CellPress

Trends in Neurosciences

Forum

A call for more clarity around causality in neuroscience

David L. Barack [1,*]
Earl K. Miller [2,*]
Christopher I. Moore, [3,*]
Adam M. Packer [4,*]
Luiz Pessoa [5,*]
Lauren N. Ross [6,*] and
Nicole C. Rust [7,*]

In neuroscience, the term 'causality' is used to refer to different concepts, leading to confusion. Here we illustrate some of those variations, and we suggest names for them. We then introduce four ways to enhance clarity around causality in neuroscience.

Causality is about understanding: given an event, what event(s) caused it? In neuroscience, we are often interested in things like the events in the brain that 'cause' behavior or the events in the brain that 'cause' other brain events. But what exactly do we mean when we say 'cause'? It turns out that when neuroscientists talk about causality, they refer to a diversity of concepts.

Consider, for instance, the following scenario:

> Following an association of a sound and a fearful experience, hearing the sound triggers a fear-related behavior.

Many neuroscientists would agree with this statement:

> The fear-related behavior triggered by hearing the sound is caused by processing in the amygdala, not the ear.

This statement follows from work showing that the amygdala plays a central role in the storage and recollection of fearful memories [1], coupled with the knowledge that the ear is no more involved in fear-related behavior than it is with any other behavior that depends on hearing. Neuroscientists who agree with this statement have internalized the concept of causality that the philosopher Ned Hall calls causal production: causes are events that produce other events [2]. The spirit behind this concept is that we do not want to map all possible influences, but instead focus in on the subset of events that are most integral. One rationale behind this is that the brain events involved in production are the most probable targets for diagnosing and treating brain dysfunction; if we want to help people with post-traumatic stress disorder, we probably want to focus on the amygdala, not the ear.

A second concept of causality in neuroscience is that causes are factors that events depend on. Consequently, any event that influences another event is causally related to it. Hall calls this broader definition 'causal dependence', and Woodward has analyzed this with an interventionist account [2,3]. In the example above, the ear does cause fear behavior insofar as the ear exists in the causal chain leading up to fear. This concept of causality was recently championed for mapping human brain function and identifying therapeutic targets [4], and it is widely present in statistics [5]. One rationale behind it is that therapeutic targets do not need to be limited to those involved in production but can act through other means as well. Some therapies will act in ways that compensate for dysfunction via a route that is not involved in production. Other therapies will target events that lie outside the brain, such as strategies to prevent the intake of addictive substances in the case of substance-use disorders. In both cases, these targets will be missed if neuroscience focuses too narrowly.

Another rationale for causal dependence is that narrower definitions of causality, like production, often oversimplify the brain by assuming feedforward causal chains and localized processing, whereas the brain is full of complex recurrent loops and distributed processing. These oversimplifications can lead researchers astray: for example, to erroneous interpretations of brain perturbation experiments. Proponents of causal dependence argue that the best path forward for neuroscience begins by defining causality as dependence, followed by understanding the specific ways that events influence one another.

There are a number of other concepts of causality prevalent in neuroscience. For example, neuroscientists often ground causal claims by the gold standard for establishing causality: that causal influences hold up to randomization [6]. We call this causal demonstration. It begins with the widely accepted notion that correlation and causation should not be confused. Causal relationships between, for example, brain activity and behavior can be tested by perturbing brain activity states in a randomized way to differentiate those that matter (are causal) from those that do not (are epiphenomena). In the example above, randomization of activity in both the amygdala and the ear would reveal a causal influence, and so this concept conflicts with causal production. It maps more directly onto the interventionist accounts described above for causal dependence, where perturbing a cause leads to changes in its effect.

In sum, neuroscientists do not have a unified, singular concept for causality; different researchers use different definitions. To avoid confusion and facilitate progress, we offer four suggestions.

First, when using the term 'causal', researchers should do their best to define what they mean. Even better, they should consider adding modifiers to causal, such as 'causal production' for clarity. We have provided a few suggested terminologies here. When those are not appropriate, we suggest introducing others. This will help neuroscientists build a lexicon around causality.

Second, it would be beneficial for philosophers, neuroscientists, and experts in causal inference to work together to describe how these and other concepts about causality in neuroscience relate to one another. Should different concepts about causality be thought of as a hierarchy that includes one broad definition complemented by narrower ones? Or as many partially overlapping concepts? Or would some other classification be more suitable? Crucially, the outcome of those efforts should be communicated in a manner that is accessible to neuroscientists, reflective of their work, and useful to the goals of the field. Additionally, answers to these questions should be flexible enough to capture the diversity and complexity of neurobiological systems, but also rigorous in how they distinguish causal relationships from noncausal ones. We anticipate that multiple concepts will be required to capture causality for cellular-level mechanisms (e.g., neurotransmission) versus pathways (e.g., the routing of information through anatomically defined circuits) versus other types of descriptions (e.g., the geometry of population activity) [7].

