



Epistemic values of quantity and variety of evidence in biological mechanism research

Yin Chung Au¹

Received: 3 March 2020 / Accepted: 26 March 2021 / Published online: 12 June 2021
© The Author(s) 2021

Abstract

This paper proposes an extended version of the interventionist account for causal inference in the practical context of biological mechanism research. This paper studies the details of biological mechanism researchers' practices of assessing the evidential legitimacy of experimental data, arguing why quantity and variety are two important criteria for this assessment. Because of the nature of biological mechanism research, the epistemic values of these two criteria result from the independence both between the causation of data generation and the causation in question and between different interventions, not techniques. The former independence ensures that the interventions in the causation in question are not affected by the causation that is responsible for data generation. The latter independence ensures the reliability of the final mechanisms not only in the empirical but also the formal aspects. This paper first explores how the researchers use quantity to check the effectiveness of interventions, where they at the same time determine the validity of the difference-making revealed by the results of interventions. Then, this paper draws a distinction between experimental interventions and experimental techniques, so that the reliability of mechanisms, as supported by the variety of evidence, can be safely ensured in the probabilistic sense. The latter process is where the researchers establish evidence of the mechanisms connecting the events of interest. By using case studies, this paper proposes to use 'intervention' as the fruitful connecting point of literature between evidence and mechanisms.

Keywords Biological mechanism · Evidence · Practice · Causation · Independence

✉ Yin Chung Au
z10608041@email.ncku.edu.tw

¹ Department of History, National Cheng Kung University, No. 1 University Road, 70101 Tainan City, Taiwan

1 Introduction

This paper proposes an extended version of the interventionist account for causal inference in the practical context of biological mechanism research. While there is an enormously rich body of literature on scientific data, evidence and phenomena, my interest in the epistemic roles of evidence in the search for biological mechanisms arose from an asymmetry between data and evidence: data can ‘become’ evidence, whereas evidence cannot become data. The existing literature has discussed data and evidence somehow separately, though these two terms seem to be interchangeable in some scholars’ works. However, at least in the practice of biological mechanism research, researchers do not confuse data with evidence and know clearly what data are useful for causal inference. This is to say that the practitioners know what sort of data, and not all, have legitimate evidential associations with the phenomena of interest and support the reliability of causal conclusions, despite that fact that they also use these two terms interchangeably.

This paper argues that two criteria for assessing the evidential legitimacy of biological data, i.e. quantity and variety, help to guarantee robust causal conclusions because two kinds of causal independence operate in experimental interventions. There are other criteria for assessing data in more specific contexts, and this paper does not aim to exhaust them. For example, whether the data represents phenomena on an in-situ basis, namely showing the phenomena of interest as closely to their actual biological environments as possible, is a criterion for assessing how convincing the data is to the exert community. Success in in-situ experimentation implies that the experimentation is able to reconstruct the relevant environments and that the results are promising for medical and/or therapeutic applications. Another criterion is direct visualisation and often relates to in-situ representation. That is, data that visualises the phenomena of interest in more direct and intuitive ways is considered more convincing. This criterion has been increasingly important due to the rapid advancements in imaging technologies since the turn of the century. Nonetheless, I consider that quantity and variety are central to data assessment. Researchers tend to determine whether data can become evidence based on the quantity and the variety *before* they seek to obtain data satisfying the other criteria. This is to say that quantity and variety are central to data assessment. Therefore, I focus on these two criteria in this paper.

I will argue to put ‘intervention’ in the structure of the philosophical understanding of evidence for biological mechanisms, so that the epistemic values of the seemingly common-sense requirements for quantity and variety can be appreciated. Variety is particularly what I seek to elaborate, where I need to discuss quantity first because researchers use quantity to validate difference-making between events of interest before they seek various evidence of mechanisms. The validity of claims on difference-making depends on whether an intervention is effective enough to produce an adequate amount of reliable data. Then, the confirmation of mechanisms follows this validation of difference-making and requires various evidence to ensure the robustness of the conclusions. I will show

that researchers intervene in not only the causal processes in question but also the components of the mechanisms of interest without the need to worry about causal interactions between the expected outcomes of interventions. I maintain that this subtle distinction between ‘interventions in the causal processes in question’ and ‘interventions on individual components’ should not be taken for granted.

Section 2 is dedicated to the first criterion: quantity. I will first introduce two kinds of evidence for biological mechanisms, which have been argued by some philosophers. I will suggest that this characterisation of evidence for biological mechanisms applies to the discussion of the basic research of biology. I will then explain why quantity is an important factor of decisions on whether a set of data is legitimate in becoming evidence. In short, by obtaining an adequate amount of data, researchers simultaneously examine the evidential associations between data and phenomena (Bogen & Woodward, 1988; Woodward, 1989, 2000) and confirm the validity of the difference-making revealed by the data.

Section 3 elaborates on the other criterion: variety. This is the criterion for assessing whether the data can become the evidence of mechanisms (Illari, 2011). This type of evidence contributes more directly to the causal inference of mechanisms than difference-making evidence. I will argue for a distinction between interventions and techniques. The variety of independent evidence ensures the reliability of mechanisms because, in the practical context, various evidence is obtained through independent interventions. Independent interventions do not equal independent experimental techniques because they test different hypotheses in different and independent contexts. I will use examples to demonstrate that the decompositional method for studying biological mechanisms tends to treat different components of the mechanisms as different and independent hypotheses. This special feature saves the reliability of mechanistic explanations from the risk of pseudorobustness caused by incomplete ontic independence in the Bayesian sense. This feature also supports the reliability of mechanisms more strongly than merely using empirical investigations of data generation procedures (Kuorikoski & Marchionni, 2016; Woodward, 2000).

In conclusion, this paper seeks to show that appreciating the details of interventions by studying the practice of biology has both empirical and formal meanings.

2 Assessing the effectiveness of interventions

This section argues to place ‘intervention’ in the structure used to understand biological evidence (Fig. 1). I distinguish between two kinds of causal processes: one is responsible for data generation, and the other serves as a part of the final conclusions. This section will make two points: 1) the causation involved in data generation procedures in biological mechanism research is not normally related to the final causation of interest. This point extends the existing studies (Kuorikoski & Marchionni, 2016; Woodward, 1989) by examining the details of the distinction between the two kinds of causal processes; 2) this section will explain this distinction by pointing out that experimental interventions in biological research are not always intended to determine the causation that directly supports the conclusion but to confirm specific

