Found Sci
DOI 10.1007/s10699-016-9511-x

Disentangling Mechanisms from Causes: And the Effects on Science

John Protzko[1]

Abstract Despite the miraculous progress of science—it's practitioners continue to run into mistakes, either discrediting research unduly or making leaps of causal inference where none are warranted. In this we isolate two of the reasons for such behavior involving the misplaced understanding of the role of mechanisms and mechanistic knowledge in the establishment of cause-effect relationships. We differentiate causal knowledge into causes, effects, mechanisms, cause-effect relationships, and causal stories (narrative accounts of the whole process). Failing to understand the role of mechanisms in this picture, including their absence of knowledge or incorrect specification, leads to errors of wrongly excluding cause-effect relationships or wrongly inferring their existence. We highlight the primary of causality over mechanistic knowledge and causal stories in scientific reasoning.

Keywords Epistemology · Causes · Mechanisms · Scientific errors

1 Introduction

Understanding how we come to learn *causes* and *mechanisms* is of central importance to proper scientific practice. It appears, however, that the theoretical and philosophical discourse entangles causes and mechanisms into that of being the producer of, or responsible for, an effect. Scientists are more likely to see these two as separate, although uncertainties do continue to arise. Here I wish to argue that causes and mechanisms are distinct elements within a broader picture of causality.

This essay concerns the epistemology of causality and mechanisms, not necessarily the metaphysical nature of causes and mechanisms (see Mackie 1974). In many scenarios, the knowledge of the presence, validity, or absence of mechanisms are used as justifications

✉ John Protzko
protzko@gmail.com

[1] University of California, Santa Barbara, CA, USA

Published online: 19 December 2016

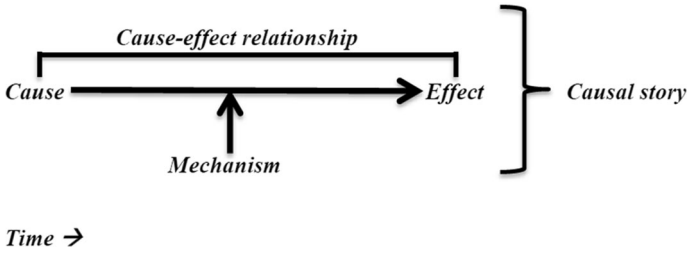

Fig. 1 Schematic of causes, effects, mechanisms, cause–effect relationships and causal stories

for the existence of causal connections (see examples in Koslowski 1996; Rouder and Morey 2011). Furthermore, the absence of a known mechanism is often brought up in scientific debates to question the validity of a causal finding (e.g. Shipstead et al. 2012; Rouder et al. 2013). We argue why this step from knowledge of mechanism to knowledge of cause–effect is unwarranted and potentially misleading. This thesis has a number of implications for the practice of science.

In addition, I present two types of errors that are often encountered by both the scientific and lay communities: faulty causal exclusion and faulty causal inference. Faulty causal exclusion occurs when a cause–effect relationship is denied due to a lack of a known mechanism. Faulty causal inference occurs when a cause–effect relationship is assumed to exist from the knowledge of how the proposed cause would bring about the effect. I argue that these errors occur because of the entanglement of causes and mechanisms.

2 The Current Model Contrasted with Past Models

The argument which I propose is schematized in Fig. 1 and includes the following: causes, mechanisms, effects, cause–effect relationships, and causal stories. By arguing about how these elements are distinct and how they interact, it is my hope to clarify the confusion about causes and mechanisms and explain how science comes to faulty causal exclusions and inferences.

This schema entails the central argument of this paper. The first aspect of this theory is that causes, effects, and mechanisms are distinct elements within a causal story. This is in contrast to many theories which posit mechanisms and causes to be functionally, if not explicitly, synonymous. The second aspect of this theory is that causes, mechanisms, and effects occur over time. Causes are the first in the temporal chain, effects are the end of the temporal chain, and mechanisms occur in between the two. The third aspect is that the cause–effect relationship is a distinct element from the causal story.[1] In this model I am not making a decisive claim about what causes or mechanisms are, I am making a claim about how they interact to bring about effects and expound on the implications of this for scientific practice. The heuristic definitions used here are as follows: causes are what bring about effects; mechanisms are how causes bring about those effects; a causal story is the overall narrative that is given to the interaction of the causes, mechanisms, effects, and the cause–effect relationship. A causal story encompasses the respective elements, but is not

[1] These elements make up part of what is considered a scientific explanation. I do not address whether they are the entirety of a scientific explanation or merely a part of one. This is a large departure from some conceptions of explanations (e.g. Salmon 1994).

necessarily said to exist in the world. It is simply a constructed narrative of extant facts, assertions, and assumptions.

