Manipulation is key – On why non-mechanistic explanations in the cognitive sciences also describe relations of manipulation and control

Lotem Elber-Dorozko

The Edmond & Lily Safra Center for Brain Sciences, The Hebrew University of Jerusalem.

To contact the author: lotem.elber@mail.huji.ac.il

This is a post-peer-review, pre-copyedit version of an article published in *Synthese*. The final authenticated version is available online at: https://doi.org/10.1007/s11229-018-01901-3

Abstract

A popular view presents explanations in the cognitive sciences as causal or mechanistic and argues that an important feature of such explanations is that they allow us to manipulate and control the explanandum phenomena. Nonetheless, whether there can be explanations in the cognitive sciences that are neither causal nor mechanistic is still under debate. Another prominent view suggests that both causal and non-causal relations of counterfactual dependence can be explanatory, but this view is open to the criticism that it is not clear how to distinguish explanatory from non-explanatory relations. In this paper, I draw from both views and suggest that, in the cognitive sciences, relations of counterfactual dependence that allow manipulation and control can be explanatory even when they are neither causal nor mechanistic. Furthermore, the ability to allow manipulation can determine whether non-causal counterfactual dependence relations are explanatory. I present a preliminary framework for manipulation relations that includes some non-causal relations and use two examples from the cognitive sciences to show how this framework distinguishes between explanatory and non-explanatory, non-causal relations. The proposed framework suggests that, in the cognitive sciences, causal and non-causal relations have the same criterion for explanatory value, namely, whether or not they allow manipulation and control.

Keywords

explanation; non-causal explanations; manipulation and control; the cognitive sciences; counterfactual dependence;

1 Introduction

Philosophers have characterized various types of explanations in the cognitive sciences. Functional analyses (Cummins 1983, 2000), mechanistic models (Craver 2007a), computational models (Chirimuuta 2014; Egan 2017; Rusanen and Lappi 2016; Shagrir 2006; Shagrir and Bechtel 2017) as well as network, topological, and mathematical models (Chirimuuta 2017; Huneman 2010; Silberstein and Chemero 2013) have all been said to have explanatory value. This poses a challenge to philosophers – how does one present a framework for explanation in the cognitive sciences, when said explanation is so deeply diverse in range[1]?

One prominent - albeit highly contended - view is the mechanistic view of explanations in the cognitive sciences. According to proponents of this view (henceforth: "mechanists") (Craver 2007a, 2007b, 2016, Kaplan 2011, 2017; Kaplan and Craver 2011; Milkowski 2013; Piccinini 2015; Piccinini and Craver 2011), generally, models in the cognitive sciences are explanatory to the extent that they describe relevant causal structures. These relevant causal structures are those "that produce, underlie, or maintain the explanandum phenomenon" (Kaplan and Craver 2011, p. 602). On this view, explanations in the cognitive sciences are often mechanistic - the phenomenon is explained by appeal to its underlying causal structure, a mechanism. The appeal of this view is strong: it implies that many explanations in the cognitive sciences

[1] One can also be a pluralist and argue that there is no single, unifying framework that can accommodate all these explanations. In this paper, I assume that, were it to be possible, such a framework would be preferable.

have a unifying feature, namely, the description of relevant causal structures. Nonetheless, in recent years, mechanists have had to defend their view against claims that some models in the cognitive sciences explain phenomena in ways that are outside the scope of the mechanistic framework (Bechtel and Shagrir 2015; Chirimuuta 2014; Egan 2017; Huneman 2010; Rusanen and Lappi 2016; Shagrir and Bechtel 2017; Shapiro 2017; Silberstein and Chemero 2013). The mechanists reply that these models are either explanatory because they describe relevant causal structures or they are not explanatory at all (Craver 2016; Kaplan 2011, 2017; Kaplan and Craver 2011; Piccinini and Craver 2011). This debate is still ongoing. Furthermore, the mechanistic view has been criticized on the grounds that it diminishes the explanatory value of non-mechanistic models such as functional analyses (Shapiro 2017) and computational models (Shagrir and Bechtel 2017).

Another approach, which is geared towards scientific explanation in general, is the counterfactual-dependence view of explanation. Woodward and Hitchcock (Hitchcock and Woodward 2003; Woodward 2003; Woodward and Hitchcock 2003) suggest that explanations provide the resources for answering a variety of what-if-things-had-been-different questions. The counterfactuals implied in these questions are described by appeal to intervention, a procedure formally and extensively set forth in (Woodward 2003) as part of an account of causal relations. Many mechanists also adopt Woodward's framework for causal relations. Diverging from the mechanists who focus on the explanatory value of causal relations, several philosophers have extended this framework and asserted that explanations reveal relations of counterfactual dependence more generally, so that some explanatory counterfactuals cannot be described as the result of interventions. In this way, non-causal counterfactual dependences, too, can be taken as explanatory (Baron et al. 2017; Bokulich 2011; Chirimuuta 2017; Jansson 2015; Jansson and Saatsi 2017; Pexton 2016; Reutlinger 2016; Saatsi and Pexton 2013; Woodward 2018; Ylikoski and Kuorikoski 2010).

This approach has promise, but it faces a challenge confronted by many frameworks that describe non-causal relations as explanatory: some counterfactual dependence relations are symmetrical (e.g., mathematical relations), yet in many of those cases, only one direction of dependence is taken to be explanatory (Craver 2016; Craver and Povich 2017)[2].

In this paper, I take a different route and combine an important feature of the mechanistic framework with the notion that non-causal dependences can also be explanatory. Mechanists often trace the initial interest in mechanistic explanations in the cognitive sciences to a desire to manipulate and control[3] neural and cognitive phenomena (in this, they follow Woodward, who makes a similar argument for causal explanation in general (2003)). Here, I suggest that some relations can allow manipulation of cognitive and neural phenomena even when these relations are not part of causal structures that produce or underlie these phenomena (henceforth, "non-causal relations"). Therefore, the motivation to manipulate cognitive and neural phenomena can be extended to account for the explanatory value of some dependence relations that do not comply with mechanistic requirements. Moreover, I argue that a framework that links explanation with the motivation to manipulate phenomena can account for some of our intuitions about the type of non-causal counterfactual dependences that are explanatory in the cognitive sciences[4].

---

[2] This issue has been addressed in several papers that develop such frameworks. (Saatsi and Pexton 2013) reply that the explanation of regularities, rather than a singular event, can be symmetrical, and therefore non-causal. For example, the fact that the length of pendulums is proportional to the gravitational field can be explained by the mathematical equation that relates the two. (Jansson 2015; Jansson and Saatsi 2017) describe specific dependence or determination relations and argue that they are not symmetrical.

[3] Throughout the paper, I treat 'manipulation' and 'control' as having the same meaning in this context. They are often found together in the literature. To avoid redundancy, generally, I will only speak of manipulation.

[4] Although they do not discuss manipulation and control directly, some of the studies that address the issue of the asymmetry of explanation in symmetrical dependence relations suggest solutions that seem consistent with this idea. (Woodward 2018) proposes, when describing one example, that if one side of a dependence relation can be explained by other ordinary causes, the direction of explanation runs from this side to the other. (Jansson and Saatsi 2017), for their part, claim that, in some mathematical relations, the dependence runs only in one direction when fixing a value of one variable determines the value of the other, but not vice versa.