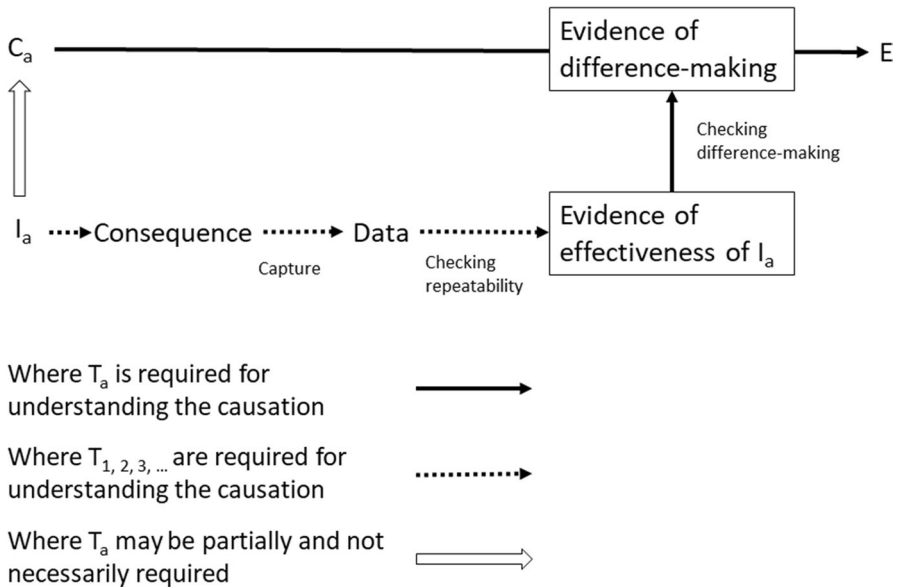


Fig. 1 Processes of determining both data repeatability and difference-making

components of causal mechanisms. Meanwhile, researchers actually seek to know the effectiveness of interventions before they start the causal inference process that will lead to the conclusion of a research project. While intervention is not a new topic in the literature, the reliability of intervention seems to be a presumption and thus black-boxed. I suggest that how researchers ensure the reliability of intervention has something to do with how they determine the evidential status of data produced by the intervention.

2.1 Two kinds of evidence for biological mechanisms

I borrow the classification between ‘difference-making evidence’ and ‘evidence of mechanisms’ from the existing literature on the health sciences (Dragulinescu, 2017; Illari, 2011; Russo & Williamson, 2007) to explore whether they are both required in the basic research of biology. Difference-making evidence means the kind of evidence that, to quote Illari (2011), ‘the effect does indeed vary with the postulated cause’ (144). In biology, whether such variance is observable is normally determined after an intervention. Russo and Williamson refer to this relationship between an effect and the postulated cause as ‘probabilistic dependence’, meaning that an increase in the probability of the cause is expected to result in an increase in the probability of the effect. Because the following section will discuss biological evidence from a formal aspect and mention the term ‘probabilistic independence’ between various kinds of evidence, a terminological demarcation is required here to avoid confusion between the different terms and meanings. Thus, I adopted Illari’s

term ‘difference-making evidence’ (2011, 139), which she uses in her paper of disambiguating Russo and Williamson’s concepts, to refer to this kind of evidence.

The other kind of evidence is ‘evidence of mechanism’. Illari (2011) lists several possible types of evidence of mechanism. For the focus of this paper, I paraphrase a type that is useful for this paper as follows: evidence of mechanism supports the claim on the existence of a mechanism that physically connects two or more events of interest. I add that biological mechanisms constructed in basic research necessarily contain causalities while clarifying that not all the relationships between the components are causal. The components are entities and activities as adopted from the characterisation argued by Bechtel (2006), Craver and Darden (2013) and Illari and Williamson (2012). I maintain that while some activities result in causal relationships between the components, others result in other kinds of relationships, such as temporal sequences, synchronicity, regulatory interactions and biochemical bindings. This does not contradict the idea that the overall effect of causal and non-causal relationships must be able to explain the event of interest.

While scholars have addressed the problems of the Russo and Williamson Thesis (RWT) that difference-making evidence and evidence of mechanism are both required for causal inference in the health sciences,¹ I assume these dual requirements to be true for basic biological research and will support my assumption by studying the practice. I seek to show that investigatory practices in biological mechanism research are conducted to obtain both kinds of evidence. The following sections will show how biological researchers determine these two evidential statuses of data and why these two kinds are both required for the confidence regarding the final causal conclusions in biological basic research.

How do biological mechanism researchers decide whether a set of data is legitimate evidence, given the abovementioned asymmetry between data and evidence? I consider the answer has two parts. The first is the decision on whether the data reveals a valid difference-making between two events of interest. This is discussed in this section. The second is the decision on whether the data serves together with data obtained from other interventions as the evidence of mechanisms *while* being causally independent of the other data. This will be discussed in Sect. 3. The former decision is made against quantity, and the latter is made against variety in a formal sense. I suggest that both decisions have much to do with the effectiveness of interventions.

When a set of data satisfies the requirement for repeatability, it is used as the evidence of the reliability of the data generation procedures. Whereby researchers determine that the difference-making revealed by the data is *valid*. Note that when mentioning the changes produced by interventions, I do not use the phrase ‘between cause and effect’ but ‘between the events of interest’. This nuance is

¹ The two kinds of evidence argued by Russo-Williamson Thesis are not always considered both necessary in medicine. Howick (2011) focuses on how important mechanisms are in providing evidence for medical decision-making and argues that tightly controlled clinical studies are sufficient for establish causation. He considers that the incomplete and complicated nature of mechanisms result in some problems of using mechanisms to connect clinical interventions and the outcomes. Meanwhile, Howick and Glasziou (2013) address four problems of using mechanisms to either understand target populations that are beyond controlled studies or solve the problems of simple induction.

important to my argument because I do not consider most experimental interventions in biological mechanism research to directly approach the final causation. As explained above, this is because the relationships between the components contained by mechanisms do not have to be all causal. Thus, interventions are normally designed to check the *roles*, which can have a variety of sorts, of specific components within the mechanistic explanations that the researchers aim to establish eventually.

In this sense, while difference-making needs to be determined valid to contribute to the final explanation, the events involved in the difference-making under investigation are not necessarily considered by the researchers as all causes in the final explanation. Thus, I tend to consider difference-making to occur between two events, not necessarily a cause and an effect. Because this saying might appear inconsistent with the RWT, I clarify that I am not against their view but seek to introduce Glennan's (2002) causal relevance to this discussion. What I term as 'non-causal relationships between events of interest' is loosely characterised as a form of causal relevance. That is, one event can be a necessary background condition for the occurrence of another event from a particular perspective without being the sufficient cause of the latter. This is due to the complexity of biological mechanisms. There can always be underdetermined or partially determined involvements of other variables in the causal process among the events under investigation. Moreover, the involvement of an event in difference-making can be causally relevant to the effect in non-positive and very indirect ways, e.g. being either confounding variable or a part in a negative feedback cycle. Therefore, evidence of mechanism is required for revealing what roles the events play in the explanation of interest, and mechanism researchers only conclude that an event involved in a difference-making is a cause when they understand its role in the *production* of the effect. This is consistent with Russo and Williamson's point that difference-making evidence 'needs to be accounted for by an underlying mechanism before the causal claim can be established' (Russo & Williamson, 2007, 159, as cited in Illari, 2011, 139).

In other words, events involved in difference-making are not always considered productive but relevant causes in the practice. Now I analyse the practice to show the following. First, the determination of these relevant causes is a necessary step before the determination of productive causes. Second, this determination normally relies on obtaining and assessing an adequate amount of repeatable data *because* there are two kinds of causal independence operating in the interventions that researchers conduct to obtain the data. One is the causal independence between different interventions for determining different components, and the other is the causal independence between the background theory for an intervention and the theory for the final conclusion. The existing version of the interventionist account has not revealed the value of such independence.

