
The Causal Role(s) of Intentions

A Reply to Andrea R. Dreßing

Elisabeth Pacherie

In her commentary ([Dreßing this collection](#)) on my target article ([Pacherie this collection](#)), Dreßing suggests that the story I offer is not just a creation myth but also an attempt to give an explanation of the function of conscious intentions in the physical world and as such answerable to both metaphysical and empirical constraints. Here, I try to clarify which of my claims should be understood as simply speculations about the origins of our capacity of intentions and which I take to be empirical claims. In response to the metaphysical and empirical challenge Dreßing raises, I argue that Dretske's distinction between structuring and triggering causes may help us see how explanations in terms of physical properties and explanations in terms of mental properties may not compete but rather complement each other. I argue that this distinction may also help us assuage certain worries raised by neuroscientific findings.

Keywords

Causal exclusion | Conscious agents | Conscious intention | Creation myth | Intentional action | Intentions | Joint action | Mental causation | Neuronal correlates of intentions | Structuring causes | Triggering causes

Author

[Elisabeth Pacherie](#)

elisabeth.pacherie@ens.fr

Ecole Normale Supérieure
Paris, France

Commentator

[Andrea R. Dreßing](#)

andrea.dressing@uniklinik-
freiburg.de

Klinik für Neurologie und
Neurophysiologie
Universitätsklinikum Freiburg
Freiburg, Germany

Editors

[Thomas Metzinger](#)

metzinger@uni-mainz.de

Johannes Gutenberg-Universität
Mainz, Germany

[Jennifer M. Windt](#)

jennifer.windt@monash.edu

Monash University
Melbourne, Australia

1 Introduction

In her commentary, [Andrea Dreßing](#) ([this collection](#)) suggests that I might have been too timid in calling the story I tell about the social function of intentions in my target article ([Pacherie this collection](#)) a creation myth. She encourages me take a bolder stance, claiming that the story I offer is not just a myth but also an attempt to give an explanation of the function of conscious intentions in the physical world. Indeed, part of my story is intended as more than a myth and so my first task here will be to clarify where I

draw the line between empirical claims and myths.

Dreßing also points out that an explanation, as opposed to a mere myth, has to fit into the framework of current scientific knowledge and is therefore subject to both metaphysical constraints and empirical constraints. I concur. In what follows, I will argue, however, that my general predicament with regard to conceptual or metaphysical constraints is not so different from the predicament of the other myth-tellers I

discuss in my article, as Dreßing suggests. Nor indeed is it direr than the predicament all philosophers of mind working within a naturalistic framework face. Finally, certain empirical findings have been interpreted as showing that conscious intentions play no role in action initiation. I also try to address this challenge.

2 Myths vs. empirical claims

In my target article, I use the phrase “creation myth” first as a dramatization device. Typically, we do not feel the urge to formulate myths about things we deem insignificant. Talking of a social creation myth was thus a way of emphasizing the importance of the social function of intentions, a function largely neglected in traditional accounts of intentions. Second, I also wanted, following Velleman (2007), to convey a note of caution. A myth, as Dreßing points out, can neither be falsified nor empirically verified. It offers a possible explanation about a phenomenon, without making a claim about truth. But I perhaps wasn’t clear enough what I was trying to be cautious about and where I drew the line between empirical claims and ultimately unverifiable explanations. So let me now draw this line more firmly.

To do this, let me distinguish three different questions about intentions and examine how they may relate. The three questions are: *what* roles or functions (in a non-teleological sense) do intentions play in human agency? *How* can intentions play these roles? *Why* do we have intentions in the first place? In my view the *what*- and *how*-questions are both empirical questions for which mythical answers won’t do. The *why*-question, as I understand it, is a question about the origins of capacity for intention. How come we have such a capacity? Why was it established?

The focus of the account I proposed, as well as the focus of the alternative accounts by Bratman (1987), Anscombe (1963), and Velleman (2007) with which I contrast it in my article, is on the *what*- and *why*-questions. However, I offered my story as a creation myth only to the extent that it was meant to address the *why*-question. As answers offered to the *what*-

question, my claims were meant as empirical claims. I take it that the claims made by Bratman, Anscombe, and Velleman about the epistemic and pragmatic functions of intentions, when understood as answers to the *what*-question, should also be interpreted as empirical claims.