This suggestion can contribute to both frameworks. Regarding the counterfactual dependence view, associating explanation with manipulation provides a way to distinguish explanatory from non-explanatory counterfactual dependences that is applicable to both causal and non-causal dependences in the cognitive sciences. In the future, this suggestion can be extended to other fields. Regarding the mechanistic framework, the point that non-causal dependences can also allow manipulation of the explanandum may be a good reason to extend this framework to include some explanatory dependences that are not causal or mechanistic.

In this paper, I analyze two examples of mathematical relations in the cognitive sciences, aiming to show that they allow manipulation in the direction of explanation. In the first example, the fact that an estimator that combines inputs from two modalities is optimal is explained by the statistics of the inputs (Ernst and Banks 2002). In the second example, the magnitude of fluctuations in the input to a neuron is explained by the ratio between its excitatory and inhibitory incoming inputs (Softky and Koch 1993; van Vreeswijk and Sompolinsky 1996).

A variety of models in the cognitive sciences have already been presented as explanatory despite the fact that they do not satisfy the mechanistic requirement of describing relevant causal structures (Chirimuuta 2014; Egan 2017; Huneman 2010; Rusanen and Lappi 2016; Silberstein and Chemero 2013). Unlike these studies, I do not aim to argue that some explanations in the cognitive sciences are non-causal. According to some manipulability frameworks for causation, relations of manipulability simply are relations of causal dependence (Woodward 2003). Proponents of such views may interpret the argument of this paper as showing that some dependence relations that were previously taken to be non-causal allow manipulation and therefore are, in fact, causal. Those who adopt such a view of causation for the examples presented here will have to concede that cause and effect can be mathematically related and occur simultaneously. Furthermore, if they accept that constitutive relations in mechanisms allow manipulation, then they must take constitutive relations to be causal. Such consequences are usually understood as undesirable

(Baumgartner and Gebharter 2016; Craver and Bechtel 2007; Romero 2015). Nonetheless, an interpretation of this paper that takes mathematical relations to be causal is possible, and I will not argue against it here.

The paper is organized as follows: section 2 will describe the role of manipulation and control in Woodward's framework and in the mechanistic framework. Section 3 will present a preliminary formulation of a manipulation relation that can accommodate causal and non-causal relations. Section 4 will provide two examples of non-causal explanations in the cognitive sciences that describe relations that allow manipulation. Finally, section 5 will discuss a few possible objections and counter-objections to the proposed framework.

2 manipulation and control in Woodward's and the mechanists' writings

Woodward develops an 'interventionist' or 'manipulationist' framework for causal relations and explanation, which is based on the notion that causal relations can potentially be used to manipulate the environment. He writes: "…our interest in causal relationships and explanation initially grows out of a highly practical interest human beings have in manipulation and control" (2003, p. 10) and states the following conditions for *X* to be a cause of *Y*:

> (M) *X* causes *Y* if and only if there are background circumstances *B* such that if some (single) intervention that changes the value of *X* (and no other variable) were to occur in *B*, then *Y* or the probability distribution of *Y* would change…An *intervention* on *X* with respect to *Y* [is] an idealized experimental manipulation of *X* which causes a change in *Y* that is of such a character that any change in *Y* occurs only through this change in *X* and not in any other way (Woodward 2010, p.4, italics in the original; for a more detailed description see Woodward 2003)

Craver, developing a framework for mechanistic explanation, writes: “Explanations in neuroscience are motivated fundamentally by the desire to bring the CNS [central nervous system] under our control.” (Craver 2007a, p. 160) . Building on Woodward’s framework, he states that a component is relevant to the behavior of a mechanism when “the two are related as part to whole and they are *mutually manipulable*” (2007a, p. 153 italics in the original).

Woodward (2003) and Craver (2007a) describe causal and mechanistic explanations, respectively, in terms of manipulability. I draw on this work and take causal relations and constitutive relations in mechanisms to allow manipulation. There is also a second, weaker, sense in which Woodward and Craver tie together explanation and manipulation - namely, both present explanation as motivated by the desire to be able to manipulate and control phenomena. This relation between manipulation and explanation has been echoed in other philosophical (Dretske 1994) and scientific (Lazebnik 2002) writings.

Inspired by this suggestion, I continue by arguing that, in the cognitive sciences, there are explanatory counterfactual dependence relations that allow manipulation of the explanandum and are neither causal nor mechanistic. Therefore, it may be possible to treat all these manipulation-allowing relations similarly, forming a more unified framework for explanation in the cognitive sciences. I begin by presenting a framework for manipulation.

3 Relations of manipulation and control (manipulation*) as explanatory relations

Ideally, I would use Woodward’s (2003) interventionist framework to describe manipulation relations. However, such a framework might not be able to accommodate non-causal dependence relations. In the case of constitutive relations, it is argued that ideal interventions on the part with respect to the whole, and vice versa, are not possible. In an ideal intervention, according to Woodward (2003), the intervention variable that changes *X* must not be a cause of *Y* through a path that does not include *X*. Arguably, however, any manipulation of the part can also be considered a direct manipulation of the whole, and vice versa, thus ruling out the possibility of an ideal intervention (Romero 2015). Similar claims can be made regarding supervenience, mathematical and other dependence relations in which the variables cannot be considered distinct.

Therefore, I suggest a slightly different account that is intended to also fit cases where the variables are not distinct. To differentiate this extended manipulation from that of Woodward, I term it manipulation*.

Take two non-identical variables, X and Y. Then Y can be manipulated* through X iff:

(1) There is at least one manipulation* variable $\mathbf{M}$ that can be used to manipulate* X[5]. So that in the counterfactual scenario in which $\mathbf{M}$ is used to change the value of X, while all variables are held constant except for {$\mathbf{M}$, X, Y, the variables on the path from $\mathbf{M}$ to X, and the variables that are manipulated through X}, the value of Y changes as well.

(2) The influence of $\mathbf{M}$ on the value of Y is completely mediated through X: if $\mathbf{M}$ is used to manipulate X as in (1), while any other manipulation* variable M is used to keep X constant and all variables are held constant except for {$\mathbf{M}$, M, Y, the variables on the path from M to X, the variables on the path from $\mathbf{M}$ to X, and the variables that are manipulated through X}, Y will remain constant[6].

The first requirement cannot tell us whether the change in Y occurred because of the change in X or because of the change in $\mathbf{M}$ directly. To meet the second requirement, the change in Y must occur only because of the change in X. When both requirements are met, the implication is that there is some dependence of Y on X that can be used to change the value of Y by changing X.

---

[5] This requirement makes the manipulation* framework non-reductive; a manipulation* relation cannot be described without appeal to other manipulation* relations. In this respect, it is similar to Woodward's framework (Woodward 2003).

[6] The requirement that *any* manipulation* variable can be used to keep X constant may seem too strong. However, note that for causal relations, the effect of the intervention variable on Y must be mediated through X by definition. Hence, however we keep X constant, while keeping all other variables that can manipulate* Y constant, Y will not change. This is also the case for mathematical relations, where the value of Y is determined by the value of X and by the other variables that mathematically define Y.