The checking of data generation procedures has been discussed by a rich body of literature. The most relevant works to this paper are Bogen and Woodward (1988) and Woodward (1989, 2000) on causal inferences among data, phenomena and theory. I fully agree with and have adopted Woodward's notion of 'empirical investigations'. Such investigations eliminate shared systematic errors and biases between applications of experimental detection by checking the data generation procedures.

They are crucial to the repeatability² of data, and the reliability of the evidential association between data and the phenomena claims need to be assessed based on enough repeatable legitimate data. Although Woodward's serial papers on the interventionist account cannot be captured by an oversimplified summary, I need to simplify an idea here for convenience because it is essential to my arguments: the assessment of the reliability of phenomena claims based on their evidential relations to the data relies upon the assessment of the reliability of the data generation procedures. However, referring to the checking of the reliability of data generation procedures as mere 'empirical investigations' seems to black-box this practice, though the existing studies have named a lot of real-world examples. This black-boxing makes the empirical investigations on devices, instruments, materials and other experimental settings misleadingly seem to be safe presumptions once they have been conducted.

2.2 Independent causal processes

I expand upon the interventionist view by showing the epistemic values of checking the effectiveness of interventions. I intend to argue that, at least for biological mechanism research, highlighting the epistemological role of interventions helps philosophers understand how data becomes evidence and what biologists seek to do by turning data into evidence. This section explains the role of effective interventions in turning enough quantity of data into evidence. That is, the checking of data generation procedures helps to consolidate difference-making phenomena, or determine causal relevance, not because interventions straightforwardly ensure the procedures to be reliable but because the reproduction of data can be causally independent of the causation of interest. Meanwhile, by explaining why the determination of difference-making is the first step of approaching a mechanistic explanation, this section shows that difference-making evidence, i.e. evidence of causal relevance, is required in biological research.

Manipulability is a dominant value in the practice of biological research as it requires 'hands-on' approaches to discovering things. I consider this trend to be closely tied to the pragmatic feature that Craver and Darden (2013) have discussed in their book on biological mechanisms. Contemporary biology is greatly inclined to serve as the basis for producing real-world impacts, such as therapeutics. The understanding of the components of mechanisms is expected to inform researchers of practical ways to produce the desired effects by controlling and manipulating these components. In this context, it is arguable that intervention is one of the few

² The literature on the philosophy of science has covered whether there are distinctions between repeating, reproduction and replication, as well as what exactly is to be reproduced. The distinctions exist because these actions can be applied to different subjects such as experimental procedures and data, where different disciplines also have different definitions of these actions and seek to fulfil different purposes by conducting the actions. Fidler and Wilcox (2018) document relevant literature in detail. Such discussions are far beyond the scope of this paper. Due to the obvious practice-focused feature of my study, I choose to follow the practitioners' terminology and consider these terms as interchangeable. Nonetheless, this paper uses 'repeatability' thoroughly for consistency. This is also the term that Woodward uses in his series of papers.

ways, if not the only way, to produce data and make phenomena observable in the search of biological mechanisms.³ Following up this view, I maintain that experimental inventions in biological mechanism research do not necessarily manipulate the direct causes of the events of interest. In addition, they do not necessarily aim to reveal a causal relationship between particular components.

Interventions are designed to determine the ways of particular components are *involved* in the final mechanistic explanations. By conducting what Woodward calls empirical investigations or what this paper generally calls ‘checking of data generation procedures’, biological mechanism researchers simultaneously learn about whether a set of data can become evidence for a phenomenon and whether the newly obtained evidence can serve as the difference-making evidence for a specific component of the mechanism in question. There is nonetheless a step missing from this description of the process, namely the determination of the validity of the difference-making revealed by the data produced during the interventions. I consider that the researchers need to know whether their interventions have been effective in terms of ensuring the validity of difference-making.

In practice, the checking of data repeatability and the determination of difference-making are usually conducted through ‘wrapped’ procedures in which two purposes are intertwined. One purpose is to obtain enough repeatable data, and the other is to consolidate the involvements of the components under investigation. The latter will contribute to the tentative conclusion of causal relevance, while the components are not necessarily productive causes of the final effects. The common procedures can be generalised as:

- (1) Obtaining data regarding the effect of an intervention.
- (2) Quantifying the data.
- (3) Statistical analysis of the quantified data.
- (4) A decision on whether the data is representative of the effect produced by the intervention based on whether this ‘observation’ can be consolidated.
- (5) Making claims on both the repeatability of the data and difference-making based on the consolidated observation.

Note that the actual procedures vary with the context of the research projects. Steps (1) and (2) are to check the repeatability of data, to confirm the exclusion of errors and to collect sufficient amounts of data for step (3). Step (4) refers to the inference from the statistical results of step (3). Finally, at step (5), both the observability of the phenomenon (i.e. the reproducibility of the experimental results) and the difference-making of interest are confirmed.

My point is that in the abstract sense, the observability of the phenomenon does not automatically warrant the validity of the difference-making. That is, the researchers ensure the effectiveness of their interventions by checking the quantity of reproduced

³ This paper sets aside the discussion of dry-lab practices, such as computational simulations and high-throughput screening. Here, I point out that although dry-lab practices appear to be remote from physical manipulation of biological materials, it is still arguable that computer-aided practices of biological mechanism research are manipulability-dominant. Dry-lab data are normally used to inform physical interventions, such as identifying promising components for intervention and designing new methods to conduct interventions. This shows that the ultimate interest of biological mechanism research is manipulability.

data before they determine the phenomenon observed to be a valid difference-making. The interventions themselves are also twofold: on the one hand, they manipulate a component of the whole mechanism to examine its possible involvement; on the other, they intervene in a causal process that produces the data. Yet, it does not directly contribute to the mechanism of interest. This is to say that the involvement of the component in the mechanism and the causation of it being observed in the experiment (i.e. the causation of Woodward's 'phenomenon') are independent of each other. The philosophical recognition of the details of this independence is not trivial, since biological mechanism research normally checks things that look nothing like the mechanisms of interest at a glance. Moreover, interventions on specific components are usually conducted via means that look nothing like the means relevant to the mechanisms of interest.

Now I use an example to show the above in a concrete way. A relatively complicated case study will be discussed in Sect. 3.3. The example here is raised from an established field of biological mechanism research: apoptosis (programmed cell death). The advantage of drawing on the apoptosis field is that both the theories and the techniques have reached their maturation in the past decades. A hallmark sign of apoptosis is the release of the molecule 'cytochrome c' from the mitochondria. The released cytochrome c forms a complex with some other molecules; then, the complex will initiate further processes that physically and more directly cause cell death (Mignotte and Vayssiere, 1998; Santucci et al., 2019; Ott et al., 2002). In other words, the release of cytochrome c is a necessary causal component of the final mechanistic explanation of apoptosis given that the formation of the complex produces apoptosis via physically connected events.

