the intuition is that the causal relevance of mental states is a capability that differentiates humans from animals. In the following sections, I will present two objections against my argument: one about the importance of phenomenology and the other concerns the complexity of human mental states. I plan on responding to each of these objections and showing that although there are cognitive differences between humans and animals, the way mental causation works is not one of these differences.

Phenomenology and Complexity

It can be argued that some human mental states are far richer than many animal mental states, and such experiences ought to play an important role in the discussion of causal relevance. For example, we have no way of knowing the way in which animal phenomenal mental states work. Humans cannot experience these states firsthand (as we can our own), and we do not yet have the ability to study or perform experiments on animals to evaluate and measure their phenomenological states. With humans, we can perform such experimentation because we have a shared language that can be used to express how these mental states feel. If animals do not have phenomenological mental states, then there does exist one way in which human and animal mental states differ.

Humans also have complex mental states – second-order desires are often thought to be a unique human cognitive capability (Frankfurt, 1971). Not only do humans have functional mental states, such as desires, but humans can also have desires about these desires. For example, I could (a) have the desire to eat ice cream and (b) have the desire to desire to eat something healthy, like an apple. Many assume that animals might have desires like (a), but they do not have functional mental states like (b). Because this level of cognitive complexity is not found in animals, this must indicate that there exists a difference between the causal relevance of human and animal mental states.

Response

There are several ways to respond to these objections. First, we can only ever truly know our own phenomenological experiences; no amount of experimentation can conclusively

determine that other human beings have phenomenological experiences like ours. Language cannot be the concluding factor because language can be deceiving. Currently, we have technology that speaks to us (i.e. the Siri application on iPhones), but this is not enough evidence to conclude that computers or cell phones experience phenomenological mental states. If one believes that other humans have phenomenological mental states, it cannot be due to language alone. If language is not sufficient to prove other human minds exist, it cannot be used to exclude the possibility that animals have minds and can experience phenomenological mental states.

Furthermore, there are humans who lack language – children and the severely cognitively disabled. Most do not believe that infants lack mental states because they do not yet have language. When a baby cries after undergoing surgery, it is reasonable to conclude that the baby is experiencing the mental state of pain, despite his inability to report the experience verbally. Similarly, when a dog whines after undergoing surgery, it is reasonable to conclude that the dog is experiencing the mental state of pain, despite her inability to report her experience verbally.

Let us assume for the sake of argument that animals do not have phenomenological experiences. This still does not refute my conclusion. I have shown animals have functional mental states, and these mental states work in the same way as human functional mental states. If humans do have additional mental states, this does not undermine the causal relevance of animals' functional mental states. Having other mental states only proves human and animal minds differ in some respects, but it does not challenge the idea that, fundamentally, animal, and human mental causation works in the same way.

Under Woodward's theory of causation, the complexity of mental states does not matter, although there is some evidence that some animals like primates have second order functional mental states (Osvath & Osvath, 2008). Woodward's theory encompasses all cases of causation – from (c) 'my desire to eat ice cream causes me to buy some ice cream' to (d) 'my desire not to desire to eat ice cream causes me to feel guilty after I consume the ice cream.' Even if animals are only capable of having mental states like (c), the causal relevance of these types of mental states still work in the same way as many human desires. I can concede that humans have more complex mental states and still maintain that fundamentally, the way in

which animal mental causation works is the same as the way in which human mental causation works. Complexity does not make a difference.

Conclusion

I have argued that some human mental states are causally irrelevant if and only if this is also true of some animals, and mental causation occurs in some humans if and only if it also occurs in some animals. I addressed objections concerning phenomenology and complexity. These objections do not undermine my arguments about the way mental causation works in animal and human minds, although they do provide evidence that there are metaphysical differences between the content or sophistication of human and animal mental states. Ultimately, I hope my argument provides more evidence that, at least in terms of mental causation, humans and animals are not as different as they are often perceived.

References

- Brentano, F. (2014). *Psychology from an empirical standpoint*. Routledge. (Original published in 1874)
- Chapin, J. K., Moxon, K. A., Markowitz, R. S., & Nicolelis, M. A. L. (1999). Real-time control of a robot arm using simultaneously recorded neurons in the motor cortex. *Nature Neuroscience*, 2(7), 664–670. <https://doi.org/10.1038/10223>
- Collinger, J. L., Kryger, M. A., Barbara, R., Betler, T., Bowsher, K., Brown, E. H. P., Clanton, S. T., Degenhart, A. D., Foldes, S. T., Gaunt, R. A., Gyulai, F. E., Harchick, E. A., Harrington, D., Helder, J. B., Hemmes, T., Johannes, M. S., Katyal, K. D., Ling, G. S. F., McMorland, A. J. C., ... Boninger, M. L. (2013). Collaborative approach in the development of High-Performance Brain–Computer interfaces for a neuroprosthetic arm: Translation from animal models to human control. *Clinical and Translational Science*, 7(1), 52–59. <https://doi.org/10.1111/cts.12086>
- Davidson, D. (1982). Rational Animals. *Dialectica*, 36(4), 317–327.
- Descartes, R. (2003). Animals are Machines. In S. J. Armstrong & R. G. Botzler (Eds.), *Environmental Ethics: Divergence and Convergence*. McGraw-Hill Humanities, Social Sciences & World Languages.
- Flesher, S. N., Downey, J. E., Weiss, J. M., Hughes, C. L., Herrera, A. J., Tyler-Kabara, E. C., Boninger, M. L., Collinger, J. L., & Gaunt, R. A. (2021). A brain-computer interface that evokes tactile sensations improves robotic arm control. *Science*, 372(6544), 831–836. <https://doi.org/10.1126/science.abd0380>