Now, how do the *what*- and the *why*-questions relate? One way to relate them is by assuming that intentions do not just have a function or functions in a value-neutral sense—things that they do—but a teleofunction in the evolutionary sense, that is, something that they do that confers some benefit or advantage on creatures with a capacity for intentions, and in this sense explains why these creatures have this capacity.

Velleman cautions us against this teleofunctional move. First, as his discussion of Bratman’s and Anscombe’s accounts makes clear, the *what*-question about intentions can be given complementary answers in terms of both pragmatic and epistemic roles, leaving us with several possible teleological stories. Second, Velleman also warns us against assuming direct links between answers to the *what*-question and answers to the *why*-question. The spandrel story he tells is meant to suggest that a capacity for intentions may only be a by-product of other capacities and thus that our capacity for intentions could be nothing more than an (admittedly very fortunate) accident. Finally, in calling his own story a creation myth as well, Velleman is also pointing out that our speculations about the origins of intentions are most likely beyond falsification or empirical verification.

Similarly, in offering my social function story as an answer to the *why*-question, I was not making a claim to truth. Rather, I was trying to broaden the terms of the debate to also include consideration of the social dimension of intentions. If we are considering what possible teleofunction intentions could have, then we should pay more attention to the benefits we derive from being able to act jointly in a flexible manner. If we are tempted by a story that views a capacity for intention as simply a by-product of more general capacities, then, among these more general capacities, we should pay

serious heed to our capacity for sociality and cooperativeness.

Turning now to the relations between the *what*-question and the *how*-question, I take it that the empirical standing of an answer to the *what*-question ultimately depends on whether this answer can be backed up by a convincing answer to the corresponding *how*-question. The validity of any empirical claim about the causal roles of intentions in human agency will remain in doubt unless one can see how it is at all possible for intentions to play these roles (Dreßing's metaphysical constraints), and it will also remain in doubt if appears to be in contradiction with well-established empirical facts (Dreßing's metaphysical constraints).

Since my claims about the functions of intentions qua answers to the *what*- rather than the *why*-question are intended as empirical claims, they are not insulated from these metaphysical and empirical worries. Let me address them in turn.

3 Metaphysical worries

Dreßing points out that my claim that intentions have a causal role to play in the online control of action confronts me with the problem of mental causation. She also suggests that this problem is more pressing for me than it is for the accounts of the functions of intentions proposed by Bratman and Anscombe. While I agree that the problem of mental causation is an issue for me, I disagree with her assessment that it isn't as serious a worry for these accounts.

First, let me clarify that when I talk about conscious intentions and their causal role, I am concerned with what [Ned Block \(1995\)](#) calls access consciousness rather than with phenomenal consciousness. In other words, my claims are about intentions qua conscious states exploiting and conveying information globally available in the cognitive system for the purposes of reasoning, speech, and high-level action control. My account thus faces the “easy” problems of consciousness rather than the “hard” problem ([Chalmers 1995](#)). I share [Chalmers](#) sanguine assessment

about phenomena pertaining to access consciousness:

There is no real issue about whether these phenomena can be explained scientifically. All of them are straightforwardly vulnerable to explanation in terms of computational or neural mechanisms. (1995, p. 201)

This is not to say, however, that in confining oneself to phenomena of access consciousness one can eschew all metaphysical conundrums. In particular, as pointed out by Dreßing, the mere fact that Cartesian dualism has fallen out of favour and that the vast majority of philosophers and cognitive scientists are nowadays willing to embrace some form of materialist monism doesn't insure the dissolution of philosophical worries about mental causation. The version of the problem of mental causation that non-reductive physicalists, whatever their exact persuasion, are confronted with is the Causal Exclusion Problem: how could mental properties play a causal role given that they appear to be screened off by their physical realizers?

Dreßing argues that this problem is more pressing for my view than for the pragmatic (Bratman) or the epistemic (Anscombe) creation myths, the reason being that these latter two teleological myths are about prior intentions and that neither “require any assumption about causality, as they do not involve a mind-world directed causality, but rather an intramental mental causality” ([Dreßing this collection](#), p. 6). Dreßing also claims that Velleman's view does not imply any explicit claim about causality either, since on this view intentions are a spandrel or a by-product.