In apoptosis research, there are many ways to detect the release of cytochrome c. Measurements of potential changes in mitochondria are demonstrative of the independence between the causation of apoptosis and the causation responsible for data generation. In healthy cells, there is an electrical mitochondria membrane potential ($\Delta\Psi_m$) maintained by the electric polarisation that exists across the two sides of the mitochondrial membrane. In cells undergoing the apoptotic process, the release of cytochrome c causes an increase in the permeability of the mitochondrial membrane, which results in consequent depolarisation. Therefore, a common way of detecting cytochrome c release is to use charged fluorescent dyes, which steadily accumulate on one side of the mitochondrial membrane of healthy cells. During apoptosis, the dyes exhibit colour changes due to the collapse of $\Delta\Psi_m$.

If researchers want to know whether cytochrome c release is involved in a special kind of apoptosis induced by a hypothetical cause (C_a), they may treat the cells by intervening in C_a (I_a) and look to determine the effectiveness of this intervention by observing the colour change of a type of fluorescent dye. The observable phenomenon is the colour change, and researchers seek to consolidate this phenomenon by repeating the experiment with different groups of cells. These actions would be steps (1) and (2) in the above list, where researchers would aim to check if the experimental result (colour change) is reproducible. The purpose of this checking is the first part of the decision on the evidential status of the data, namely, the legitimacy and reliability of the data in terms of data generation.

Then, the determination of difference-making between $\Delta\Psi_m$ and I_a requires step (3)–(5). This will fulfil the purpose of assessing whether the data is reliable evidence for the mechanism of C_a -inducing apoptosis. In this part, the researchers would

determine whether I_a is effective enough to produce an apoptotic process that contains the release of cytochrome c. The effectiveness of I_a can now be determined based on the statistical significance of difference-making data, where the validity of this difference-making is simultaneously confirmed. By making colour changes in fluorescent dyes observable in a reproducible manner, which has been considered by previous studies as satisfying the elimination of errors and biases, the researchers would have *concurrently* obtained enough data for statistical analysis. The statistical analyses can determine if there is difference-making caused by I_a . After the difference-making has been determined, the researchers will treat ‘cytochrome c release’ as a component of the mechanistic explanation of C_a -inducing apoptosis. The causal process relevant to mitochondria depolarisation no longer matters.

At the stage of inferring a claim from the evidence, the collapse of $\Delta\Psi_m$, namely the local causation that is responsible for both assessing the repeatability of data and determining local difference-making, does *not* become a component of C_a -inducing apoptosis. This isolation of the local causation of data-generation from the causal inference to the final conclusion is normal in biological mechanism research. In this example, several background theories are required for the establishment, calibration and verification of the procedures for generating observable/repeatable images of colour changes in the dye. These theories contain a number of causal processes that can explain this colour change as an effect, including at least: the electric theory that explains the potential loss and movement of charged particles, the biochemical theory that explains the relation between increased permeability and particle movements and the electrochemical theory that explains the visualisation of fluorescence excited by light of particular wavelengths. A more detailed empirical investigation of the data generation procedures may also require knowledge of the biochemical theory that ensures that the dye really accumulates in the mitochondria as desired. Nonetheless, none of these theories supports the causal processes that are directly responsible for C_a -inducing apoptosis. The depolarisation of the mitochondrial membrane does not contribute to the effect (apoptosis) that is hypothetically caused by C_a . The researchers thus would not seek to examine the involvement of membrane depolarisation in the mechanistic explanation of C_a -inducing apoptosis. Nor do they, obviously, seek to determine the involvement of colour changes in the dye as causally relevant to apoptosis, as colour change is just an operational indicator of depolarisation.

The above example aims to show how the assessment of the effective interventions can simultaneously be isolated from the final causation of interest and contribute to revealing relevant causes. The recognition of these dual statuses of the assessment of interventions arguably helps to solve the problem of vicious circularity.⁴ Woodward has stated in his response to some misunderstandings of the ‘involvement’ of theory in the reasoning process from data to phenomena⁵: “[...] this sort of involvement of T in data to phenomena reasoning does not necessarily mean that T

⁴ This paper discusses biology only and is not concerned with the scenario that circularity might still challenge the interventionist account for knowledge in some other sciences.

⁵ Regarding this quote, D = data, P = phenomena, and T = theory, I need to make clearer that what Bogen and Woodward call ‘phenomena’ is very close to (if not the same as) ‘experimental result’ in this paper. My use of the term is intended to be consistent with biologists’ common terminology. In the passages where I cite and directly use Woodward’s theses, I adopt his term ‘phenomena’.

is being used to explain D or that D cannot be evidence for P unless T is conceived as playing this explanatory role” (2011, 177). This distinction between a theory (T) being involved in the data-phenomena reasoning and T explaining the data is particularly obvious in biological mechanism research because the determination of the components of the mechanisms of interest normally involves a number of theories that have nothing to do with the theory that supports the researchers’ confidence in designing interventions on the components. I have used the example of cytochrome c release to show this distinction. In the proposed structure (Fig. 1), $T_{1,2,3\dots}$ are the theories that explain the electric, chemical and biochemical reactions leading to the visualisation of fluorescence colour changes. T_a is the theory that makes the researchers confidently think C_a -inducing apoptosis is possible. The main idea is not that T_a and $T_{1,2,3\dots}$ are independent of each other but that the *uses* of these theories are independent of each other.

I have proposed to put intervention in the structure of understanding how data becomes the evidence of difference-making in biological mechanism research. The next section will elaborate on how the causal independence between data generation and the final conclusion also contributes to the legitimacy of evidence in terms of confirming reliable mechanisms.

3 Various independent evidence for reliable mechanisms

This section extends the existing literature on evidential variety and robustness by pointing out an epistemic demarcation between different experimental interventions. The main idea of this section is to create a distinction between (a) the independence between experimental techniques and (b) the independence between experimental interventions. To ensure that the variety of independent evidence supports the reliability of the biological mechanisms, an appropriate philosophical view for evidence-gathering methods should refer to interventions and not techniques. The independence of interventions is a safer and more complete form of ‘ontic independence’ among diverse evidence regarding robustness analysis. I will suggest that this form of evidential independence is a result of the nature of biological mechanism research.

3.1 Independent evidence and robustness

The previous section has introduced the two necessary and complementary kinds of evidence for biological mechanisms. I have described procedures of both assessing data repeatability and confirming valid difference-making where quantity is used as a criterion. This section will discuss how data are given the status of the other kind of evidence, i.e. evidence of mechanism (Illari, 2011) or mechanistic evidence (Russo & Williamson, 2007), namely evidence directly contributing to causal inferences regarding the mechanism. Diverse data obtained from biological interventions need to satisfy the requirement of independence so that they gain the epistemic status of being the ‘evidence of mechanism’. I argue that in biological research, the

independence between different interventions conducted to obtain evidence in the examination of a single component of the mechanism has richer abstract meanings than what previous studies have considered when referring to it as ‘ontic independence’ (OI) among evidence-gathering methods.