The first aspect of this model is that causes, mechanisms, and effects are distinct elements within a causal story. While possibly intuitive to many practicing scientists, the treatment of these elements in the philosophical literature has blended causes and mechanisms. If causes occur first (in time) and are responsible for bringing about an effect, it is reasonable to say that they produce the effect (being first in time of the causal story). Philosophical treatment of causes and mechanisms, however, can blend these together. Mechanisms have been argued to be what is responsible for an effect or phenomenon (Bechtel and Abrahamsen 2005; Illari and Williamson 2012) and have been argued to be what produce a behavior or effect (Machamer et al. 2000; Glennan 2002).[2] Such formulations leave little room for causes. In fact, some have questioned whether causes and causal knowledge serve any function in light of mechanistic knowledge (Williamson 2013). What is referred to as a cause in the scientific sense is only a cause of an effect in a derivative sense; that "It is not the penicillin that causes the pneumonia to disappear, but what the penicillin does" (Machamer et al. 2000, p. 6). Put differently, Y, (the presence of pneumonia) is not affected by X (the presence or absence of penicillin) but by the mechanism (what X does). This can easily lead to confusion, for example, in the job description of coroners.

The job of a coroner is to determine the cause of death. Death from a myocardial infarction, or self-inflected wrist cutting, or decapitation, all represent very different causes of death. They all, however, share the same mechanism—a lack of oxygenated blood flow to the brain. Under formulations of mechanism as causes, decapitation would only cause death 'in a derivative sense'. To reformulate the previous statement: It is not the decapitation that causes death, but what the decapitation does (severs the flow of oxygenated to the brain). Since most causes of death happen by way of a lack of oxygenated blood flow to the brain, such formulations of causes and mechanisms makes the job of the coroner exponentially simpler. It is always a lack of oxygenated blood flow to the brain. Therefore, we can see that in the course of defending the active processes of causes and the 'activities' of mechanisms, such focuses on mechanisms may be misconstrued as a commitment to the claim that the presence of the cause does not affect the outcome.

The second aspect of this theory is that causes, mechanisms, and effects occur over time. Causes are the first in the temporal chain, effects are the end of the temporal chain, and mechanisms occur in between the two. Causes come before mechanisms which also come before effects.[3]

The third aspect of this theory is that the cause–effect relationship is a distinct element from a causal story. The difference here has to do with realist interpretations. Cause–effect relationships are here believed to exist in the world. How we come to know these cause–effect relationships is part of the general practice of science. In contrast, causal stories are constructed to tie together causes, mechanisms, effects, and cause–effect relationships (when known) to form a narrative pulling together the elements into a consistent account. It

[2] One such definition, for example: "Mechanisms are entities and activities organized such that they are productive of regular changes from start or set-up to finish or termination conditions." (Machamer et al. 2000, p. 3). It is not just the parts, but the behavior (or activities) of those parts.

[3] There is a deeper question of the possibility of instantaneity in general that is beyond the scope of this paper, though I take it as given that two events cannot occur instantaneously at a near-infinitely small scale (e.g. defined as a plank time value of 0). Some amount of time must pass between the cause and the effect. Whether pure instantaneous causation is possible is a different question, altogether.

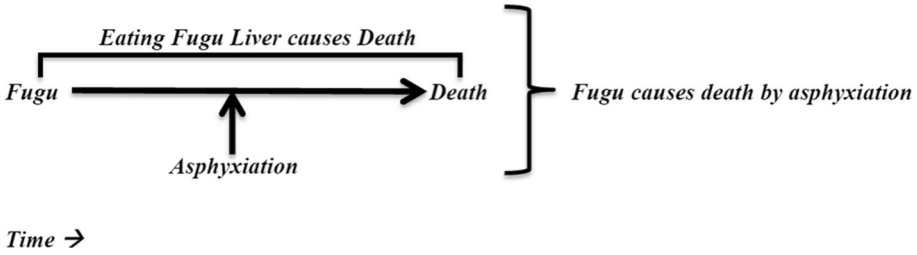

Fig. 2 Alternate causal story and mechanism

can therefore be said that the cause–effect relationship exists in the world while the causal story is a narrative construction.

3 Concrete Examples

In what follows I provide a concrete set of examples to help show how each of the elements (causes, effects, mechanisms, cause–effect relationships, and causal stories) play out given a set of facts.

Suppose we randomly assign a large number of people to two conditions, half of who eat Fugu (*Takifugu rubripes*) liver and the other half who eat Salmon (*Oncorhynchus gorbuscha)* liver (and we control for relevant differences in the two groups afterwards). We then observe that everybody who ate the Fugu liver died while nobody who ate the Salmon liver died. This would give us strong reason to believe that there exists a cause–effect relationship between eating Fugu liver and death. This justification rests on the counterfactual model of causality as applied to randomized controlled trials (e.g. Rubin 2005). It is important to point out that while this is one method of causal inference, the theory presented here need not require a randomized controlled trial (RCT).[4]

The effect under question is the deaths of the participants. The cause of death here is eating the liver of a Fugu fish. We create a causal story pulling these facts together by stating that Fugu liver is deadly [call this the "deadly fish" explanation]. This experiment allows us to establish the cause–effect relationship, but it does not give us the mechanism.[5] Not knowing how a cause brings about an effect, however, does not alter whether we know the truth of the cause–effect relationship. It simply has not yet been investigated.

Suppose we now want to know a mechanism—*how* eating Fugu liver brings about death. We send the 100 corpses to a pathologist who concludes these participants died from asphyxiation. We have the cause and the effect, as well as the cause–effect relationship; now, we also have a mechanism. This is schematized in Fig. 2.

