**A brief and straightforward summary of our discussions about causality in brain research**

Channeled by: Nicole Rust

Based on: https://twitter.com/NicoleCRust/status/1524750391224676355

No doubt everyone can agree that science works best when we can eliminate (or at least minimize) confusion. This discussion about causality began with the observation that the field of brain research uses the term 'causality' in at least two distinct ways. These different ways are reflected in the work of our very thoughtful colleagues Dani Bassett and Konrad Kording, but it's certainly not just them. It turns out that 'causality' is not just one thing, and we lack consensus definitions in brain research about how to define its variations.

This fact probably has important implications - after all, we throw around the term 'causal' quite a lot these days, as we place increasing emphasis on its importance and we are emboldened with a host of new tools to test it, like optogenetics. But what exactly do we mean when we say 'causal'? It's worth noting that brain research isn't the only field wrestling with these issues – many other fields are too. Including philosophy, which has been debating this topic for centuries and continues to do so today. This document is intended to be a simple and straightforward summary of the issues we've been discussing, as a community, on Twitter. The goal here is to highlight what the issues are in an effort to arrive at consensus, where possible. And where what is described below is off point, let's work together to fix it (progress!).

*

Here are two examples of the different ways that that brain researchers use the term 'causality':

On the Nature of Explanations Offered by Network Science: A Perspective From and for Practicing Neuroscientists
Maxwell A. Bertolero, Danielle S. Bassett
https://onlinelibrary.wiley.com/doi/full/10.1111/tops.12504

> *For example, consider a thought experiment in which we destroy a particular brain region that functional neuroimaging has implicated in a particular cognitive process. Because the animal would no longer be able to engage in that cognitive process, one might (wrongly) say that we have uncovered evidence that that region causes that function. However, this is where neuroscientists equating necessity with causality can lead to failure; it is entirely possible that that region is in fact upstream of the region actually performing the relevant computation, and thus the lesion study provides some evidence but not sufficient evidence for a causal mechanistic explanation.*

Causal mapping of human brain function
Shan H. Siddiqi, Konrad P. Kording, Josef Parvizi & Michael D. Fox
https://www.nature.com/articles/s41583-022-00583-8

> *.. a single axiomatic definition for causality: an event is causal if, with all else being equal, its presence or absence affects the probability of an outcome. Even if the cause is not necessary or sufficient to induce the effect, this probabilistic definition can still reveal treatment targets. For*

*instance, smoking does not always cause lung cancer and lung cancer is not always caused by smoking, but smoking single-handedly increases the risk of lung cancer. Following this definition, we can reduce the risk of lung cancer with interventions that reduce smoking.*

Note that the example that Bertolero & Bassett regard as 'not casual' would be regarded by Siddiqi et al. as 'causal'. So it appears that brain researchers are using the same term to reflect different concepts. To minimize confusion going forward, it would be great to arrive at a better understanding of the ways in which this term is used, as well as consensus definitions for these concepts. Below are three questions that it would be productive for the community to address.

*Question 1) Do these two concepts of causality map onto 'causal production' and 'causal dependence' by the philosopher Ned Hall? If so, should we call them that?*

**Two Concepts of Causation**
Ned Hall
https://www.fitelson.org/269/Hall_TCOC.pdf

*Causation, understood as a relation between events, comes in at least two basic and fundamentally different varieties. One of these, which I call "dependence", is simply that: counterfactual dependence between wholly distinct events. In this sense, event c is a cause of (distinct) event e just in case e depends on c; that is, just in case, had c not occurred, e would not have occurred. The second variety is rather more difficult to characterize, but we evoke it when we say of an event c that it helps to generate or bring about or produce another event e, and for that reason I call it "production".*

Setting aside whether you think either of these concepts are relevant or useful in brain research, it's helpful to acknowledge that these are in fact concepts that exist in brain research as well as philosophy. By defining them, we can have a constructive discussion about them. And as suggested by Chris Moore, by adding modifiers to 'causality', we clarify both that causality is not just one thing as well as define what it is we are talking about. Hall's concepts seem to map onto the examples outlined above, where the Bertolero/Bassett usage reflects causal production and the Siddiqi et al. usage reflects causal dependence. Does everyone agree? Below I assume that the answer is 'yes'.

*Question 2) What type(s) of causality are we after in brain research? And why?*

*The spirit behind defining causality as production:* This spirit is captured in an example like reading. What does it mean to understand reading and the brain areas that are causally involved in it? It means that we are interested in identifying brain areas like the *human visual word form area* (which does something akin to transforming simple shapes into letters and words) and *Broca's area* (which

supports language comprehension). We aren't interested in including the eye, because processing in the eye doesn't support reading any more than it does any other type of seeing. One motivation behind the emphasis on production in our investigations of reading is to determine why some people have trouble with reading in the absence of other visual deficits. This knowledge is the type of understanding that we need to diagnose reading problems as well as develop new treatments for them.

*The spirit behind causal dependence:* In short, the spirit behind causal dependence is that brain researchers cannot just focus on causal production to provide the foundation for developing new treatments for brain dysfunction; a broader definition is required. Yes, some therapeutic targets will be the ones involved in production - dysfunctional circuits that can be modified using therapeutic brain stimulation. Or at a lower level, serotonin transporters, which can be modified via medications. However, some effective treatments will not act through production – for example, some will act by compensating for dysfunction via a route that does not act through production. Still other treatments targets will originate outside the brain, such as stressful life events that lead to depression or anxiety, or addictive substances that can interfere with everyday function. In at least some cases, behavior-based therapies may be more effective that brain-based approaches. By focusing exclusively on *how the brain produces behavior*, these crucial targets will be missed. Consequently, the more expansive view of causality reflected by causal dependance is required to meet our ultimate goal of diagnosing and treating brain dysfunction. (Thanks to Shan Siddiqi for help with this explanation).