I borrowed some concepts from the existing frameworks proposed by Kuorikoski and Marchionni (2016) and Stegenga and Menon (2017). These works have extensively elaborated on the strengths and weakness of many conditions where different types of evidential independence are expected to warrant robust arguments in the logical sense. I maintain that it is philosophically fruitful to study the epistemological role of interventions in biological mechanism research not only in the empirical sense but also in the formal sense. Stegenga and Menon examine various conditions in which causal inferences made from seemingly independent evidence can fall short of the ideal Bayesian scenario of robustness. Kuorikoski and Marchionni defend the reliability of causal claims inferred from the triangulation of multiple kinds of independent evidence. I recognise that these two accounts of the epistemic value of independent evidence are not fully compatible with each other. However, I consider that appreciating the independence between interventions helps to reconcile these two theses.

What I have shown in Sect. 2, i.e., the independence both (1) between the causation of data generation procedures and the final causation under investigation and (2) among various kinds of causal processes that are responsible for obtaining diverse kinds of evidence, seems to only match ontic independence (OI). OI means that the materials and theories, and thus potential biases and systematic errors, of diverse experiments are independent of each other. Stegenga and Menon question OI by discussing some situations of ‘pseudorobustness’ where their biological examples seem to be very convincing. To introduce the problems briefly, there are two scenarios of robustness being in danger. The key point regards whether a hypothesis (H) ‘d-separates’ two consequences (two variables). D-separation means that, ideally, in the only trail connecting the two consequences, H is neither a collider itself nor containing a collider, where neither of the two consequences is an ancestor of any variable in H. The first scenario of pseudorobustness is that an auxiliary assumption influences the evidence of both consequences. The second scenario is dyssynergy. That is, the evidence together is less confirmatory than either of them, where the two consequences may be causally dependent. Following these problems, Stegenga and Menon acknowledge the possibility that conditional probabilistic independence (CPI) does provide some more support for a hypothesis, and they have also clarified that such support does not equal warranting realism regarding a hypothesis. I fully agree with the realism part. Yet, I will point out that independence between different experimental interventions supports biological mechanisms to a greater extent than CPI because the OI of evidence in biological mechanism research itself does not exhibit the weakness that CPI can supposedly avoid.

3.2 Gaining workable knowledge via decomposition

I suggest that it is important for philosophers to appreciate two features of biological mechanisms in the interventionist aspect: one is the practice of decomposition

(Bechtel, 2006), and the other is the importance of confirming the workability of mechanisms. They manifest the nature of biological mechanism research that the existing literature has elaborated on. Below I first discuss the workability of mechanisms, which highlights the importance of the overall outcome of the organisation of components, and then explain why decomposition deserves an in-depth discussion in the study of evidence.

Biological mechanisms, though being theoretical and idealised frameworks,⁶ can be in some senses analogous to real machines and thus have to be ‘workable’. That is, when constructing a mechanistic explanation, researchers need to confirm the capability of the mechanism to produce the effects of interest. Daston and Galison’s notion of ‘workable knowledge’ (2007) contributes to my extended interventionist account of biological evidence, though the focus of their influential book is by no means on biological mechanisms. The main idea of their discussion about the role of interventions in many sciences follows Ian Hacking’s interventionist view and is broadly in the Baconian sense. That is, in my interpretation, the confirmation of a scientific object being real depends on whether the object works. This idea comes up in Daston and Galison’s arguments about the increasingly engineering-oriented science since the late twentieth century where they maintain that the ethos of scientific research has been increasingly a scientific-engineering hybrid. In this context, researchers use scientific objects, such as simulation images, to both produce real-world effects and persuade the audience to believe and use novel theories. One needs to intervene in something to know both (1) that it works and (2) how it works, and the knowledge of how this thing works should contribute to the action of making it work in the desired ways. I consider this orientation to describe well the research trajectories of contemporary biology.

Craver and Darden’s view of the pragmatic value of biological mechanism research explicitly reveals this emphasis upon real-world effects. By combining Daston and Galison’s notion of workable knowledge and Craver and Darden’s view of pragmatic value, we can now consider explaining and controlling biological effects in a mechanistic way in accordance with the two sides of the realist understanding of these effects:

It was a strong salvo in a long-standing debate over whether and under what conditions scientific objects may be taken as real. On the side of representation: we should take as real that which offers the best explanations. On the side of intervention: we should accept as real that which is efficacious. (Daston & Galison, 2007, 392)

I argue that the development of biological mechanistic explanations relies significantly on the confirmation of the workability of the mechanisms and that this

⁶ This paper does not deal with the realism of biological mechanisms. My basic position is that while biological mechanisms indeed produce physical effects and can be both investigated and manipulated physically, they themselves are theoretical and idealized frameworks established to explain biological events. Such theorisation and idealisation are obvious in the way that their boundaries are highly flexible and adjustable according to the purpose of the study (Bechtel, 2015). To put my position in an oversimplified way, I consider biological mechanisms to be conceptual constructs composed of physical components.

confirmation is conducted via interventions in the components. Since I consider mechanisms as mechanistic *explanations* of the phenomena of interest, disambiguation is needed here in terms of what exactly is the thing that must be workable. Is it the explanation or the effect produced by the mechanism? My answer is the former. Researchers confirm that a mechanism (i.e. a mechanistic explanation) works by confirming that the effect (phenomenon) presents in the way they have expected. Section 2 has described how the phenomenon is confirmed by assessing the evidential status of data in terms of difference-making. Yet, biological research needs another kind of evidence, namely the evidence of mechanism, to know *how* the mechanism works and thus has an explanatory function.

I argue that in biological research, the confirmation of such workability does not necessarily require interventions in the causal processes of the final mechanisms. Section 2 has discussed independence between the causation of data generation and the causation of final interest. Now, I take a step further to suggest that different experimental interventions for exploring the same component are intended to test different and causally independent hypotheses. Thus, the reliability of the final mechanistic claims is exempt from the challenges regarding pseudorobustness. This exemption is a result of the mechanistic thinking style of biological research. Such a thinking style shapes the ways that biological effects are understood, speculated about, intervened in and explained, thus leading to the decomposition method of obtaining evidence for causal inference. I consider the independence between diverse biological evidence to result from the decomposition mode of testing the components because there is a temporary causal separation between the component and the target mechanism. In the vein of Bechtel (2006), I view decomposition as both conceptual and physical. Conceptual decomposition means that the researchers deal with parts of the mechanisms separately, such as speculating about, understanding and making hypotheses about them without physical interventions. Physical decomposition means that the researchers investigate the components by physically manipulating them.

Consider the study of a real machine as an analogy. If one wants to know whether the connection between an engine and the wheels of a carriage causes movement of the carriage in a particular way (i.e., the final effect awaiting explanation) when a specific kind of material is used in making the wheels, what they check is whether the special movement of the machine occurs when the condition of ‘the wheels being made of that material’ presents. The workability of the explanation (that the wheels made of that material play a role in the special movement of the carriage) and the confirmation of the effect (i.e. the special movement) are tied to each other. In the scenario that the movement occurs in the expected way, one can decide that the wheels made of the specific material are the objects of investigation because they plausibly contribute to the causation of interest.