[4] For example, the counterfactual model is logically equivalent to certain structural equation models (Galles and Pearl 1998). We could easily build a structural equation model with death as the outcome and what food eaten as exogenous variables with the errors uncorrelated (due to random assignment); this model would contain all of the relevant variables in our causal model and fulfill the requirement of causality (death responds to variations in the fish liver variable; Pearl 2009).

[5] In this manuscript I occasionally refer to *the* mechanism. This is not to suggest that only one mechanism may exist between a cause and effect. It is of course possible for a cause–effect relationship to follow along multiple mechanisms. Mechanisms occur in-between the cause and effect, the cause and effect are not part of the mechanism.

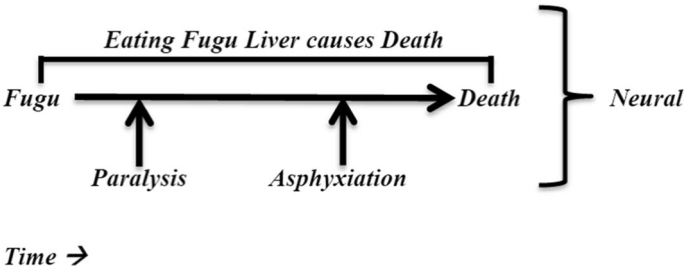

Fig. 3 Alternate causal story and mechanism II

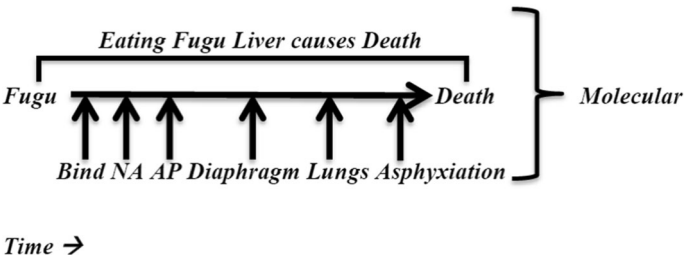

Fig. 4 Alternate causal story and mechanism III

Suppose we now want to know how Fugu liver causes death via asphyxiation (a mechanism of asphyxiation). It is possible all of those who dies of asphyxiation choked on the liver, rather than a chemical process. To rule this out, we run another randomized trial and observe that everyone in Fugu group became paralyzed, then asphyxiated, then died (nobody in the Salmon group did). We now have a better understanding of the phenomenon under investigation (see Fig. 3).

We also see how the more detailed mechanism is separated in time; paralysis occurs first, followed by asphyxiation. This will be referred to as the "neural" causal story.

Now suppose we want to know a mechanism of how eating Fugu liver leads to paralysis, asphyxiation, and death. Through a series of studies, we learn there is a chemical in the Fugu liver, tetrodotoxin. Tetrodotoxin binds to the sodium channels outside the nerve cell membranes, blocking the passage of sodium ions, inhibiting the generation of action potentials of the nerves, leading to the stopping of contractions of the diaphragm, leading to the lungs failing to operate, leading to asphyxiation and death (see Clark et al. 1999). We can call this the "molecular" causal story; its causal graph can be seen in Fig. 4.

We now have a much more in depth knowledge of how eating Fugu liver brings about death. This does not necessarily mean that knowledge of a mechanism at a more micro level is better. The level to which a mechanism is 'better' than a mechanism at another level more likely comes down to the level of analysis of the problem (Noble 2008).

It is my hope that these concrete examples help make explicit the argument I am forwarding. Causes, effects, and mechanisms are distinct elements within a causal story. Causes, mechanisms, and effects occur over time. The cause–effect relationship is a distinct element from a causal story. It is with this model of causes and mechanisms that I wish to explain the implications for scientific practice.

4 Implications for Scientific Practice

4.1 Implication 1: The Knowledge of How the Cause Works (Mechanisms) is not Necessary to Establish a Cause–Effect Relationship

In the special sciences (especially the Life Sciences), causality is among the most prominent of pursuits. Even simple correlations among variables often require a causal interpretation if an underlying latent variable is proposed (e.g. Borsboom et al. 2003). The most common method of establishing a causal relationship is through counterfactual dependence (e.g. Rubin 2005; Pearl 2014), though not the only way. The counterfactual system of discovering causality rests on the assumption of holding all variables fixed in their current state except the one under consideration (Pearl 2009; Woodward 2003). If this intervention on a single variable leads to a change in the outcome variable (DV), we have good reason to believe a cause–effect relationship exists between them. This counterfactual model of causal inference is the reason we perform RCTs (Rubin 2005), with the assumptions of randomization across groups met, we can be sure the difference between the groups are due to the manipulation.