I disagree with this assessment for three reasons. First, as I explained in section 2, while the speculative character of these stories qua answers to the *why*-question may justify labelling them as creation myths, the stories also offer answers to the *what*-question. In that regard their claims about the epistemic or pragmatic roles of intentions should be taken as empirical claims. Thus, even if we go along with Velleman's claim that a capacity for intentions

is a spandrel and that the epistemic and pragmatic functions of intentions are not teleofunctions, they are nevertheless functions in the ordinary functionalist sense and we still need an explanation of how intentions can play these epistemic and pragmatic roles.

Second, the Causal Exclusion Problem is a problem for anyone espousing a non-reductive form of materialist monism,¹ whether their primary concern is with intra-mental causation or with mind–world causation. Suppose that a state S has the mental property M (e.g., the property of being an intention to go to London on Monday) and a physical basis P , suppose that S' has the mental property M' (e.g., the property of being an intention to buy a train ticket to London) and a physical basis P' , and suppose that S'' has the mental property M'' (e.g., the property of being a belief that one will go to London on Monday) and a physical basis P'' . On a Bratmanian pragmatic account of intentions, I would want to be able to say that my intention to go to London on Monday causes, via further means-end reasoning, my intention to buy a ticket to London. But how can the mental property M of S play a causal role in bringing about a state S' with mental property M' , given that they appear to be screened off by the physical properties P and P' ? Similarly, with regard to the epistemic function of intentions, how could I say that my intention to go to London on Monday causes my belief that I will go to London on Monday, given that mental properties M and M'' appear to be screened off by the physical properties P and P'' ?

Third, while it is true that on Bratman's account future-directed intentions may only cause behaviour through the mediation of present-directed intentions, still Bratman insists that the whole point of having a capacity for intentions is to produce behaviour that contrib-

utes in the long run to our securing greater desire-satisfaction. Similarly, on a reliabilist reading of Anscombe's epistemic claim that intentions embody knowledge of our actions, they do so because intentions reliably cause what they represent. As Velleman puts it, “[u]nless an intention with the content ‘I’m going to move my toe’ reliably causes my toe to move, it won’t amount to practical knowledge” (Velleman 2007, p. 201). Thus, Bratman, Anscombe and Velleman cannot be exonerated from the task of explaining how mental states can cause behaviour.

With respect to the problem of mental causation, we are all in the same boat. The metaphysical standing of my account is no less or more precarious than the standing of these other accounts. Are we then all metaphysically doomed? Readers should not hold their breath; I have no new, unassailable solution to the problem of mental causation to offer. Yet, it would certainly be premature to claim that the problem of mental causation is insoluble. Many lines of response have been proposed and are currently being explored (for a review, see Robb & Heil 2014). I cannot discuss all these accounts here. Let me just say that the approach I find most congenial stems from Fred Dretske's work (Dretske 1988, 2004) on psychological explanations of behaviour. Dretske distinguishes between triggering causes and structuring causes, where a triggering cause is an event that initiates or triggers a causal chain of events, and a structuring cause the cause of the process or setup that makes a given triggering cause produce the effect it does. To take an example from Dretske (2004), moving a computer mouse is the triggering cause of cursor movement, but hardware and programming are the structuring causes of cursor movement. Dretske's central claim is that mental states and events are best analysed as structuring rather than triggering causes of behaviour. On this view there is no competition between physical and psychological or mental explanations, since they have different explananda. While the triggering physical properties explain bodily motion, i.e., explain why bodily motions occur at a certain point in time, the structuring mental properties explain

¹ While this issue is not at the heart of Bratman's preoccupations, I think we can safely assume that he would want his account of intentions to be compatible with physicalism. I won't dwell here on Anscombe's metaphysical view, except to say that she was no materialist herself but was also highly suspicious of Cartesian dualism (Anscombe 2008). Suffice it to say that many of the philosophers who nowadays embrace the view that intentions have an epistemic function, would want this claim to be compatible with a physicalist stance.

behavior, i.e., they explain why in circumstances of a certain sort, bodily motions of this kind rather than that kind are produced.