I take it that, in the cognitive sciences, if *Y* can be manipulated* through *X*, then, to some extent, *X* and the dependence relation explain *Y*. This is a counterfactual framework because the change in **M** describes a counterfactual scenario. However, the counterfactuals discussed here differ slightly from those found in Woodward (2003) and describe different possible manipulations* of *X*, which are not ideal interventions. I will assume here that in the counterfactual scenarios of the manipulations*, the mathematical relations of the factual world still hold[7].

Several points are worth noting here. First, I am certainly *not* suggesting that "if I can manipulate it I can explain it". Instead, the relation between manipulation and explanation is such that manipulation* relations and manipulating* variables can be used to explain the dependent explanandum. Second, like Woodward's manipulability for causal relations, the manipulation* relation does not have to be practically possible but only conceptually so. Finally, I focus on the cognitive sciences. It may be possible to extend this framework to other sciences, but I suspect that there are some fields, such as fundamental physics, that may not be as concerned about manipulation of their investigated phenomena. Therefore, I refrain from making a more general claim.

4 manipulation and control in mathematical explanations in the cognitive sciences

In this section, I use the manipulation* framework to analyze two examples of explanations in the cognitive sciences that appeal to mathematical relations. As a warm-up, I will take the well-known - albeit not from the cognitive sciences - example of a mathematical explanation: Königsberg's bridges (Craver and Povich 2017; Lange 2013; Reutlinger 2016).

Euler's theorem states that it is possible to walk through a graph traversing each edge exactly once (an Euler walk) iff exactly zero or two nodes in the graph are connected to an odd number of edges. Therefore, the fact that it is impossible for someone to take an Euler walk in Königsberg is explained by the fact that

[7] See (Baron et al. 2017) for a discussion of counterfactuals regarding mathematical relations.

Königsberg has four parts that are connected to an odd number of bridges. In this example, although in some conditions the organization of Königsberg's bridges (in terms of whether it meets Euler's criterion) and the possibility of an Euler walk there can each be derived from the other, we take the organization of Königsberg's bridges to explain the impossibility of an Euler walk there and not vice versa (Craver and Povich 2017). Intuitively, the direction of manipulation in this example coincides with the direction of explanation; we can manipulate the possibility of someone taking an Euler walk by changing the organization of Königsberg's bridges, but we cannot manipulate the organization of Königsberg's bridges by changing the possibility of someone taking an Euler walk. This intuition can be explicated in the manipulation* framework.

Let us consider a manipulation* variable **M** that can change the organization of Königsberg's bridges. For example, we can tear down a bridge in Königsberg with the purpose of having only two parts with an odd number of bridges. Such a change is expected to manipulate* both the organization of Königsberg's bridges and whether someone can take an Euler walk there. The change in possibility of taking an Euler walk is mediated via the change to the organization of Königsberg's bridges; any manipulation* to keep the organization of Königsberg's bridges constant (e.g., quickly build a new bridge) will make this walk impossible again. Thus, both requirements for a manipulation* relation are met: the possibility of an Euler walk can be manipulated* via the organization of Königsberg's bridges. When considering whether this manipulation* relation can also work in the other direction, we must seek a manipulation* that can change both variables such that when the possibility of an Euler walk is held constant, the organization of Königsberg's bridges would remain constant as well. However, we can hold the possibility of an Euler walk constant by barricading the city so that it is impossible for someone to take an Euler walk there. It is difficult to fathom a manipulation* that would change the organization of Königsberg's bridges when the city is not barricaded but would not affect this organization when the city is barricaded. Considering the destruction of bridges, it would change both variables, but the change in the organization of Königsberg's bridges would remain regardless of whether the city is barricaded. Therefore, until someone comes up with an example that fits this requirement, this framework does *not* imply that we can manipulate* the organization of Königsberg's bridges via the possibility of an Euler walk. In this example, the direction of manipulation* fits the direction of explanation.

4.a Optimal integration of information from two modalities

Consider a task where you are asked to estimate the length of a wooden bar. You have both visual and haptic inputs that reflect the length of this bar, but because both these inputs are noisy, the visual and haptic inputs differ slightly. What will your answer be? Ideally, you would like your answer to be optimal in the sense that, given the information you have, it will minimize the difference between your estimate and the true bar length. Measurements of this difference are called 'cost functions'.

It can be shown mathematically that when the inputs from the two modalities are independent and normally distributed around the true bar length (see Fig. 1a), the following estimate minimizes three common cost functions (number of errors, mean absolute error (L1) and mean squared error (L2))[8]. This estimate is a weighted mean of the inputs, so that the weight of each modality is inversely related to the variance of the input noise (see Fig. 1b):

$$(1)\ \ Estimate(\mu) = \frac{S_V \cdot \left(\frac{1}{\sigma_V^2}\right) + S_H \cdot \left(\frac{1}{\sigma_H^2}\right)}{\left(\frac{1}{\sigma_V^2}\right) + \left(\frac{1}{\sigma_H^2}\right)}$$

Where μ is the real length of the bar, S_V and S_H are the inputs that we get from the visual and the haptic modalities, respectively, and σ_V^2 and σ_H^2 are the variance of the noise of visual and haptic inputs (i.e., $S_V \sim N(\mu, \sigma_V^2)$, $S_H \sim N(\mu, \sigma_H^2)$).

[8] Throughout this discussion I assume that we have no information about the prior probability of the bar length.

Therefore, if one posits that the inputs of the different modalities are distributed as described and the cost of errors in the task is one of the three common cost functions, an optimal strategy would be to answer in accordance with (1). Indeed, Ernst and Banks (2002) discovered that, in such a task, people gave answers that were similar to the answers equation (1) would yield.

Now, one can ask ‘why is it the case that eq. (1) is optimal?’. We can answer this question by referring to a mathematical relation. It can be shown that when it is assumed that the distributions of the inputs are independent and normal with an expected value that is the real bar length μ (as in Fig. 1a), then equation (1) can be mathematically derived as minimizing the common cost functions. But (1) may not be optimal if these assumptions about the inputs are not correct. For example, if the expected value of the visual and haptic inputs, S_V and S_H, is not the actual bar length μ (i.e., they are biased estimates) then equation (1) will not yield an optimal answer (see Fig. 1c-d). The optimal estimate will be one that takes this bias into account. Therefore, the optimality of (1) depends on the probability distributions of the inputs. The probability distributions of the inputs and the mathematical derivation that yields (1) as the optimal estimate together explain the optimality of (1).

The probability distributions of the inputs explain (1)’s optimality even though (1)’s dependence on the probability distributions of the inputs would generally not be considered causal: the probability distributions of the inputs and the optimality of (1) occur simultaneously, and the dependence is between variables that are mathematically connected rather than between two distinct variables (Craver and Bechtel 2007)[9].

[9] As discussed in the introduction, non-causal by popular opinion that considers simultaneous, mathematical relations to be non-causal.