It is important to note that nowhere in counterfactual dependence is establishing a mechanism required. The causal effect of X on Y can be established without any reference to *how* X causes Y; cause–effect relationships are aimed to investigate *whether* X causes Y. Therefore, there are ways, like RCTs, that can be shown to be good (at least when ideally carried out) at establishing cause–effect relationship without reference to how the cause brings about the effect. So cause–effect relationships can be discovered without the need of knowing a mechanism. This is equally true if we adopt a structural equation modelling (SEM) framework.[6]

Take the first example of a simple very well-conducted RCT only examining death from different fish livers. We randomly assigned a large number of people to eat Fugu liver or Salmon liver. The people who ate the Fugu liver died. According to the scientific approach to causal inference in the experimental sciences, we now have strong reason to believe that eating Fugu liver causes death. In that causal story we have no idea (yet) of a mechanism. That knowledge or lack of knowledge has no bearing on whether we know if the cause–effect relationship holds (cf. Russo and Williamson 2007). Knowledge of how the cause would have brought about the effect (had it brought it about) is not knowledge that it did bring it about. To show the full cause–mechanism-effect chain, each step in the chain must be independently investigated. It is not sufficient to argue upwards from mechanistic knowledge to infer the existence of the cause–effect relationship (Howick 2011) So what is different when we gain an understanding of a mechanism?

4.2 Implication 2: Changing the Knowledge of a Mechanism Changes the Causal Story but Does not Change our Knowledge of Whether a Cause–Effect Relationship Holds

How a cause brings about an effect (mechanism) plays an important role in scientific understanding. As I argued before, it need not play a role in our knowledge of a cause–effect relationship (see also Psillos 2004). Consider the four causal stories presented earlier: “deadly fish”, “asphyxiation”, “neural”, and “molecular”. In each of these instances

[6] We could add in mechanisms into our SEM (see Bollen and Pearl 2013), but they would not be necessary to establish the first step of causal inference.

we gained a better understanding of *how* the cause brings about the effect. In "asphyxiation" we understood how eating Fugu liver led to death. In "neural" we understood that it was not because all of our participants choked on the fish liver, but instead became paralyzed and were unable to breathe. In "molecular" we understood a mechanism of paralysis and how the Fugu liver affects the body at a molecular level. Notice that all of these events occur distributed across time, all starting with the eating of the toxic fish.

In each of these four stories the mechanism changed and as the mechanism changed, so did the causal story. What did not change, however, was the cause–effect relationship. In each of these situations, the cause–effect relationship, 'eating Fugu liver causes death', was preserved. Since the inference of a causal connection from counterfactual dependence does not require a mechanism, altering the knowledge of how the cause would bring about the effect has no bearing on our knowledge of the existence of that causal dependence.

What happens when we are wrong about a mechanism?

4.3 Implication 3: The Knowledge of a Proposed Mechanism Between Two Variables has no Bearing on our Knowledge of Whether the Cause–Effect Relationship Holds—It Only Affects the Causal Story

To illustrate this implication, I will introduce a final causal story. Suppose we again randomly assign 200 people to eat Fugu liver or Salmon liver; the 100 who ate Fugu liver all drop dead. Let's now say that we posit the following mechanism: eating Fugu liver leads to death because the liver of that fish is sacred to the Gods and they punish those who flout their will. As per this model and implication 2 we have the same cause, the same effect, and the same cause–effect relationship. We now have a new mechanism and overall a new causal story: "sacred fish" (the Fugu liver is sacred and the Gods strike down those who eat it). This is not such an absurd mechanism or causal story at that; indeed, the Abrahamic religious tradition of eschewing pork and shellfish may be traced back to self-preservation from eating spoiled meat.

It turns out, unfortunately, that this mechanism is incorrect. Let us assume that this mechanism was incorrect because it was based on a mistranslation of the sacred text (the liver of a different fish is sacred). This new information falsifies our mechanism; it also falsifies our causal story (since the liver is no longer sacred we cannot tell a causal story recounting this information). The cause–effect relationship, however, remains untouched. Our 100 people who ate Fugu will not come back to life because the proposed mechanism was false. The truth of a mechanism can only affect the causal story, not the cause–effect relationship. So overall we can see that our knowledge of an incorrect mechanism cannot negate our knowledge of the truth of a cause–effect relationship.

5 Two Classes of Errors

In this section I will introduce two classes of errors that occur both in the scientific literature and in daily life: faulty causal exclusion and faulty causal inference. I argue that understanding the model presented here would help prevent them from occurring. The errors are not meant to highlight *that* we are sometimes wrong in scientific endeavors, but *why* we are sometimes wrong.

5.1 Faulty Causal Exclusion

Faulty causal exclusion occurs when the existence of a cause–effect relationship is argued to be false from either not knowing or being wrong about the mechanism. Many practicing scientists have encountered such an argument in their own work from a reviewer who insists on a mechanism or argues the falsity of a finding due to mechanistic arguments. As explained before, the existence of, or truth of, a mechanism has no effect on what we believe about the truth of the cause–effect relationship. As a brief historical example, this error occurred when a nineteenth century doctor first proposed and experimented with doctors washing their hands.

Ignaz Semmelweis was a medical doctor who first proposed that doctors should wash their hands before delivering babies. This was in response to the observation that women and children were more likely to contract and die from puerperal fever if they gave birth in a hospital than if they gave birth to midwives (Wykticky and Skopec 1983). Semmelweis then experimented with chlorine and hand washing to reduce puerperal fever. After multiple introductions of such policies and dramatic decreases in puerperal fever (Routh 1849), the scientific community should have had reason to believe that a cause–effect relationship existed between hand washing with chlorine and puerperal fever.