- Fodor, J. A. (1968). *Psychological explanation: An introduction to the philosophy of psychology*. Random House Trade.
- Frankfurt, H. G. (1971). Freedom of the will and the concept of a person. *The Journal of Philosophy*, 68(1), 5. <https://doi.org/10.2307/2024717>
- Gozzano, S. (2022). Blocking kripke's argument against the type-identity theory of mind. *Acta Analytica*, 38(3), 371–391. <https://doi.org/10.1007/s12136-022-00525-0>
- Hitchcock, C. (2001). The intransitivity of causation revealed in equations and graphs. *The Journal of Philosophy*, 98(6), 273. <https://doi.org/10.2307/2678432>
- Hochberg, L. R., Serruya, M. D., Friebs, G. M., Mukand, J. A., Saleh, M., Caplan, A. H., Branner, A., Chen, D., Penn, R. D., & Donoghue, J. P. (2006). Neuronal ensemble control of prosthetic devices by a human with tetraplegia. *Nature*, 442(7099), 164–171. <https://doi.org/10.1038/nature04970>
- Huxley, T. (1886). On the Hypothesis that Animals Are Automata, and its History. In *Select Works of Thomas H. Huxley*. John B. Alden.
- Kim, J. (1966). On the psycho-physical identity theory. *American Philosophical Quarterly*, 3(3), 227-235.
- Lewis, D. (1973). Causation. *Journal of Philosophy*, 70(17), 556-567.
- Machamer, P., Darden, L., & Craver, C. (2000) Thinking about mechanisms. *Philosophy of Science*, 67(1), 1-25.
- Mackie, J. L. (1980). *The cement of the universe: A study of causation*. Oxford University Press.
- Menzies, P., & Price, H. (1993). Causation as a secondary quality. *The British Journal for the Philosophy of Science*, 44(2), 187–203. <https://doi.org/10.1093/bjps/44.2.187>
- Mumford, S., & Anjum, R. L. (2011). *Getting causes from powers*. Oxford University Press, USA.
- Osvath, M., & Osvath, H. (2008). Chimpanzee (*Pan troglodytes*) and orangutan (*Pongo abelii*) forethought: Self-control and pre-experience in the face of future tool use. *Animal Cognition*, 11(4), 661–674. <https://doi.org/10.1007/s10071-008-0157-0>
- Pearl, J. (2009). *Causality*. Cambridge University Press.
- Polger, T. W. (2009). Identity theories. *Philosophy Compass*, 4(5), 822–834. <https://doi.org/10.1111/j.1747-9991.2009.00227.x>
- Putnam, H. (1967). Psychological predicates. In W. H. Capitan & D. D. Merrill (Eds.), *Art, Mind, and Religion* (pp. 37–68). University of Pittsburgh Press.

- Rizzoglio, F., Altan, E., Ma, X., Bodkin, K. L., Dekleva, B. M., Solla, S. A., Kennedy, A., & Miller, L. E. (2023). From monkeys to humans: Observation-based EMG brain-computer interface decoders for humans with paralysis. *Journal of Neural Engineering*, 20(5), 056040. <https://doi.org/10.1088/1741-2552/ad038e>
- Stavisky, S. D., Kao, J. C., Nuyujukian, P., Pandarinath, C., Blabe, C., Ryu, S. I., Hochberg, L. R., Henderson, J. M., & Shenoy, K. V. (2018). Brain-machine interface cursor position only weakly affects monkey and human motor cortical activity in the absence of arm movements. *Scientific Reports*, 8(1). <https://doi.org/10.1038/s41598-018-34711-1>
- Tanner, J. E., Patterson, F., & Byrne, R. (2006). The Development of Spontaneous Gestures in Zoo-Living Gorillas and Sign-Taught Gorillas: From Action and Location to Object Representation. *The Journal of Developmental Processes*, 1(1).
- Wessberg, J., Stambaugh, C. R., Kralik, J. D., Beck, P. D., Laubach, M., Chapin, J. K., Kim, J., Biggs, S. J., Srinivasan, M. A., & Nicolelis, M. A. L. (2000). Real-time prediction of hand trajectory by ensembles of cortical neurons in primates. *Nature*, 408(6810), 361–365. <https://doi.org/10.1038/35042582>
- Willett, F. R., Avansino, D. T., Hochberg, L. R., Henderson, J. M., & Shenoy, K. V. (2021). High-performance brain-to-text communication via handwriting. *Nature*, 593(7858), 249–254. <https://doi.org/10.1038/s41586-021-03506-2>
- Woodward, J. (2008). Mental causation and neural mechanisms. In *Being Reduced* (pp. 218–262). Oxford University Press. <https://doi.org/10.1093/acprof:oso/9780199211531.003.0013>