Much work remains to be done in order for us to understand more precisely how structural causes operate and in particular how they can do so in the dynamic way needed to account for the plasticity and flexibility of human behaviour. In this respect, Dretske's account remains largely under-developed (for recent work on this issue, see e.g., [Slors 2015](#); [Wu 2011](#)). Dretske's approach in terms of structuring causes has the great merit, however, of offering a potential solution to the Causal Exclusion Problem and to let us see how explanations in terms of physical properties and explanations in terms of mental properties may not compete but rather complement each other. As we will now see, thinking of intentions as structural causes of action rather than triggering causes can also help us assuage certain empirical worries.

4 Empirical worries

The claim that conscious intentions play a causal role in action production should be compatible with our best empirical knowledge on how action is produced. The main empirical worries this claim confronts come from neuroscientific findings that have been interpreted as showing that the time of onset of conscious intentions is not compatible with their being the initiators of actions.

The most famous of these experiments are Libet's studies on "readiness potential" ([Libet et al. 1983](#); [Libet 1985](#)). In these studies, subjects were asked to flex their wrist at will and to note when they felt the urge to move by observing the position of a dot on a special clock. While subjects were both acting and monitoring their urges (intentions, decisions) to act, Libet used an EEG to record the activity of prefrontal motor areas. On average, participants reported the conscious intention to act, which Libet called the W-judgement, about 200ms before the onset of muscle activity. By contrast, the EEG revealed that preparatory brain activity, termed by Libet type II readiness potential

(RP), preceded action onset by about 550ms. In other words, their brains started preparing the action at least 350ms before the participants became aware of their intention to act. This led Libet to the conclusion that the wrist-flexing actions in his experiments were not initiated by conscious intentions but were initiated instead by the (unconscious) RPs.

These experiments and Libet's interpretation of his findings have been widely discussed (see e.g., [Banks & Pockett 2007](#); [Bayne & Pacherie 2014](#); [Mele 2009](#); [Nahmias 2002](#); [Pacherie 2014](#); [Roskies 2011](#)) and commentators have pointed out a number of methodological problems with Libet's paradigm as well as conceptual problems with his interpretation of his results. Let me focus first on one methodological problem and one attempt to address it. I will then consider one conceptual problem

Libet argues that it is the RP rather than the conscious intention that initiate the agent's action. If RPs are the initiators of the action, there should be a robust correlation between them and the actions they cause: we should expect RP events to be "immediately" followed by the appropriate action, or, to put it the other way round, we should expect that when there is no movement, there is also no RP event. As several commentators have observed (e.g., [Mele 2009](#); [Roskies 2011](#)), the back-averaging techniques used in the experiment do not allow us to ascertain whether this is indeed the case. Because the RP on any one trial is obscured by neural noise, what is presented as "the RP data" is determined by averaging the data collected on a large number of trials. In order to compute this average, the EEG recordings on different trials need to be aligned, and this requires some fixed point that can be identified across trials. Since in Libet's experiments action onset serves as the needed fixed point for the alignment of EEG recordings, any RPs that are not followed by an action simply won't be measured, and so we don't know how robust the correlation between the RP and Libet-actions is.

In a recent experiment, Schurger and colleagues ([Schurger et al. 2012](#)) used a modified Libet task to circumvent the limitations of back-averaging techniques. Their aim was to

test the proposal that RPs correlate with pre-decision activity rather than, as Libet proposed, with activity that coincides with, or is subsequent to, the agent's decision. Schurger and colleagues proceeded on the assumption that the decisions of the participants in Libet's experiment can be modelled—as neural decision tasks typically are—in terms of an accumulator-plus-threshold mechanism: decisions are made when relevant evidence accumulated over time reaches a certain threshold. Given that in Libet's task subjects are explicitly instructed not to base their decision on any specific evidence, Schurger and colleagues proposed in this instance that the decision process amounts to simply shifting premotor activation closer to the threshold for initiation of the movement and waiting for a random threshold-crossing fluctuation in RP. Thus, Schurger and colleagues predicted the same premotor activation build-up as Libet when a movement is produced. However, whereas on Libet's post-decision interpretation of this build-up there should be no premotor activity (and hence no RPs) when no movement is produced, on Schurger and colleagues' stochastic decision model there should be continuous random fluctuations in RPs even when no movement is produced. Schurger and colleagues reasoned that it should be possible to capture these fluctuations by interrupting subjects in a Libet task with a compulsory response cue and sorting trials by their reaction times. On the assumption that the interrupted responses arise from the same decision accumulator as the self-initiated ones, and on the assumption that close-to-threshold activity reflects spontaneous fluctuations of RPs rather than mounting preparation to move building over the course of the entire trial, slow and fast reaction times should be distributed equally within trials. In their *Libetus Interruptus* task, they found, as they had predicted, that slow and fast responses to interruptions were distributed equally throughout the time span of the trial.