Semmelweis' work, however, was largely ignored or ridiculed. Among the many sociological reasons for rejecting his findings, an additional reason given was that there was no proposed mechanism between chlorinated hand washing and disease (Hauzman 2006). It was not until the work of Pasteur and Lister that a mechanism given for Semmelweis' findings. Although I do not claim that understanding the model here would prevent all sociological rejections of scientific work, understanding that mechanism does not play a part in the establishment or truth of cause–effect relationships removes one source of scientific justification for those sociological drives. If causal evidence is judged on the merit of causal inference and not on mechanistic knowledge, such errors of faulty causal exclusion may be avoided.

5.2 Faulty Causal Inference

More insidious than excluding a cause–effect relationship is the faulty generation of them. This error occurs when a cause–effect relationship is argued to exist based on knowledge of how the cause would bring about the effect (mechanism) and a convincing causal story. Another historical example of these errors comes from the use of class 1c anti-arrhythmia drugs after a myocardial infarction.

Since the 1950s, class 1c anti-arrhythmia drugs were given to patients after a myocardial infarction (heart attack; MI) to prevent tachycardia and death during recovery. How these drugs worked was well known; the drugs stabilized sodium channels inside the cells of the heart and normalized heart rhythm (Milne et al. 1984). Since recovery after a MI entails the possibility of the heart to become tachycardic—combatting this with administration of anti-arrhythmia drugs was extensively used. The causal story was established (preventing tachycardia with anti-arrhythmia drugs after a MI saves lives) and justified on the basis of the activities of the drugs.

The error, however, was in presuming the truth of the cause–effect relationship from the knowledge of how the drugs worked and a plausible causal story. It wasn't until the 1980s that the cause–effect relationship between anti-arrhythmia drugs and survival after a heart attack was actually tested. Over 1700 participants were randomly assigned in a double-

blind, placebo-controlled manner to test this cause–effect relationship. The study had to be stopped when it was discovered that anti-arrhythmia drugs actually increased death during recovery from MI (CAS Trial 1989). This trial was later replicated and again had to be stopped as well due to the increase in deaths (CAS Trial and Investigators 1992).

The true cause–effect relationship was that administering such drugs after a MI increased deaths. The opposite cause–effect relationship was inferred, knowing how anti-arrhythmia drugs worked and a convincing causal story. However, as I have argued, a causal story is simply a narrative of facts and does not necessarily reflect the truth of the world.

What happened in the CAS trial and the decades leading up to it? The activities and operations of class 1c anti-arrhythmia drugs were known (stabilizing sodium channels inside the cell walls of the heart). Yet when put into practice the cause–effect relationship directly implied by the activities and operations of the drug failed to hold and in fact, were reversed.[7] Remember in this example the cause was taking the class Ic anti-arrhythmia drugs, the mechanism was the activities of the drugs on the sodium channels, and the proposed effects was preventing death through preventing tachycardia. The mechanism was known, yet the faulty causal inference was to believe that knowledge of the mechanism means the cause–effect relationship holds (in this example, that taking the drugs prevented tachycardia). The problem is, while class Ic anti-arrhythmia drugs do block sodium channels inside heart cell walls, they also *increase* the time when refractoriness and non-refractoriness overlap, increasing the probability that an odd wavelet can turn into re-entrant arrhythmia (Starmer 2002; Xia et al. 2009). The paradox is that they both prevent and cause tachycardia.[8] Such an outcome was not known for decades because the cause–effect relationship, that taking the drugs increased death, was inferred from the operations and activities of the drugs; it was not directly tested.

This presumption of cause–effect relationships from knowledge of how the drugs worked and a convincing causal story is a dangerous outcome that is not easily explained by existing theories of mechanisms. Tens of thousands of people are estimated to have died from this presumption of a cause–effect relationship with anti-arrhythmia drugs (Silverman 1997). Under the model presented here, we can see that mechanisms are a distinct element of the cause–effect relationship and cannot attest to its truth or our knowledge of its truth.[9]

Another recurring example of faulty causal inference is whether eating sugar affects children's behavior or cognition. The answer to this question, based off of 23 randomized controlled trials, is no; sugar consumption does not affect acting out behavior, sugar consumption does not affect attention, sugar consumption does not affect mood, it does not affect motor skills, levels of aggression, or academic achievement (Wolraich et al. 1995).

But now let us propose a mechanism for why sugar consumption actually does affect behavior, despite the evidence. We can argue that "a diet high in added sugar reduces the production of a brain chemical known as brain-derived neurotrophic factor (BDNF);

[7] The fact that the mechanism of the drug was known, yet it acted contrary to its mechanism has bene referred to as Sanderson' paradox (Sanderson 1996; Starmer 2002). This is also similar to conditional causal effects, discussed below and in Hesslow (1976).

[8] The difference comes down to what the drugs due at the singular versus the multicellular level.