Then, in a decomposition mode of investigation, one can examine the components of the machine separately. Each examination can be derived from a theory that is *both* independent from the background theories of the other examinations and independent of the theory of the machine as a whole. When the researchers study the friction between the wheels and the surface, the wheels need not be driven by the same engine so long as the power provided is the same. The friction can be

studied in many different settings if the conditions for producing the same form of movement of the carriage are satisfied. Temporarily, the power of the movement of the carriage is assigned to be an operational local cause, no matter what the power source is. This temporary separation between the actual cause and the effect can be applied to the examination of other parts of the machine. In other words, when one claims the workability of this machinery, they normally do not worry if their evidence for understanding the wheels depends on the presence of the whole machine. Although the final confirmation of this workability requires the assembly of the wheels and the engine, and although the assembly will require some extra knowledge of the coherent relationships between the engine, the wheels and other parts of the machine, these requirements do not contradict the operational setting in which the wheels were once studied independently of any theory that supports the performance of the whole machine.

Such decomposition is very similar to how biological mechanisms are studied. The previous example of ‘cytochrome c release’ implies this idea. In the case of apoptosis, an increase in mitochondrial membrane permeability is used as a temporary and operational cause, which helps researchers detect and obtain evidence about cytochrome c release. In *this* experimental intervention, the increase of mitochondrial permeability need *not* be causally associated with apoptosis (the final effect). Once the involvement of ‘cytochrome c release’ has been confirmed as a component of the target mechanism of apoptosis, the final assembly of this and other components will be required to establish the conclusion. This is similar to the final assembly of the car machine. This requirement for final assembly does not contradict the operational setting that ‘the involvement of cytochrome c release’ in the target mechanism is *once* studied independently of any theory that supports the causal relationship between cytochrome c release and the apoptotic effect in question.

Because of this decomposition mode, experimental interventions designed to obtain evidence about a component can be and normally are independent of each other in terms of instrument, material and background theory. I have mentioned the worry that according to Stegenga and Menon, this sort of independence is OI and can be violated by either dyssynergy or auxiliary assumptions. These possible failures suggest that the robustness of causal conclusions supported by OI is at risk. This paper does not aim to develop a unifying account for the independence of diverse evidence across sciences. Instead, this paper focuses on biology and argues that the decomposition mode for studying mechanisms guarantees a ‘safer’ and more complete form of OI. This paper also notes that the independence between diverse kinds of evidence in biological mechanism research both satisfies and yet cannot be categorised as Stegenga and Menon’s CPI.⁷ This is because the conditions for CPI are

⁷ As this paper does not focus on the probabilistic analysis of evidence, I quote Stegenga and Menon’s conditions for CPI in this footnote:

- i. $\Pr(H|E1) > \Pr(H)$,
- ii. $\Pr(H|E2) > \Pr(H)$,
- iii. $\Pr(E1 \& E2|H) = \Pr(E1|H) \times \Pr(E2|H)$,
- iv. $\Pr(E1 \& E2 | \sim H) = \Pr(E1 | \sim H) \times \Pr(E2 | \sim H)$.

For CPI, all four conditions need to be satisfied to ensure the avoidance of pseudorobustness.

set up to avoid pseudorobustness of the conclusion only when OI is at risk. Since the OI of evidence in biological mechanism research is itself satisfactory, characterising such independence as CPI is unnecessary.

3.3 Case study: independent interventions

Now I analyse a slightly complicated case study to demonstrate the proposed distinction between (a) the independence between various experimental interventions and (b) the independence between various experimental techniques. I argue that this distinction explains why the triangulation of various pieces of evidence (as defended by Kuorikoski and Marchionni) supports the reliability of biological mechanisms. As reiterated, this independence is a safer and more complete form of OI because different interventions are intended to test independent local hypotheses. Although the local hypotheses are nested under the main hypothesis (i.e., the mechanism as a whole), the final conclusion inferred from the evidence is still reliable.

In Joffre et al. (2015), the researchers sought to know if a protein kinase, STK38, is involved in the mechanism of autophagy. Their first step was to determine whether STK38 interacts with any part of a complex that is known necessary for autophagy, i.e. the Exo84-Beclin complex. The main hypothesis about the mechanism is:

H_m : STK38 interacts with some part of Exo84-Beclin complex during autophagy.

If H_m is true, the researchers would be confident in drawing a conclusion regarding the mechanism that STK38 is positively involved in the causal mechanism of autophagy. That is, STK38 is a component of the mechanistic explanation of autophagy. The background theory directly relevant to this mechanism includes the following parts:

- (a) STK38 activation has a positive role in autophagy
- (b) Exo84-Beclin is necessary for autophagy
- (c) Beclin has the potential to bind STK38
- (d) Exo84 is necessary for autophagosome formation
- (e) Autophagosome formation produces autophagy

Here, (c) is used to test H_m , and the other parts are used to design the interventions.

I skip the analysis of how the researchers determined valid difference-making phenomena by checking the effectiveness of interventions. This section concentrates on how the researchers gained the confidence to claim a mechanism connecting STK38 and autophagy.

At a glance, both the data and the interventions seem to be subject to technique-based classification. According to the original paper (Joffre et al. 2015, Fig. 1), there are three sets of evidence classified by techniques: immunoprecipitation (IP), western blotting (WB), and confocal microscopic imaging of immunofluorescence staining. We have:

Evidence 1 (E1)=IP data, which is used to infer the phenomenon ‘STK38 binds to Beclin’

Evidence 2 (E2)=WB data, which is used to infer the phenomenon ‘STK 38 binds to Beclin’

Evidence 3 (E3)=confocal microscopic imaging data, which is used to infer the phenomenon ‘STK38 and Beclin co-localise in the cell’

The three sets of evidence are used to respectively consolidate two phenomena that are the consequences of H_m :

C1 = STK38 binds to Beclin

C2 = STK38 and Beclin co-localise

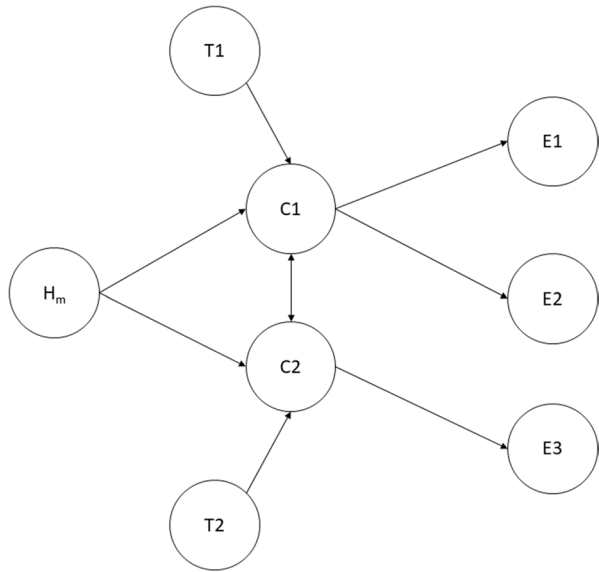
The applications of these three techniques are causally independent of each other because of their different and independent materials and background theories. Woodward’s ‘empirical investigations’ on the techniques appear sufficient to ensure the independence between the three sets of evidence obtained from the applications of these techniques. This is how Kuorikoski and Marchionni use empirical investigations to argue that different experiments are independent of each other in terms of systematic errors and biases:

First, if the processes of data generation (methods) are causally independent (being based on different kinds of causal mechanisms), then any token random causal disturbance of one method should not have an effect on the other method. Second, if the methods are based on different kinds of causal processes, the presence of any systematic error (bias) in one method should not affect the probability of an error occurring in the other. (2016, 232)

I fully agree with (1) their use of Schupbach’s (2018) ‘reliability independence’ for describing this sort of independence and (2) that the notion of reliability independence captures a significant part of the reason why the triangulation of various evidence obtained from applying different techniques guarantees the reliability of the conclusion. However, this evidential independence, as well as the reliability of the conclusion thereby guaranteed, can still be challenged because the abovementioned possibilities of pseudorobustness have not been eliminated.