Given that this relation is not causal, it seems that the mechanistic framework cannot account for it. How can the manipulation* framework elucidate this case? Intuitively, the optimality of (1) can be manipulated via the probability distributions of the inputs. I will show that this is indeed a manipulation* relation.

Let us consider a variable that can be used to manipulate* the probability distributions of the inputs. It is possible to change the probability distributions of the inputs by changing the experimental conditions. (Ernst and Banks 2002) used specialized lab equipment to simulate visual and haptic inputs that differed in their variance. Thus, it is possible to change experimental conditions so that the probability distributions of the inputs become biased (their expected value is no longer the true bar length) and by this to render (1) no longer optimal. We can show that the optimality of (1) can be manipulated* via the probability distributions of the inputs by finding a manipulation* variable that cannot change *Y* when *X* is held constant. Consider the aforementioned manipulation* variable, where the experimental conditions are changed with the purpose of biasing the visual and haptic inputs. It is possible to counter the change to the probability distributions of the inputs, for example by giving subjects special glasses that will remedy the bias. In such a case, the probability distributions of the inputs as well as the optimality of (1) will both remain constant. In fact, because of the mathematical dependence relation, we know that if we change the experimental conditions, however we choose to keep the probability distributions of the inputs constant, while keeping other relevant variables such as the cost function constant, the optimality of (1) will remain constant as well. Therefore, the two conditions for manipulation* are met. We can conclude that we can manipulate* the optimality of (1) via the probability distributions of the inputs, and therefore the latter, together with the mathematical dependence relation, explain the former.

What about the asymmetry of the direction of explanation, despite the symmetrical mathematical dependence (Craver 2016; Craver and Povich 2017)? The optimality of (1) is mathematically related to the probability distributions of the inputs. So, one might argue that the manipulation* relation should be symmetrical. Yet, it would seem very odd to say that the probability distributions of the inputs are explained by (1)'s optimality. Luckily, this direction of explanation is not a consequence of the manipulation* framework.

To see if the probability distributions of the inputs can be manipulated* via the optimality of (1), we search for a manipulation* variable **M** that can change the value of both variables, but if some variable is used to hold the optimality of (1) constant, the probability distributions of the inputs do not change. One way to hold the optimality of (1) constant is by changing the cost function. However, it is difficult to imagine how some manipulation* can change the probability distributions of the inputs for one cost function but not for another. For this reason, until someone comes up with such a variable, in this example, the manipulation* framework implies that manipulation* and explanation go only in one direction: the probability distributions of the inputs can be used to manipulate* and explain the optimality of (1), but not vice versa. The manipulation* framework yields the desired results: a symmetrical mathematical relation allows manipulation* only in one direction, which is the direction we would also take to be the direction of explanation.

4.b Cortical neurons spike irregularly despite having a large number of incoming synaptic connections

Generally speaking, neurons in the cortex fire irregularly (Softky and Koch 1993): their inter-spike intervals (the time between two consecutive spikes) vary greatly. A common regularity measure is the coefficient of variation (CV):

$$(1)\quad CV = \frac{\sigma_{\Delta t}}{\overline{\Delta t}}$$

Where Δt is the inter-spike interval, $\overline{\Delta t}$ is the mean of Δt and $\sigma_{\Delta t}$ is the standard deviation of Δt (for a period where many inter-spike intervals are measured). The CV of many cortical neurons tends to be between 0.4 and 1.2, while for regular firing we would expect $CV \ll 1$ (i.e., the CV should be an order of a magnitude smaller than 1; see simulated examples in Fig. 2a) (Softky and Koch 1993). Given that the number of input synapses on cortical neurons is on the order of thousands, this finding is bewildering. Usually, the firing of a neuron is viewed as reflecting an approximate summation of synaptic inputs. According to the Central Limit Theorem, when the number (n) of independently and identically distributed (iid) random variables is very large, the sum of these random variables has an asymptotically normal distribution with an expected value proportional to n and a standard deviation proportional to $\sqrt{n}$. Formally:

(2) $\lim_{n \to \infty} \sum_{i=1}^{n} x_i \sim N(n \cdot E(x), n \cdot \sigma_x^2)$

Where n is the number of inputs, x_i is a random variable i, $E(x)$ is the expected value of x and σ_x^2 is the variance of x. According to this formula, when the number of summed random variables is very large, the standard deviation of their sum (also called fluctuations in the signal) is equal to $\sqrt{n} \cdot \sigma_x$, which is negligible relative to the signal (i.e., the sum itself)[10]. To illustrate, if $E(x) = \sigma_x$, for a thousand inputs and total signal size of 1, the fluctuations will be around 0.03. This means that when the number of iid inputs is very large, we can mathematically derive that the total input will be approximately constant (Fig. 2b, left). Studies have shown that it is not likely that the irregular firing is an intrinsic property of the neurons (Mainen and Sejnowski 1995), and therefore the irregular firing is likely produced by large fluctuations in the inputs (Fig. 2b, right). So, the puzzling question is this: why do neurons with many input synapses receive highly fluctuating inputs, despite what we know from the Central Limit Theorem?

One possible explanation for the surprising irregularity of the neurons' firing is that the inputs are not independent. Instead, neurons are a part of a network in which excitatory and inhibitory synaptic inputs to each neuron are balanced such that most of the excitatory and inhibitory inputs cancel out and the total input is reduced to the order of magnitude of the fluctuations. Indeed, (van Vreeswijk and

[10] This occurs because the sum is proportional to n and the fluctuations are proportional to $\sqrt{n}$, so the sum and its fluctuations differ by a magnitude of $\sqrt{n}$.

Sompolinsky 1996) have shown that such a balance can be achieved in a network that has some general connectivity properties (e.g., one requirement is sparse connectivity). In this way, the number of inputs to each neuron is still very large but the total input fluctuates strongly. The theory of excitatory-inhibitory balance has received experimental support (Wehr and Zador 2003; Xue et al. 2014). According to this theory, the magnitude of fluctuations in the neurons' total input depends on the balance between excitatory and inhibitory synaptic inputs and therefore this inhibitory-excitatory (henceforth IE) balance explains the fluctuations.

As in the previous example, the relation between the IE balance and the fluctuations in the total input does not comply with our usual description of a causal relation; the variable 'fluctuations in total input to the neuron' is simultaneous with the variable 'IE balance', and the relation between the IE balance and the fluctuations in total input is a mathematical relation: without IE balance, the Central Limit Theorem yields a barely fluctuating input, and when there is IE balance in accordance with the model from (van Vreeswijk and Sompolinsky 1996), the mathematical model yields a highly fluctuating input.

According to the manipulation* framework, this relation is explanatory. There is a manipulation* variable that changes the IE balance and the fluctuations in the neuron's input. For example, we can block many of the inhibitory inputs, disturbing the IE balance in the network, and this will yield a barely fluctuating input. Furthermore, however we choose to restore the IE balance (e.g., by blocking many excitatory inputs or by increasing the firing rate of the remaining inhibitory inputs), we will also restore the fluctuations in the input. Therefore, this example meets the two requirements for manipulation* and, according to the manipulation* framework, the fluctuations in total input are explained by the IE balance.