[9] It has been argued that the statistical observation of the path through which one variable affects another can be used to infer cause-effect relationships (Pearl 2009). The statistical and observational requirements of this *front-door* approach are, however, considered to be unrealizable in real data (see Morgan and Winship 2007; Bullock et al. 2010).

without BDNF, our brains can't form new memories and we can't learn (or remember) much of anything (Molteni et al. 2002)."[10] This type of research is a more recent example of faulty causal inference. The evidence from *how* sugar would affect memory and learning is used to infer *that* consuming sugar affects memory and learning, despite the experimental evidence that it does no such thing.

There is another danger in causal stories: they are incredibly simple to formulate. When you tell people that the most successful firefighters are those who make the riskiest choices, they can immediately construct a causal story[11] about how the successful firefighter of course braves the risks to save people from a burning building, while the unsuccessful firefighter plays it safe and stands by as people die (Anderson et al. 1980). The problem is, if you tell people that the most successful firefighters are those who make the most conservative and least risky choices, they have no trouble coming up with an equally plausible causal story: "the successful firefighter carefully weights the relevant risks before taking appropriate and decisive action, while the unsuccessful trainee plunges headlong into danger, risking both his own and others' lives by foolhardy actions" (Anderson et al. 1980, p. 1047).

What is dangerous is that people do this spontaneously (e.g. Anderson 1983) and when we construct these causal stories, the beliefs remain even after the initial evidence is discredited. People who construct the causal story of the successful risky firefighter continue to believe that cause–effect relationship exists even after you tell them there is no such evidence and was simply constructed for a psychological experiment; those who construct the causal story of the successful conservative firefighter believe that cause–effect relationship instead (Anderson et al. 1980; Anderson 1983; Anderson and Sechler 1986).

5.3 Another Possible Error

There is another situation which arises that I only touch on here. When the same cause brings about the same effect, but it does so through a different mechanism, can it be considered the same cause–effect relationship? Turning the key in a combustion engine car starts it through process x. Turning the key in an electric car starts it through process y. They have different mechanisms but the same cause–effect relationships (turning the key starts the car). It would be mistaken to believe that because they have different mechanisms, the causes and cause–effect relationships are not the same or cannot be equated. We may be tempted to conclude that turning a car's key starts the car (same cause–effect relationship). This can get us into tricky situations, however. Increasing load (cause) on the muscular system can produce increases in proliferation of system-specific cells (effect). Increasing load on the brain (by introduction of new skills) causes increases in proliferations of system-specific cells (generation of new neurons, see Draganski et al. 2004). The mechanisms, however, are radically different. It is unclear whether we can say that these two systems have the same cause–effect relationships. This could also be a difference between standard conversational use and scientific use of the terms. We do not delve into

[10] https://www.psychologytoday.com/blog/neuronarrative/201204/what-eating-too-much-sugar-does-your-brain. Also the book by Perlmutter (2013).

[11] This line of research is called the explanation effect; however, since the use of the term 'explanation' in the psychological literature is so different than the use of 'explanation' in philosophy, we avoid the term explanation effect for simplicity.

the topic here but raise it as a possible error of excluding the same cause–effect relationships because of different mechanisms.

6 Counterarguments/Problems

I have presented what I believe to be the correct understanding of the relationships between causes, mechanisms, effects, cause–effect relationships, and causal stories. This model carries with it certain implications for the implementation of science that have gone overlooked or misunderstood. The model also corrects for two classes of errors that are not explained by other theories of causes and mechanisms[12]; provided we do not hold the belief that mechanism needs to be a physical process (as do Salmon 1998; Glennan 2002).

6.1 Probabilistic Causation

In the examples of casual stories presented above we assumed determinism between the cause and effect for simplicity. Probabilistic causality, presents another side to causal inference. Probabilistic causality can be separated into metaphysical probabilistic causation (MPC) and epistemic probabilistic causation (EPC). Metaphysical probabilistic causation refers to an action having a *possible* effect. Thus, the effects under question are probabilistic in the sense that, given knowledge about all possible antecedents, the outcome is still a probability. Certain functions of quantum mechanics fall into this class, so may be whether a cell becomes cancerous. This can be contrasted with EPC, which would be that our *knowledge* of an event is a probability. This can occur because we do not know all of the antecedents of an event, or that the effect under question is a conditional causal effect and we do not know the key moderating instances. Most research involving living organisms and their behavior fall into EPC. The outcome of a coin flip, for example, is an instance of EPC. The outcome is determined given numerous conditions, such as size and shape of the coin, force and angle applied to the flip, gravitational pull at the location of the toss (Diaconis et al. 2007). When we say the probability of heads is .5, what we mean is the probability of heads *given no additional information* is .5 ($p(\text{Heads}|\emptyset) = .5$).

In the examples used in this manuscript, we assume EPC. The processes under investigation are likely to be deterministic with unknown antecedents and conditional causal properties. This lack of perfect information can render the investigation probabilistic, despite the underlying processes being deterministic (see Hesslow 1981, for a similar argument). In cases such as this, the above arguments do not change at all. It simply means that more evidence is required to understand the underlying cause–effect relationship due to the extent of the uncertainties.