Third, there is a need to create a better framework for executing and interpreting brain perturbation experiments. Recent progress on this has been made [4,6,8,9], but more work remains. One source of confusion is diaschisis: the change in the function of a brain area that results from the loss of its input due to perturbation in a distant brain area. This can happen in acute or chronic lesion studies, and when it does it can lead to erroneous conclusions about the function of the site of perturbation. We need better and more broadly agreed-upon ways to disambiguate causal relationships in the brain, given its highly interconnected networks. Another challenge is randomization. While the importance of randomizing one variable relative to all others is conceptually clear, in practice this is often difficult to achieve. Modern perturbation methods – such as targeted perturbation (e.g., [10,11]) – are powerful, but the results of these experiments can be difficult to interpret. To avoid erroneous conclusions, the causal relationships between the variables of interest need to be carefully considered. Given the complexities of inactivation and activation brain perturbation experiments, they should be regarded as one tool for inferring brain function, but not prioritized at the expense of other approaches such as correlative measures. As in solving all hard problems, triangulating evidence is ideal.

Finally, neuroscience would benefit by developing a better vision for the path forward and its relationship with causality. Some would argue, for instance, that causal production is misguided, and that we should define causality exclusively as dependence. But we clearly do not seek to describe all possible causal dependencies. For example, to help people with attention-deficit disorders, there is no need to exhaustively document all possible stimuli and processes that might distract these individuals. So how then do we conceptualize what it is that we are trying to achieve with regard to causality in neuroscience? The immediate answer will be different for different researchers, depending upon their goals; some may seek to identify a therapeutic target to treat a particular type of brain dysfunction, whereas others may seek to determine the contribution of a particular circuit component to normal function. In both cases, pinpointing what causes what is a central challenge.

None of these issues is easily addressed. But addressing them is crucial for moving neuroscience forward.

Acknowledgments

We thank Arash Afraz, Konrad P. Kording, Michael D. Fox, and Shan H. Siddiqui for helpful comments. This work was supported by the National Institute on Drug Abuse at the National Institutes of Health (K99DA048748 to D.L.B.), the National Science Foundation (CAREER award 1945647 to L.N.R. and award 2043255 to N.C.R.), the Office of Naval Research (N000142212453 to E.K.M.), the JPB Foundation (to E.K.M.), and the Wellcome Trust (A.M.P.).

Declaration of interests

The authors declare no competing interests in relation to this work.

[1]Departments of Neuroscience and Philosophy, University of Pennsylvania, Philadelphia, PA, USA
[2]The Picower Institute for Learning and Memory and Department of Brain and Cognitive Sciences, Massachusetts Institute of Technology, Cambridge, MA, USA
[3]Carney Institute for Brain Science, Department of Neuroscience, Brown University, Providence, RI, USA
[4]Department of Physiology, Anatomy, and Genetics, University of Oxford, Oxford, UK
[5]Department of Psychology and Maryland Neuroimaging Center, University of Maryland, College Park, MD, USA
[6]Department of Logic and Philosophy of Science, University of California, Irvine, CA, USA
[7]Department of Psychology, University of Pennsylvania, Philadelphia, PA, USA

*Correspondence:
dbarack@gmail.com (D.L. Barack),
ekmiller@mit.edu (E.K. Miller),
christopher_moore@brown.edu (C.I. Moore),
adampacker@gmail.com (A.M. Packer),
pessoa@umd.edu (L. Pessoa), rossl@uci.edu (L.N. Ross), and
nrust@psych.upenn.edu (N.C. Rust).

https://doi.org/10.1016/j.tins.2022.06.003

References

1. Josselyn, S.A. *et al.* (2015) Finding the engram. *Nat. Rev. Neurosci.* 16, 521–534
2. Hall, N. (2004) Two concepts of causation. In *Causation and Counterfactuals* (Collins, J. *et al.*, eds), pp. 225–276, MIT Press
3. Woodward, J. (2003) *Making Things Happen: A Theory of Causal Explanation*, Oxford University Press
4. Siddiqi, S.H. *et al.* (2022) Causal mapping of human brain function. *Nat. Rev. Neurosci.* 23, 361–375
5. Pearl, J. (2009) *Causality* (2nd edn), Cambridge University Press
6. Jazayeri, M. and Afraz, A. (2017) Navigating the neural space in search of the neural code. *Neuron* 93, 1003–1014
7. Ross, L.N. (2021) Causal concepts in biology: how pathways differ from mechanisms and why it matters. *Br. J. Philos. Sci.* 72, 131–158
8. Jonas, E. and Kording, K.P. (2017) Could a neuroscientist understand a microprocessor? *PLoS Comput. Biol.* 13, e1005268
9. Wolff, S.B. and Ölveczky, B.P. (2018) The promise and perils of causal circuit manipulations. *Curr. Opin. Neurobiol.* 49, 84–94
10. Rickgauer, J.P. *et al.* (2014) Simultaneous cellular-resolution optical perturbation and imaging of place cell firing fields. *Nat. Neurosci.* 17, 1816–1824
11. Packer, A.M. *et al.* (2015) Simultaneous all-optical manipulation and recording of neural circuit activity with cellular resolution *in vivo*. *Nat. Methods* 12, 140–146