Figure 2 visualises how these three sets of evidence and phenomena are fitted into the Bayesian network of the ideal scenario for robustness. This structure has an obvious problem that C1 and C2 seem to be interrelated or even interdependent: if STK38 binds to Beclin, they should appear in the same location; if STK38 and Beclin co-localise, it is possible that they bind to each other under particular circumstances. Hence, the probability of E1 and the probability of E2 are very likely to affect each other. In this sense, the independence between T1 and T2 (i.e. the two underlying theories of the techniques) does not guarantee the independence of E1 and E2 from each other. The ontic independence between the evidence obtained via employing IP, WB and confocal microscopy does not help warrant a robust conclusion. Nor does the independence between different applications of the same techniques support the reliability of the conclusion.

Fig. 2 Network representation of evidence classified by techniques of the case study



To solve this, I propose not only empirically but also formally ensuring the causal independence between various sets of evidence by considering the role of intervention in the practice. The modified structure (Fig. 3) of triangulation is that various pieces of evidence obtained in different interventions exhibits promising independence between each other. I classify the pieces of evidence by the ways they are actually obtained and used in the scientific practice:

C1 = phenomenon 1 (P1) = the thing to be captured after intervention 1 = interaction between Beclin and wild-type STK38 in HEK293 cells, where autophagy does *not* occur.

C2 = phenomenon 2 (P2) = the thing to be captured after intervention 2 = interaction between Beclin and overexpressed STK38 in HEK293 cells, where autophagy does *not* occur.

C3 = phenomenon 3 (P3) = the thing to be captured after intervention 3 = STK38-Beclin co-localisation in Hela cells, where autophagy *occurs*.

The above definitions of $C_{1,2,3}$ show that, in both the scientific and the philosophical senses, they are actually three independent hypotheses to be tested. The presence of the final effect (autophagy) is not necessary for all these tests. I name the three independent hypotheses as ‘local hypotheses’ and represent them as $HL_{1,2,3}$. Being local only means that they are nested under H_m —this naming is only to keep H_m in the structure because the confirmation of H_m eventually leads to the final conclusion. What makes $HL_{1,2,3}$ independent of each other is that they do not need to be tested in the presence of the condition for H_m . Importantly, the probabilistic independence between $C_{1,2,3}$ results from not only this independence between

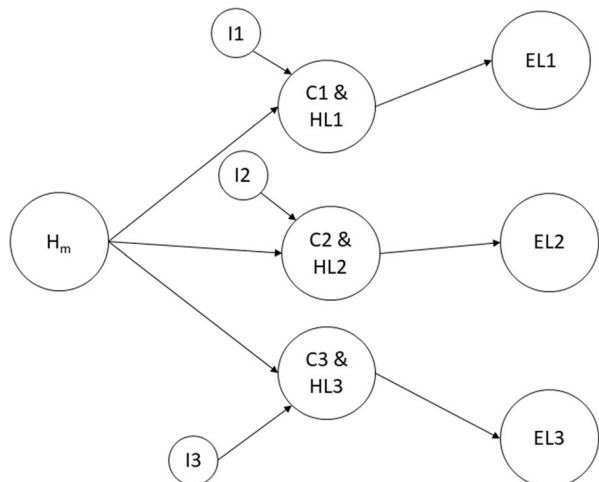
$HL_{1,2,3}$ but also that they could have been the results of three independent projects. For example, one could conduct a small project that only tests whether P1 can be consolidated without intending to test P2 and P3 because P2 and P3 do not affect the occurrence of P1 whether or not the conditions remain.

The above description might sound redundant to some, as it is close to the notion of ontic independence. Figure 3 does not seem to solve the worry about pseudorobustness. In response, I emphasise that the techniques and materials are independent precisely because the researchers could aim to intervene in three different and independent phenomena. That is, the ontic independence between techniques and materials is a result of the independence (1) between the local hypotheses and (2) between the research goals of the fictional ‘small projects’, not the reason why the consequences are independent of each other. The ontic independence between techniques and materials does not need to guarantee the robustness of the conclusion because it is a by-product of the independent interventions.

There are two additional points about this case. First, cell types in biology represent physiological environments in which the events of interest are to be observed. Meanwhile, the two independent sources of STK38 ensure that the physiological behaviour of STK38 is neither mutually affected nor shaped by the presence of autophagy. In Fig. 3, $P_{1,2,3}$ are observed after $I_{1,2,3}$, and $I_{1,2,3}$ do not affect each other because any of them could be isolated from this research project and contribute to another new and irrelevant study. While the employment of different techniques indeed guarantees the empirical independence between the different sets of evidence, it is the independence between the interventions for testing different local hypotheses that guarantees the probabilistic independence between the diverse sets of evidence.

The case study has shown that Fig. 3 is the more appropriate illustration of the practice. The interventions are independent because of the mode of decomposition for investigating biological mechanisms. The previous engine–wheels analogy is useful for clarifying my point: the examination of the friction between the wheels

Fig. 3 Network representation of evidence classified by interventions of the case study



and the surface can be isolated from the study of the wheels as a part of the whole machine. In biological research, phenomena observed in one cell type (or a type of physiological environment) does not rely on the occurrence of the phenomena in another cell type, despite the possibility that the techniques for detecting them may be the same. Later, I will answer a possible objection to this statement. My point is simply that these local hypotheses are independent of each other not merely because the confirmation of them relies on different techniques. The distinction between ‘independence between different techniques’ or ‘independence between different applications of the same technique’ and *independence between different interventions* explains why they render two different forms of OI.

Note that in the structure presented by Fig. 3, the researchers could obtain more kinds of evidence for each local hypothesis, such as $EL_{1a, 1b, 1c...}$ and $EL_{2a, 2b, 2c...}$. The reliability of the local conclusions about the local hypotheses is supported by at least two actions. First, as the previous studies have explained well, the researchers conduct empirical investigations to avoid shared errors and biases. Second and importantly, since both these local hypotheses and the interventions were inspired by established knowledge, the researchers can always gather extra evidence from the literature and are confident about the independence between the existing data and their data. This practice is normal and is explicitly stated in countless biological papers.