Therefore, we do not see any changes whether the epistemic investigation is deterministic or probabilistic. Those who eat the Fugu liver and do not die may have a genetic protective factor or be part of a population who has an immunity (moderated effect or part of a conditional causal claim), may have been assigned to eat the liver but did not actually eat the liver (non-complier), or may simply have gotten very sick but did not die (possible individual differences). In EPC, all of these issues can arise without falsifying the cause–effect relationship until further investigation is carried out. If the reason an individual did not die is because of individual differences or a conditional causal claim, knowing about

[12] I do not address the problem of causes without mechanisms because under this model, it does not present a problem, there is no such thing.

the cellular effect will not change our knowledge of the cause–effect relationship as, by definition, the mechanism does not bring about the effect in the conditional causal situation where the condition does not hold.

7 Conclusion

The model presented here disentangles causes from mechanisms and places them in a framework that is able to take into account narrative explanations while keeping a realist scientific interpretation of causes, mechanisms, and effects. I argue that there needs to be an acknowledgement of the separation of causes, mechanisms, effects, cause–effect relationships, and causal stories. This model carries with it three implications for the practice of science: (1) knowledge of how a cause would bring about an effect (mechanisms) is not necessary to establish a cause–effect relationship; (2) changing the knowledge of the mechanism changes the causal story but does not change the cause–effect relationship; and (3) the falsity of the mechanism cannot invalidate the truth of the cause–effect relationship, only the causal story. In addition, knowledge of how a cause would bring about an effect and a causal story presume, but do not demonstrate an existence of a cause–effect relationship. The reason we make this presumption comes from the strength of the causal story (see Taleb 2010; the narrative fallacy); stories are the way human beings naturally think, and we force facts into a narrative to make sense of them. To be confident that a causal story is accurate to the world, not only must the cause–effect relationship be demonstrated, but a causal connection to mechanisms must also be demonstrated. Put succinctly:

> establishing the causal status of each part of a mechanism would require finding out (or estimating) its causal effect… The causal effect can be found out, at least in favorable circumstances, *without* understanding the causal mechanisms, if any, involved; but the causal mechanisms, even if they are present, cannot be understood without the notion of the causal effect, that is without some notion of (counterfactual) dependence. (Psillos 2004, pp. 315–316).

Our ability to construct causal stories from imperfect information is not only astounding, but can cause serious scientific errors.

Acknowledgements I would like to thank Nancy Cartwright, Mordechai Juni, Taraz Lee, Andrew Maul, Rebecca Schaefer, and Margaret Tampari for their thoughts on this argument; as well as Rogier Kievit and the two anonymous reviewers for their helpful comments on improving this paper.

References

Anderson, C. A. (1983). Abstract and concrete data in the perseverance of social theories: When weak data lead to unshakeable beliefs. *Journal of Experimental Social Psychology, 19*(2), 93–108.

Anderson, C. A., Lepper, M. R., & Ross, L. (1980). Perseverance of social theories: The role of explanation in the persistence of discredited information. *Journal of Personality and Social Psychology, 39*(6), 1037–1049.

Anderson, C. A., & Sechler, E. S. (1986). Effects of explanation and counterexplanation on the development and use of social theories. *Journal of Personality and Social Psychology, 50*(1), 24–34.

Bechtel, W., & Abrahamsen, A. (2005). Explanation: A mechanist alternative. *Studies in History and Philosophy of Science Part C: Studies in History and Philosophy of Biological and Biomedical Sciences, 36*(2), 421–441.