What about the challenge from symmetry of non-causal relations (Craver 2016; Craver and Povich 2017)? We can see that the manipulation* account does *not* imply that the IE balance can be manipulated* via the fluctuations in total input. One way to keep the fluctuations in total input to the neuron constant is by using an electrode to add external current to the neuron. However, it is again difficult to fathom a variable that will change the balance between excitatory and inhibitory synaptic inputs for one value of electrode current but not for another value. Hence, there is no implication regarding manipulation* and explanation in the opposite direction, in accordance with our intuition about manipulation and explanation in this example.

I have brought three examples in which the manipulation* account can come to our aid in distinguishing explanatory from non-explanatory relations of mathematical dependence. I believe these examples show convincingly that, for some non-causal explanations, explanatory value is closely related to manipulation. In the following section, I discuss several possible objections to the proposed framework.

5 Possible criticisms of the manipulation* framework

a) The manipulation* view ignores important differences between causal and non-causal relations that make non-causal relations unfit for manipulation.

The manipulation* view bundles together causal and non-causal relations and treats them similarly. This, it may be argued, misses crucial differences between these relations. Importantly, when manipulation is discussed regarding causal relations one distinct variable is manipulated via another. However, for paradigm non-causal relations, the two variables are closely linked – they are logically or mathematically related or at least occupy the same space-time slice. What sense does it make to talk about manipulation when the two variables' values are determined simultaneously? It may make more sense to say that we are manipulating both variables together, through an external variable.

However, even though in the two examples from the cognitive sciences presented here the explanans occur simultaneously and are mathematically related to the explananda, in both cases discovering the manipulation* relation between the variables can help us manipulate the explananda in ways we could not have done before.

Considering the first example, the dependence of the optimality of an estimate on the probability distributions of the inputs allows us to organize experimental settings so that some estimate is optimal. This mathematical dependence is especially crucial since there is no way to observe the optimality of an estimate. Unlike the common case with causal relations where the values of the cause and the effect can be observed, the optimality of an estimate is a latent variable that can only be derived mathematically. Hence, this mathematical dependence is essential to the manipulation of the optimality of an estimate and cannot be replaced by causal dependences. Despite the fact that such optimal estimates are latent variables, currently, they play an important part in explaining the behavior of humans and animals (Berniker et al. 2010; Ernst and Banks 2002; Fernandes et al. 2014; Vul et al. 2014; Weiss et al. 2002) and therefore are central in the cognitive sciences[11].

Let us now consider the second example. Without the dependence of the fluctuations in total input to the neuron on the IE balance, we could still contemplate a causal manipulation of the fluctuations in total input through the activity of specific neurons, but this relation would lack systematicity and would be very difficult to generalize. In light of the mathematical dependence of the fluctuations in total input on the IE balance, we can know how the fluctuations will change when we change the activity of different pre-synaptic neurons because we can consider the change in IE balance.

[11] Some may be surprised that scientific explanations of phenomena can be given in terms of optimality. In the cognitive sciences, where behavior and neuronal activity are often explained by underlying computational models, such explanations are very common. Generally, these explanations assume that the cognitive system has evolved enough by evolution to reach some (at least locally) optimal strategy regarding perception and decision-making problems.

Therefore, while it is true that manipulations that employ a non-causal dependence of *Y* on *X* often (perhaps always) need an external variable that can causally affect *X*, I think it is undeniable that some non-causal dependences extend the ways in which we can manipulate phenomena.

b) Manipulation of variables in models is not equivalent to the manipulation of physical objects

One could argue that the manipulation* framework abuses the point that Woodward and Craver were trying to make; when Craver discusses manipulation of the CNS (central nervous system), he means that we want to manipulate and control actual physical objects: we want to cure Alzheimer's disease, treat anxiety disorders, or enable paraplegics to walk. My examples, this argument will continue, are of manipulation of abstract mathematical variables that appear only in models, and it is not clear how these variables relate to real, physical brains. In this sense, manipulation* does not truly allow us to manipulate the CNS.

It is true that, in the examples given here, the mathematical dependence relations are between abstract variables: estimates, probability distributions, random variables, etc. But these abstract relations are applied to real phenomena[12], allowing us to manipulate them. It is easy to see this point regarding the Königsberg's bridges example. Euler's mathematical theorem describes abstract phenomena, namely, graphs and paths. Nonetheless, this theorem has real, physical, implications: it would be impossible for me to take an Euler's walk in Königsberg.

In the example offered in 4.a, the mathematical dependence tells us what computation some machine or organism should perform under certain conditions to minimize estimation error. This estimation error may be related to an organism's fitness and affect its survival. In the example provided in 4.b, we can eliminate the fluctuations in the total input to the neuron by disrupting

[12] See (Kuorikoski and Ylikoski 2015) for a discussion of the relation between counterfactual dependences in models and in real phenomena.

the IE balance, and observe the results of this change. Therefore, we see that mathematical relations between abstract entities can allow the manipulation of physical phenomena.

c) manipulation* relations are explanatory relations only because both manipulation and explanation are related to more basic ontic relations, which are the interesting relations

I imagine this argument goes something like this: it may be true that explanatory relations and manipulation* relations tend to describe the same relations, but this is only because both rely on similar ontic relations such as cause-effect, part-whole, structure-function, etc. It is these ontic relations that should be examined and taken as relevant to explanations.

I cannot deny that manipulation* relations rely on some specific ontic relations – wholes can be manipulated through parts, effects through their causes, etc. The types of ontic relations that allow extended manipulation are definitely worth investigating. It is especially interesting that, in the given examples, the manipulation* is possible in exactly one direction because the manipulated* variable, *Y*, also depends on another variable that is independent of *X* and can be used to hold *Y* constant. Nonetheless, this does not diminish the importance of the fact that explanatory and manipulation* relations tend to be the same relations, and that explanation is tightly linked to manipulation, even for non-causal relations.

Moreover, while it may be possible to characterize explanatory relations as a collection of various ontic relations, such a description will not yield a reason for the explanatory value of these specific ontic relations but not others. In contrast, the notion that some relations explain a phenomenon because they allow its manipulation at least suggests a reason for the explanatory value of some relations and lack thereof of others.

d) The manipulation* framework is inferior to the mechanistic framework, which already has a clear formulation of causal relations as explanatory relations

The mechanists provide a clear and elegant framework where causal relations are explanatory. This framework accounts for causal and mechanistic explanations. Thus far, the mechanists have answered (Craver 2016; Kaplan 2011, 2017; Kaplan and Craver 2011; Piccinini and Craver 2011) most of the many challenges that have been presented to them (Bechtel and Shagrir 2015; Chirimuuta 2014; Egan 2017; Huneman 2010; Rusanen and Lappi 2016; Shagrir and Bechtel 2017; Shapiro 2017; Silberstein and Chemero 2013). It can be argued that, compared to the mechanistic framework, the manipulation* framework is overly broad and adds unnecessary complications in an attempt to answer questions already dealt with by the mechanistic framework.

My response to this criticism is twofold. First, like many other non-causal explanations found in the literature, this paper presents two non-causal explanations that the mechanistic framework does not easily accommodate. The mechanists would probably have to argue that the examples I offered are not explanations, that they appeal to some causal relation or that they are exceptions to their general framework. Alternatively, they could argue that mathematical and constitutive relations are causal. None of these options seems very natural to me, while the manipulation* framework accommodates these examples easily.