These results cast serious doubt on Libet's claim that the neural decision to move coincides with the onset of the RP, since spontaneous fluctuations of RPs happen all the time. There-

fore, they also cast doubt on his further claim that since RP onset precedes the urge to move by 350ms or more, conscious intentions can play no role in the initiation of the movement. If instead the neural decision to move coincides with a much later threshold-crossing event, it remains at least an open possibility that this event coincides with and constitutes the neural basis of a conscious urge to move. Schurger and colleagues take no stand on the exact relation between the conscious urge to move and their threshold-crossing event. They insist, however, that this threshold-crossing event should not be interpreted as *the* cause of the movement but rather as just one of the many factors involved in the causation of self-initiated movements. This leads me to my final point.

One conceptual problem with Libet's interpretation of his findings and also, as Dreßing points out, with most interpretations of neuroscientific experiments and a large part of the philosophical debates on mental causation and causal exclusion lies in the conception of causality that is assumed, “namely a temporal, linear, one-way causality” (Dreßing [this collection](#), p. 10). I agree with Dreßing's suggestion that a different concept of causation should be considered, one that allows for multiple causal processes to operate in parallel and to exert influence on one another. This is indeed the spirit of the dynamical model of intentions I have proposed elsewhere (Pacherie 2008). In particular, I insisted that a distal intention does not cease to exist and play a role once a corresponding proximal intention has been formed (and the same goes for proximal and motor intentions). What I suggested is that all three levels of intentions operate simultaneously, each exerting its own form of control, as well as operating together with unconscious processes. Following Dretske's lead, we can think of intentions as structuring rather than as triggering causes of action. On the dynamic hierarchical model of intentions I have proposed, we can further think of the structures set up by intentions as nested. This means that we don't need intentions to initiate actions for them to play a causal role in the production of action. This also means that the intentional online control that I argued was an

important pragmatic function of intention may be best conceived as a form of re-structuring, necessary only when the initial structuring is inadequate.

5 Conclusion

In her commentary, Dreßing suggested that the story I told about intentions should be viewed not just as a creation myth but as an attempt to give an explanation of the function of conscious intentions in the physical world. I tried to clarify exactly what I offered as merely a creation myth, namely the story given in answer to the question “*Why* do we have intentions in the first place?” and what I offered as empirical claims, namely my story as an answer to the question “*What* roles do intentions play in human agency?”

Dreßing also stresses that as an account of the roles intentions play in agency, my story has to meet both metaphysical and empirical constraints. In particular, she suggests that my claims about the role of intentions in action control makes the Causal Exclusion Problem more pressing for me than for other myth-tellers. I argued that the problem is actually equally pressing for all of us who want their views to be compatible with physicalism. I suggested that Dretske’s distinction between structuring and triggering causes and his view that mental properties should be understood as structuring causes may offer a solution to this metaphysical problem. Finally, Dreßing remarks that my claims concerning the role of conscious intentions appear to clash with certain findings from neuroscientific experiments. In response, I briefly discussed the most famous of these experiments, Libet’s RP experiments, and pointed out some of their limitations. I also questioned, together with Dreßing, the conception of causation with which these debates tend to operate, and suggested that Dretske’s distinction between structuring and triggering causes may also help to reconcile neuroscientific findings and claims about the causal roles of intentions.