Bollen, K. A., & Pearl, J. (2013). Eight myths about causality and structural equation models. In S. L. Morgan (Ed.), *Handbook of causal analysis for social research* (pp. 301–328). New York: Springer.
Borsboom, D., Mellenbergh, G. J., & Van Heerden, J. (2003). The theoretical status of latent variables. *Psychological Review, 110*(2), 203–219.
Bullock, J. G., Green, D. P., & Ha, S. E. (2010). Yes, but what's the mechanism? (don't expect an easy answer). *Journal of Personality and Social Psychology, 98*(4), 550–558.
Clark, R. F., Williams, S. R., Nordt, S. P., & Manoguerra, A. S. (1999). A review of selected seafood poisonings. *Undersea Hyperbolic Medicine, 26*(3), 175–184.
Diaconis, P., Holmes, S., & Montgomery, R. (2007). Dynamical bias in the coin toss. *SIAM Review, 49*(2), 211–235.
Draganski, B., Gaser, C., Busch, V., Schuierer, G., Bogdahn, U., & May, A. (2004). Neuroplasticity: Changes in grey matter induced by training. *Nature, 427*(6972), 311–312.
Galles, D., & Pearl, J. (1998). An axiomatic characterization of causal counterfactuals. *Foundations of Science, 3*(1), 151–182.
Glennan, S. (2002). Rethinking mechanistic explanation. *Philosophy of Science, 69*(S3), S342–S353.
Hauzman, E. E. (2006). Semmelweis and his German contemporaries. In *40th International congress on the history of medicine*.
Hesslow, G. (1976). Two notes on the probabilistic approach to causality. *Philosophy of Science, 43,* 290–292.
Hesslow, G. (1981). Causality and determinism. *Philosophy of Science, 48*(4), 591–605.
Howick, J. (2011). Exposing the vanities—and a qualified defense—of mechanistic reasoning in health care decision making. *Philosophy of Science, 78*(5), 926–940.
Illari, P. M., & Williamson, J. (2012). What is a mechanism? Thinking about mechanisms across the sciences. *European Journal for Philosophy of Science, 2*(1), 119–135.
Koslowski, B. (1996). *Theory and evidence: The development of scientific reasoning*. Mit Press.
Machamer, P., Darden, L., & Craver, C. F. (2000). Thinking about mechanisms. *Philosophy of Science, 67*(1), 1–25.
Mackie, J. L. (1974). *The cement of the universe: A study of causation*. Oxford: Clarendon.
Milne, J. R., Hellestrand, K. J., Bexton, R. S., Burnett, P. J., Debbas, N. M. G., & Camm, A. J. (1984). Class 1 antiarrhythmic drugs-characteristic electrocardiographs differences when assessed by atrial and ventricular pacing. *European Heart Journal, 5*(2), 99–107.
Molteni, R., Barnard, R. J., Ying, Z., Roberts, C. K., & Gomez-Pinilla, F. (2002). A high-fat, refined sugar diet reduces hippocampal brain-derived neurotrophic factor, neuronal plasticity, and learning. *Neuroscience, 112*(4), 803–814.
Morgan, S. L., & Winship, C. (2007). *Counterfactuals and causal inference: Methods and principles for social research*. Cambridge: Cambridge University Press.
Noble, D. (2008). *The music of life: Biology beyond genes*. Oxford: Oxford University Press.
Pearl, J. (2009). *Causality: Models, reasoning and inference* (2nd ed.). Cambridge: MIT Press.
Pearl, J. (2014). *Causes of effects, from philosophy to science*. Tech. Rep.
Perlmutter, D. (2013). *Grain brain: The surprising truth about wheat, carbs, and sugar–your brain's silent killers*. Hachette: Little, Brown and Company.
Psillos, S. (2004). A glimpse of the *secret connexion*: Harmonizing mechanisms with counterfactuals. *Perspectives on Science, 12*(3), 288–319.
Rouder, J. N., & Morey, R. D. (2011). A Bayes factor meta-analysis of Bem's ESP claim. *Psychonomic Bulletin & Review, 18*(4), 682–689.
Rouder, J. N., Morey, R. D., & Province, J. M. (2013). A Bayes factor meta-analysis of recent extrasensory perception experiments: Comment on Storm, Tressoldi, and Di Risio (2010). *Psychological Bulletin, 139*(1), 241–247.
Routh, C. H. F. (1849). On the causes of the endemic puerperal fever of Vienna. *Medico-chirurgical Transactions, 32,* 27.
Rubin, D. B. (2005). Causal inference using potential outcomes. *Journal of the American Statistical Association, 100*(469), 322–331.
Russo, F., & Williamson, J. (2007). Interpreting causality in the health sciences. *International Studies in the Philosophy of Science, 21*(2), 1157–1170.
Salmon, W. C. (1994). Causality without counterfactuals. *Philosophy of Science, 61,* 297–312.
Salmon, W. C. (1998). *Causality and explanation*. Oxford: Oxford University Press.
Sanderson, J. (1996). The sword of Damocles. *Lancet, 348,* 2–3.
Shipstead, Z., Redick, T. S., & Engle, R. W. (2012). Is working memory training effective? *Psychological Bulletin, 138*(4), 628–654.

Silverman, W. (1997). Equitable distribution of the risks and benefits associated with medical innovations. In A. Maynard & I. Chalmers (Eds.), *Non-random reflections on health services research*. London: BMJ.

Starmer, C. F. (2002). How antiarrhythmic drugs increase the rate of sudden cardiac death. *International Journal of Bifurcation and Chaos, 12*(09), 1953–1968.

Taleb, N. N. (2010). *The black swan: The impact of the highly improbable fragility*. New York, NY: Random House.

Trial, C. A. S. (1989). Investigators. Preliminary report: Effect of encainide and flecainide on mortality in a randomized trial of arrhythmia suppression after myocardial infarction. *New England Journal of Medicine, 321*(6), 406–412.

Trial, C. A. S., & II Investigators. (1992). Effect of the antiarrhythmic agent moricizine on survival after myocardial infarction. *New England Journal of Medicine, 327*(4), 227–233.

Williamson, J. (2013). How can causal explanations explain? *Erkenntnis, 78*(2), 257–275.

Wolraich, M. L., Wilson, D. B., & White, J. W. (1995). The effect of sugar on behavior or cognition in children. A meta-analysis. *JAMA, 274*(20), 1617–1621.

Woodward, J. (2003). *Making things happen: A theory of causal explanation*. Oxford: Oxford University Press.

Wykticky, H., & Skopec, M. (1983). Ignaz Philipp Semmelweis, the prophet ofbacteriology. *Infection Control, 4*(5), 367–370.

Xia, H., Zhao, X., Bains, J., & Wortham, D. C. (2009). Influence of channel blockers on cardiac alternans. In *2009 Annual international conference of the IEEE engineering in medicine and biology society* (pp. 2823–2826). IEEE.

John Protzko is a Cognitive Scientist who studies the development of human intelligence and the biased interpretation of scientific evidence (sometimes inclusive).