Second, according to the mechanistic framework, explanations describe relevant causal relations. However, many explanatory dependence relations in this framework are the relations between the explanandum phenomenon and the components in the mechanistic decomposition of this phenomenon. These dependence relations are not causal, but constitutive. In his seminal work, Craver (2007a) describes the relations between a phenomenon and its mechanistic components also as manipulability relations, based on Woodward's (2003) framework for causal relations.

However, many have made the point that the mechanistic framework has problems with describing manipulation and intervention in a way that fits relations between the phenomenon and its mechanistic components (Baumgartner and Casini 2017; Baumgartner and Gebharter 2016; Harbecke 2010; Harinen 2014; Leuridan 2012; Romero 2015). The arguments in these works are usually similar in spirit to the one by (Romero 2015) presented in section 3: phenomena and their mechanistic components are related in part-whole relations, and occupy the same space-time slice, so it is problematic to talk about an ideal intervention in Woodward's sense (2003) on one with respect to the other. Even if such interventions are possible, this could imply that constitutive relations are causal, a result that many believe should be avoided (Baumgartner and Gebharter 2016; Craver and Bechtel 2007; Romero 2015).

Thus, it seems that if one wished to argue that relevant mechanistic components allow manipulation of the explanandum, one might have to forgo the claim that only causal relations in Woodward's sense allow manipulation[13]. In many ways, then, constitutive relations in mechanistic explanations face similar issues to the mathematical relations I described here, and so, these too can benefit from a framework that accommodates non-causal relations.

6 Conclusions

There are two promising frameworks for explanations in the cognitive sciences. One of these takes the view that counterfactual dependences, causal and non-causal alike, are the basis for explanations. The second, mechanistic framework, emphasizes the relation between manipulation and explanation and

[13] Another baffling issue in Craver's mutual manipulability criterion is that Craver takes the direction of manipulation to go both from phenomenon to its components and from the components to the phenomenon, while the direction of explanation goes only from the components to the phenomenon. Franklin-Hall's (2016) interpretation of mutual manipulability suggests a solution to this issue: top-down manipulation amounts to manipulation of the input conditions of the phenomenon. So, we can consider this top-down manipulation a causal manipulation of components by the inputs.

takes only causal dependences to be the basis for explanations. In this paper, I suggested a view of explanation that relates to both these frameworks. I argued that some non-causal counterfactual dependence relations also allow manipulation of the explanandum. This may be a good enough reason for the mechanists to also accept some non-causal relations as explanatory. Moreover, whether counterfactual-dependence relations allow manipulation may enable us to differentiate explanatory from non-explanatory ones. A major advantage of this framework is that it suggests a general criterion for explanatory value in the cognitive sciences without relinquishing non-causal explanations. In this paper, I focused on relations of mathematical dependence in which the counterfactual dependence can be determined analytically. Future work should discuss other counterfactual dependence relations and how they can be identified. This can be a step towards a more unified view of explanations in the cognitive sciences.

Acknowledgements:

I am thankful to Nir Fresco, Jens Harbecke, Arnon Levy, Jonathan Najenson, Oron Shagrir and three anonymous referees for thoughtful and helpful comments on this manuscript. For their insightful comments, I am also grateful to the participants of 'Explanatory Reasoning in the Sciences (The Second Jerusalem-MCMP Workshop in the Philosophy of Science)' in Munich 2017, and to participants in the workshops of the German-Israeli Foundation research group 'Causation and Computation in Neuroscience' in Jerusalem 2017, Cologne 2016 and Witten 2015. This research was supported by a Grant from the GIF, the German-Israeli Foundation for Scientific Research and Development.

References

Baron, S., Colyvan, M., & Ripley, D. (2017). HOW MATHEMATICS CAN MAKE A DIFFERENCE. *Philosophers' Imprint*, *17*, 3.

Baumgartner, M., & Casini, L. (2017). An Abductive Theory of Constitution. *Philosophy of Science*, *84*, 214–233. doi:10.1086/690716

Baumgartner, M., & Gebharter, A. (2016). Constitutive Relevance, Mutual Manipulability, and Fat-Handedness. *The British Journal for the Philosophy of Science*, *67*, 731–756. doi:10.1093/bjps/axv003

Bechtel, W., & Shagrir, O. (2015). The Non-Redundant Contributions of Marr's Three Levels of Analysis for Explaining Information-Processing Mechanisms. *Topics in Cognitive Science*, *7*, 312–322.

Berniker, M., Voss, M., & Kording, K. (2010). Learning priors for bayesian computations in the nervous system. *PLoS ONE*, *5*, e12686. doi:10.1371/journal.pone.0012686

Bokulich, A. (2011). How scientific models can explain. *Synthese*, *180*, 33–45. doi:10.1007/s11229-009-9565-1

Chirimuuta, M. (2014). Minimal models and canonical neural computations: the distinctness of computational explanation in neuroscience. *Synthese*, *191*, 127–153.

Chirimuuta, M. (2017). Explanation in Computational Neuroscience: Causal and Non-causal. *The British Journal for the Philosophy of Science*, (axw034). doi:10.1093/bjps/axw034

Craver, C. F. (2007a). *Explaining the Brain*. Oxford University Press.

Craver, C. F. (2007b). Constitutive Explanatory Relevance. *Journal of Philosophical Research*, *32*, 3–20. doi:10.5840/jpr20073241

Craver, C. F. (2016). The Explanatory Power of Network Models. *Philosophy of Science*, *83*, 698–709.

Craver, C. F., & Bechtel, W. (2007). Top-down causation without top-down causes. *Biology and Philosophy*, *22*, 547–563. doi:10.1007/s10539-006-9028-8

Craver, C. F., & Povich, M. (2017). The directionality of distinctively mathematical explanations. *Studies in History and Philosophy of Science*, *63*, 31–38. doi:10.1016/j.shpsa.2017.04.005

Cummins, R. (1983). *The Nature of Psychological Explanation*. MIT Press.

Cummins, R. (2000). “How does it work?” vs. “What are the laws?” Two conceptions of psychological explanation. In F. Keil & R. A. Wilson (Eds.), *Explanation and Cognition* (pp. 117–145). MIT Press.

Dretske, F. (1994). If You Can’t Make One, You Don’t Know How It Works. *MIDWEST STUDIES IN PHILOSOPHY*, *19*, 468–482.

Egan, F. (2017). Function-Theoretic Explanation and Neural Mechanisms. In D. M. Kaplan (Ed.), *Explanation and Integration in Mind and Brain Science* (pp. 145–163). Oxford University Press.

Ernst, M. O., & Banks, M. S. (2002). Humans integrate visual and haptic information in a statistically optimal fashion. *Nature*, *415*, 429–433. doi:10.1038/415429a

Fernandes, H. L., Stevenson, I. H., Vilares, I., & Kording, K. P. (2014). The generalization of prior uncertainty during reaching. *The Journal of Neuroscience*, *34*, 11470–84. doi:10.1523/JNEUROSCI.3882-13.2014

Franklin-Hall, L. R. (2016). New Mechanistic Explanation and the Need for Explanatory Constraints. In A. Ken & G. Carl (Eds.), *Scientific Composition and Metaphysical Ground* (pp. 41–74). Palgrave Macmillan UK.