*Question 3) Conceptually, what is the best path forward for brain research?*

The tensions between *causal production* and *causal dependence* in brain research are revealed in the following example offered up by the Earl Miller (in a slightly different context) who was channeling the late and great brain researcher, Charlie Gross:

> *Imagine that a subject is performing a demanding but simple visual task in which they press one of two buttons based on a decision about some visual input (e.g. Which way are the dots moving?). We want to know: what brain areas play a causal role in supporting this task? Then, as a causal perturbation, a researcher hits the subject on the toe with a hammer and this disrupts their performance. Question: Is the toe causally involved in this task?*

Under the definition of causality as *production* the answer is no; under the definition of causality as *dependence* the answer is yes. While different brain researchers might align themselves with one or the other definitions of causality, what unites everyone is the agreement that it's not a very useful experiment. More broadly, brain researchers agree that the agenda of identifying all possible

causal dependencies that influence behavior is not a useful path forward for brain research. This is because this agenda would ultimately consider every part of the brain and body causally related to every possible behavior - for example, I could also distract you with a sound (activates your ear and auditory cortex) or by putting a pillow over your face so you can't breathe (activates alarm signals in your amygdala), ad infinitum.

So … acknowledging that important treatment targets can exist beyond parts of the brain involved in *causal production*, but also that we aren't interested in identifying all possible things that are involved in *causal dependence*, how might we conceptualize the best path forward for brain research? That's the big question. The first sketches of the answer include: 1) We prioritize our investigations by their potential for therapeutic relevance (no toe hitting experiments); 2) We acknowledge that our tools lie on a spectrum with regard to the degree to which they provide evidence for causality (with convergent evidence from lesions + stimulation providing the strongest evidence; Siddiqi et al), and 3) We use these tools to map out the ways in which causes create their influence on behavior by combining them with computational models that describe causal chains.

*

Another important discussion floating around is that the idea of causal production implicitly assumes a system with a relatively simple structure, and the concept is not very applicable in a complex system with many recurrent loops, like a brain. In that type of system, there's really no 'production' to be had. So the only sensible path forward is to proceed with definitions of causality as 'causal dependence', followed by the creation and testing of models that describe causal chains.

*

A final idea that has been expressed is complementary but more philosophical; it highlights that it is actually impossible to define 'a' singular cause for anything. As described by Shan Siddiqi:

> *I was born because the surgeon cut me out of my mother's uterus, but also because my mom's egg was fertilized with my dad's sperm, but also because my parents met, but also because my grandparents met, and so on. Of course, this can be traced to infinity, which would be absurd, so one might imagine that it actually starts with something, which many people referred to as "God" (this viewpoint was popularized by Aquinas, but originated with Ibn Sina). Regardless of where it originates, we focus on what matters – if you want to modify an outcome, you can work on any part of the causal chain.*

Some of these thoughts map onto ideas about different levels of explanation in brain research - the idea that brain activity produces behavior but what is happening at the molecular and cellular level determines brain activity. But this way of thinking also encourages us to appreciate that 'a' singular cause does not exist for anything. Moreover, it reminds us that if we think about the relationship between the brain and behavior as having 'a' cause, we are fooling ourselves.

*This discussion is distinct from (but not completely unrelated to) these issues:*

- *Causation can happen probabilistically.* Causes can influence the probability that something happens (as opposed to influencing it with certainty). This issue was central to a lot of discussion in the mid-1900s, and it was central to Bradford Hill's work extending ideas about the relationship between smoking and lung cancer to a larger conceptual framework about how the environment influences health. While probabilistic causes are highly relevant to noisy brains and behavior, that's not what we are discussing here.
- *The role of randomization in establishing causality and ways to get around it.* Randomization is emphasized as the gold standard for establishing causality, including by Jazyeri & Afraz (Neuron, 2017). But when that's not possible, there are other ways to infer causal relationships (summarized by Marinescu, Lawlor & Kording, Nature Human Behavior, 2018). While these issues are highly relevant to how we go about testing causality, we aren't talking about randomization here.
- *The differences between causality in a model and the evidence used to evaluate that model.* These issues were described by Max Bertolero and Dani Bassett in the same paper linked above to highlight (e.g.) the importance of correlation-based approaches in brain research. That discussion is very important, but it's not what we are talking about here.
- *'Necessary and sufficient' isn't very applicable to brains.* These issues were described in a paper by Yoshihara & Yoshihara (J Neurogenet., 2018). What would it mean for a brain area or pattern of brain activity to be 'sufficient' to produce a behavior? The concept does not make sense because things like behavior require many causes to happen. The same is true of the relationship between individual genes and development. The inability to use sufficiency as a criterion for causality is important, but it's not what we are talking about here.
- *The complexities of interpreting brain perturbation experiments.* Many of the complexities associated with testing causality and interpreting the results of brain perturbation experiments were nicely summarized by Jazyeri & Afraz, 2017. Among all of the 'unrelated' things in this bullet list, these issues are probably the most relevant (although not quite central) to what we are discussing here. Examples: *Off-target effects* refer to changes of variables that the perturbation directly interacts with but are not the intended target (e.g., stimulating the axons of passage when microstimulating a cluster of neurons). *Secondary effects* are effects downstream of the intended target that may or may not be the key variables influencing the behavior. *Off-manifold perturbations* refer to the complexity of interpreting the results of experiments that put the brain into unnatural states.