References

- Anscombe, G. E. M. (1963). *Intention (second edition)*. Oxford, UK: Blackwell.
- (2008). *Faith in a hard ground: Essays on religion, philosophy and ethics (Vol. 11)*. Exeter, UK: Imprint Academic.
- Banks, B. & Pockett, S. (2007). Benjamin Libet’s work on the neuroscience of free will. In M. Velmans & S. Schneider (Eds.) *The Blackwell companion to consciousness*. London, UK: Blackwell.
- Bayne, T. & Pacherie, E. (2014). Consciousness and agency. In J. Clausen & N. Levy (Eds.) *Springer Handbook of Neuroethics* (pp. 211-230). Dordrecht, NL: Springer.
- Block, N. (1995). On a confusion about the function of consciousness. *Behavioral and Brain Sciences*, 18 (2), 227-247. [10.1017/S0140525X00038188](https://doi.org/10.1017/S0140525X00038188)
- Bratman, M. (1987). *Intention, plans, and practical reason*. Cambridge, MA: Harvard University Press.
- Chalmers, D. (1995). Facing up to the problem of consciousness. *Journal of Consciousness Studies*, 2 (3), 200-219.
- Dretske, F. I. (1988). *Explaining behavior: Reasons in a world of causes*. Cambridge, MA: MIT Press.
- (2004). Psychological vs. biological explanations of behavior. *Behavior and Philosophy*, 32 (1), 167-177.
- Dreßing, A. R. (2015). Conscious intentions: Do we need a creation myth? In T. Metzinger & J. M. Windt (Eds.) *Open MIND*. Frankfurt a. M., GER: MIND Group.
- Libet, B. (1985). Unconscious cerebral initiative and the role of conscious will in voluntary action. *Behavioral and Brain Sciences*, 8 (4), 529-566. [10.1017/S0140525X00044903](https://doi.org/10.1017/S0140525X00044903)
- Libet, B., Gleason, C. A., Wright, E. W. & Pearl, D. K. (1983). Time of conscious intention to act in relation to onset of cerebral activity (readiness-potential). The unconscious initiation of a freely voluntary act. *Brain*, 106 (Pt 3), 623-642.
- Mele, A. (2009). *Effective intentions: The power of conscious will*. New York, NY: Oxford University Press.
- Nahmias, E. (2002). When consciousness matters: A critical review of Daniel Wegner “the illusion of conscious will”. *Philosophical Psychology*, 15 (4), 527-541. [10.1080/0951508021000042049](https://doi.org/10.1080/0951508021000042049)
- Pacherie, E. (2008). The phenomenology of action: A conceptual framework. *Cognition*, 107 (1), 179-217. [10.1016/j.cognition.2007.09.003](https://doi.org/10.1016/j.cognition.2007.09.003)

- (2014). Can conscious agency be saved? *Topoi*, 33 (1), 33-45. [10.1007/s11245-013-9187-6](https://doi.org/10.1007/s11245-013-9187-6)
- (2015). Conscious intentions. In T. Metzinger & J. M. Windt (Eds.) *Open MIND*. Frankfurt a. M., GER: MIND Group.
- Robb, D. & Heil, J. (2014). Mental causation. In E. N. Zalta (Ed.) *The Stanford encyclopedia of philosophy (spring 2014 edition)*.
<http://plato.stanford.edu/archives/spr2014/entries/mental-causation>
- Roskies, A. (2011). Why Libet's studies don't pose a threat to free will. In W. Sinnott-Armstrong & L. Nadel (Eds.) *Conscious Will and Responsibility* (pp. 11-22). New York, NY: Oxford University Press.
- Schurger, A., Sitt, J. D. & Dehaene, S. (2012). An accumulator model for spontaneous neural activity prior to self-initiated movement. *Proceedings of the National Academy of Sciences*, 109 (42), E2904-E2913. [10.1073/pnas.1210467109](https://doi.org/10.1073/pnas.1210467109)
- Slors, M. (2015). Conscious intending as self-programming. *Philosophical Psychology*, 28 (1), 94-113. [10.1080/09515089.2013.803922](https://doi.org/10.1080/09515089.2013.803922)
- Velleman, D. (2007). What good is a will? In A. Leist & H. Baumann (Eds.) *Action in Context* (pp. 193-215). Berlin, GER: de Gruyter.
- Wu, W. (2011). Confronting many-many problems: Attention and agentive control. *Noûs*, 45 (1), 50-76. [10.1111/j.1468-0068.2010.00804.x](https://doi.org/10.1111/j.1468-0068.2010.00804.x)