Harbecke, J. (2010). Mechanistic Constitution in Neurobiological Explanations. *International Studies in* *the Philosophy of Science*, *24*, 267–285. doi:10.1080/02698595.2010.522409

Harinen, T. (2014). Mutual manipulability and causal inbetweenness. *Synthese*, *195*, 35–54. doi:10.1007/s11229-014-0564-5

Hitchcock, C., & Woodward, J. (2003). Explanatory Generalizations, Part II: Plumbing Explanatory Depth. *nous*, *37*, 181–199.

Huneman, P. (2010). Topological explanations and robustness in biological sciences. *Synthese*, *177*, 213–245.

Jansson, L. (2015). EXPLANATORY ASYMMETRIES: LAWS OF NATURE REHABILITATED. *The Journal of Philosophy*, *112*, 577–599.

Jansson, L., & Saatsi, J. (2017). Explanatory Abstractions. *The British Journal for the Philosophy of Science*, axx016. doi:10.1093/bjps/axx016

Kaplan, D. M. (2011). Explanation and description in computational neuroscience. *Synthese*, *183*, 339–373.

Kaplan, D. M. (2017). Neural computation, multiple realizability, and the prospects for mechanistic explanation. In D. M. Kaplan (Ed.), *Explanation and Integration in Mind and Brain Science* (pp. 164–189). Oxford University Press.

Kaplan, D. M., & Craver, C. F. (2011). The Explanatory Force of Dynamical and Mathematical Models in Neuroscience : A Mechanistic Perspective. *Philosophy of Science, 78*, 601–627.

Kuorikoski, J., & Ylikoski, P. (2015). External representations and scientific understanding. *Synthese*, *192*, 3817–3837. doi:10.1007/s11229-014-0591-2

Lange, M. (2013). What Makes a Scientific Explanation Distinctively Mathematical ? *The British Journal* *for the Philosophy of Science*, *64*, 485–511. doi:10.1093/bjps/axs012

Lazebnik, Y. (2002). Can a biologist fix a radio? - Or, what I learned while studying apoptosis. *CANCER CELL*, *2*, 179–182. doi:10.1007/s10541-005-0013-7

Leuridan, B. (2012). Three Problems for the Mutual Manipulability Account of Constitutive Relevance in Mechanisms. *The British Journal for the Philosophy of Science*, *63*, 399–427. doi:10.1093/bjps/axr036

Mainen, Z. F., & Sejnowski, T. J. (1995). Reliability of spike timing in neocortical neurons. *Science*, *268*, 1503–6. doi:10.1126/science.7770778

Milkowski, M. (2013). *Explaining the Computational Mind*. MIT Press.

Pexton, M. (2016). THERE ARE NON-CAUSAL EXPLANATIONS OF PARTICULAR EVENTS. *Metaphilosophy*, *47*, 264–282.

Piccinini, G. (2015). *Physical Computation: A Mechanistic Account*. Oxford University Press.

Piccinini, G., & Craver, C. F. (2011). Integrating psychology and neuroscience: functional analyses as mechanism sketches. *Synthese*, *183*, 283–311.

Reutlinger, A. (2016). Is There A Monist Theory of Causal and Non-Causal Explanations ? The Counterfactual Theory of Scientific Explanation. *Philosophy of Science*, *83*, 733–745. doi:10.1086/687859

Romero, F. (2015). Why there isn’t inter-level causation in mechanisms. *Synthese*, *192*, 3731–3755. doi:10.1007/s11229-015-0718-0

Rusanen, A., & Lappi, O. (2016). On computational explanations. *Synthese*, *193*, 3931–3949.

Saatsi, J., & Pexton, M. (2013). Reassessing Woodward’s Account of Explanation: Regularities, Counterfactuals, and Noncausal Explanations. *Philosophy of Science*, *80*, 613–624.

Shagrir, O. (2006). Why we view the brain as a computer. *Synthese*, *153*, 393–416.

Shagrir, O., & Bechtel, W. (2017). Marr's Computational Level and Delineating Phenomena. In D. M. Kaplan (Ed.), *Explanation and Integration in Mind and Brain Science* (pp. 190–214). Oxford University Press.

Shapiro, L. A. (2017). Mechanism or Bust? Explanation in Psychology. *The British Journal for the Philosophy of Science*, *68*, 1037–1059.

Silberstein, M., & Chemero, A. (2013). Constraints on Localization and Decomposition as Explanatory Strategies in the Biological Sciences. *Philosophy of Science*, *80*, 958–970.

Softky, W. R., & Koch, C. (1993). The highly irregular firing of cortical cells is inconsistent with temporal integration of random EPSPs. *The Journal of Neuroscience*, *13*, 334–50. http://www.jneurosci.org/content/jneuro/13/1/334.full.pdf

van Vreeswijk, C., & Sompolinsky, H. (1996). Chaos in neuronal networks with balanced excitatory and inhibitory activity. *Science*, *274*, 1724–6. doi:10.1126/science.274.5293.1724

Vul, E., Goodman, N., Griffiths, T. L., & Tenenbaum, J. B. (2014). One and Done? Optimal Decisions From Very Few Samples. *Cognitive Science*, *38*, 599–637. doi:10.1111/cogs.12101

Wehr, M., & Zador, A. M. (2003). Balanced inhibition underlies tuning and sharpens spike timing in auditory cortex. *Nature*, *426*, 442–446. doi:10.1038/nature02116

Weiss, Y., Simoncelli, E. P., & Adelson, E. H. (2002). Motion illusions as optimal percepts. *Nature Neuroscience*, *5*, 598–604. doi:10.1038/nn858

Woodward, J. (2003). *Making Things Happen*. Oxford: Oxford University Press.

Woodward, J. (2010). Causation in biology: stability, specificity, and the choice of levels of explanation. *Biology and Philosophy*, *25*, 287–318. doi:10.1007/s10539-010-9200-z

Woodward, J. (2018). Some Varieties of Non-Causal Explanation. In A. Reutlinger & J. Saatsi (Eds.), *Explanation Beyond Causation* (pp. 117–140). Oxford University Press.

Woodward, J., & Hitchcock, C. (2003). Explanatory Generalizations , Part I : A Counterfactual Account. *NOUS*, *1*, 1–24.

Xue, M., Atallah, B. V., & Scanziani, M. (2014). Equalizing excitation-inhibition ratios across visual cortical neurons. *Nature*, *511*, 596–600. doi:10.1038/nature13321

Ylikoski, P., & Kuorikoski, J. (2010). Dissecting explanatory power. *Philosophical Studies*, *148*, 201–219. doi:10.1007/S11098-008-9324-Z

Figures

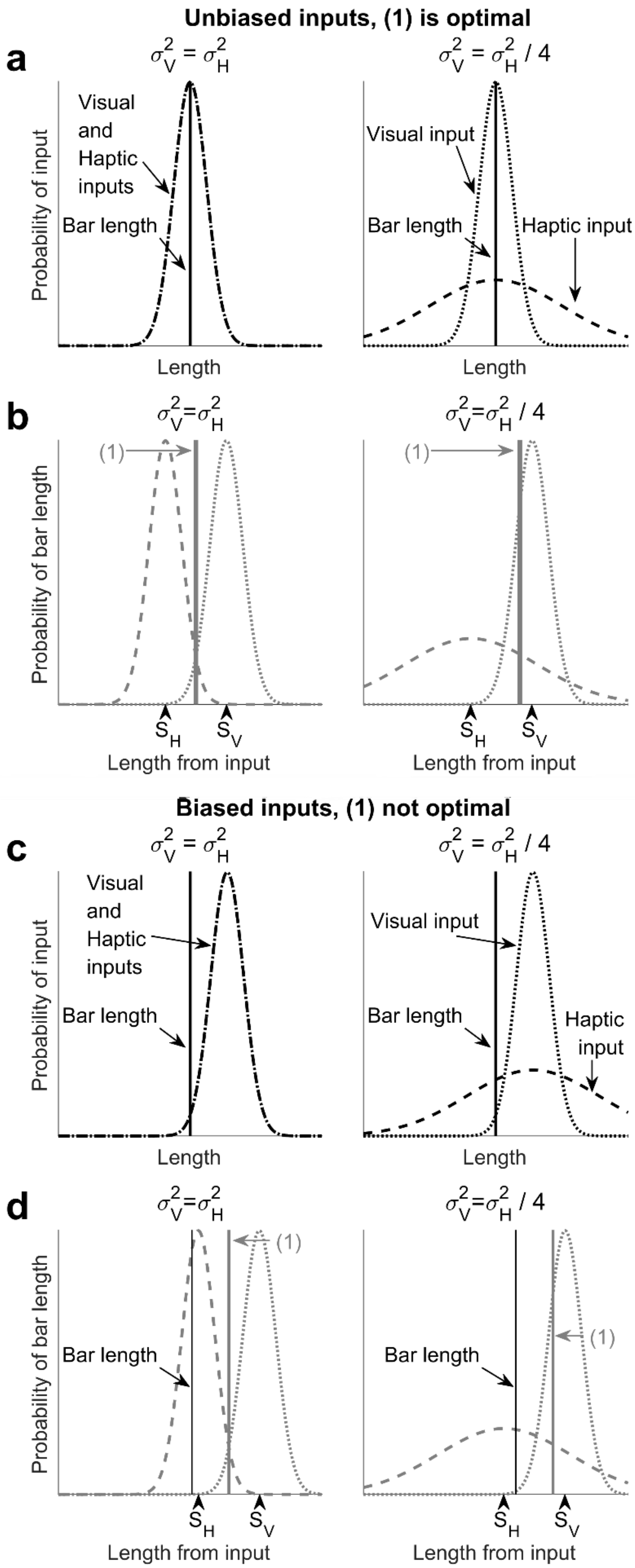

Figure 1 Equation (1) is an optimal estimate of the bar length when inputs are normally distributed around the actual bar length and is sub-optimal when inputs are biased (based on analyses from (Ernst and Banks 2002)). The probability distributions of the inputs given the actual bar length are denoted in black. The probability distributions of the bar length, given each input, are denoted in grey **(a-b)** Estimation is optimal when inputs are unbiased. **(a)** Probability distribution of inputs given the actual bar length. (σ_V^2) and (σ_H^2) are the variances of the visual and haptic inputs **(b)** Example of estimation of bar length using (1) from visual and haptic inputs. S_V and S_H are the visual and haptic inputs. Dotted and dashed gray distributions are the probability distributions of bar length from visual input alone, and haptic input alone, respectively. The line denoted by (1) is the estimated bar length according to eq. (1). On the left, the variances of the inputs are equal. On the right, the variance of the haptic input is much larger. Although the estimate from (1) is not exactly the actual bar length, it is optimal because on average it yields the minimal error. **(c-d)** same as in **(a-b)** for biased inputs. **(c)** Probability distributions are biased so that the expected values of these distributions are not equal to the actual bar length. **(d)** Example of estimates of bar length using (1) from visual and haptic inputs. True bar length is denoted in black. Legend otherwise is the same as in **(b)**. Because inputs are biased, the estimate given by (1) is not optimal.

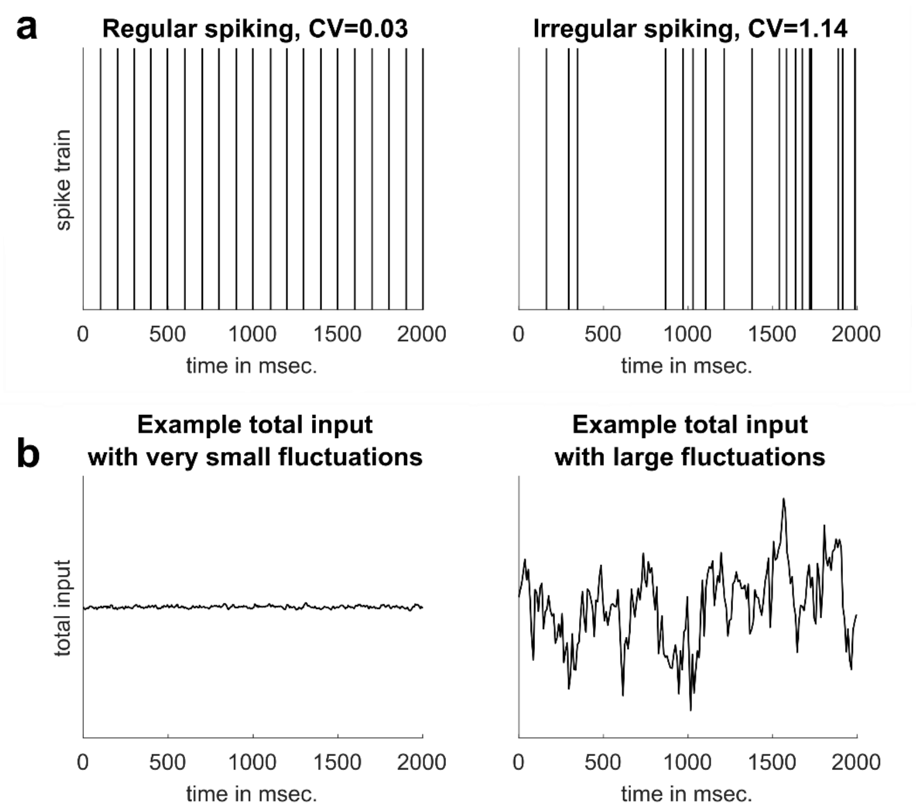

Figure 2 Simulated examples. **(a)** Spike trains of regularly and irregularly spiking neurons. Both neurons have an average firing rate of 10 spikes/s. **(b)** Simulated total synaptic input current. Left, synaptic input is barely fluctuating. This is the type of input we expect from many independent synapses. Right, synaptic input is highly fluctuating. This is the type of input we expect in the case that inhibitory and excitatory synaptic inputs are balanced.