Finally, I answer a possible objection to my statement that ‘phenomena observed in one cell type (or a type of physiological environment) does not rely on the occurrence of the phenomena in another cell type’. Since the design of interventions for biological research tends to be derived from, or at least inspired by, established knowledge, and since the assumptions that support the design are at least partially based on deductive reasoning, one might worry whether the probability of the phenomenon observed in the new intervention depends on, or is affected by, the probability of those established phenomena. I consider this worry unnecessary and that it runs the risk of denying biological research in general. For example, if one wants to know whether protein A participates in cellular autophagy in species X (i.e., physiological context X), where they design an intervention by inferring from the established phenomenon that protein A participates in cellular autophagy in species Y (i.e., physiological context Y), it would be unreasonable to question this intervention by assuming that autophagy in X and in Y exhibit probabilistic dependence on each other. Such a doubt nearly necessitates requiring the studies of X and Y to be totally unrelated in terms of any way that the philosophers understand scientific practices. This is similar to assuming that the knowledge established by studying Y is useless for the study of any other species. These doubts are not helpful to the philosophy of biology.

4 Conclusion

This paper has argued for an extended version of the interventionist account for understanding the epistemic values of two criteria (quantity and variety) for deciding the evidential status of experimental data in biological mechanism

research. I argued that these two requirements can be considered as checking the effectiveness of interventions. Interventions have twofold meanings from the beginning. On the one hand, the researchers intervene in the causation responsible only for local data generation; on the other, they intervene in the causation responsible for the production of the final effect.

Based on these two meanings, I have argued that the causal independence between various sets of evidence obtained via independent interventions ensures the reliability of evidence. I have emphasised that the independence between techniques is not equal to the independence between interventions, where the latter is a more complete form of ontic independence. Therefore, the variety of evidence that ensures the robustness of causal conclusions results from two kinds of causal independence. The first is causal independence between various theories on which interventions are based. The second is the causal independence between the theories supporting interventions and the theory supporting the final conclusion.

I have also sought to solve the black-box problem of calling the process of checking the reliability of data ‘empirical investigations’. I have argued that investigations of experimental setting ensure this reliability because the causation underlying the difference-making of interest and the causations underlying interventions can be independent of each other. This is to say that causal independence is also important to the determination of difference-making. The requirement for quantity can be understood in this aspect. An adequate quantity of data is used to determine the validity of difference-making. In practice, this determination takes place before the process of obtaining various independent evidence. The minor point is that I have speculated that the practice of decomposition may have resulted from the engineering-oriented, pragmatic and hands-on features of biology since the late twentieth century.

This paper suggests that the abstract understanding of biological knowledge production can be improved by examining the practices of intervention. By detailing how physical manipulations of biological materials contribute to theories in the logically valuable sense, these results hopefully will serve as a basis for further studies that will substantiate this extended interventionist account.

Funding This study was partially funded by the Ministry of Science and Technology, Taiwan. Grant number MOST 109-2511-H-006-013-MY2.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Bechtel, W. (2006). *Discovering Cell Mechanisms: The Creation of Modern Cell Biology*. Cambridge University Press.
- Bechtel, W. (2015). Can mechanistic explanation be reconciled with scale-free constitution and dynamics? *Studies in History and Philosophy of Science Part C: Studies in History and Philosophy of Biological and Biomedical Sciences*, 53, 84–93.
- Bogen, J., & Woodward, J. (1988). Saving the phenomena. *Philosophical Review*, 97(3), 303–352.
- Craver, C. F., & Darden, L. (2013). *In Search of Mechanisms: Discoveries across the Life Sciences*. University of Chicago Press.
- Daston, L., & Galison, P. (2007). *Objectivity*. Routledge.
- Dragulinescu, S. (2017). Mechanisms and difference-making. *Acta Analytica*, 32(1), 29–54. <https://doi.org/10.1007/s12136-016-0292-1>.
- Fidler, F., & Wilcox, J. (2018). Reproducibility of scientific results. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy (Winter 2018 Edition)*. <https://plato.stanford.edu/archives/win2018/entries/scientific-reproducibility/>. Accessed 18 Jan 2020.
- GleNNan, S. (2002). Rethinking mechanistic explanation. *Philosophy of Science*, 69(S3), S342–S353. <https://doi.org/10.1086/341857>.
- 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. <https://doi.org/10.1086/662561>.
- Howick, J., & Glasziou, P. (2013). Problems with using mechanisms to solve the problem of extrapolation. *Theoretical Medicine and Bioethics*, 34. <https://doi.org/10.1007/s11017-013-9266-0>.
- Illari, P. M. (2011). Mechanistic evidence: Disambiguating the Russo-Williamson thesis. *International Studies in the Philosophy of Science*, 25(2), 139–157. <https://doi.org/10.1080/02698595.2011.574856>.
- 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. <https://doi.org/10.1007/s13194-011-0038-2>.
- Joffre, C., Dupont, N., Hoa, L., Gomez, V., Pardo, R., Gonçalves-Pimentel, C., Achard, P., et al. (2015). The pro-apoptotic STK38 Kinase is a new Beclin1 partner positively regulating autophagy. *Current Biology*, 25(19), 2479–2492. <https://doi.org/10.1016/j.cub.2015.08.031>.
- Kuorikoski, J., & Marchionni, C. (2016). Evidential diversity and the triangulation of phenomena. *Philosophy of Science*, 83(2), 227–247. <https://doi.org/10.1086/684960>.
- Mignotte, B., & Vayssi re, J.-L. (1998). Mitochondria and apoptosis. *European Journal of Biochemistry*, 252(1), 1–15. <https://doi.org/10.1046/j.1432-1327.1998.2520001.x>.
- Ott, M., Robertson, J. D., Gogvadze, V., Zhivotovsky, B., & Orrenius, S. (2002). Cytochrome c release from mitochondria proceeds by a two-step process. *Proceedings of the National Academy of Sciences*, 99(3), 1259. <https://doi.org/10.1073/pnas.241655498>.
- Russo, F., & Williamson, J. (2007). Interpreting causality in the health sciences. *International Studies in the Philosophy of Science*, 21(2), 157–170. <https://doi.org/10.1080/02698590701498084>.
- Santucci, R., Sinibaldi, F., Cozza, P., Polticelli, F., & Fiorucci, L. (2019). Cytochrome c: An extreme multifunctional protein with a key role in cell fate. *International Journal of Biological Macromolecules*, 136, 1237–1246. <https://doi.org/10.1016/j.ijbiomac.2019.06.180>.
- Schupbach, J. N. (2018). Robustness analysis as explanatory reasoning. *The British Journal for the Philosophy of Science*, 69(1), 275–300. <https://doi.org/10.1093/bjps/axw008>.
- Stegenga, J., & Menon, T. (2017). Robustness and independent evidence. *Philosophy of Science*, 84(3), 414–435. <https://doi.org/10.1086/692141>.
- Woodward, J. (1989). Data and phenomena. *Synthese*, 79(3), 393–472.
- Woodward, J. (2000). Data, phenomena, and reliability. *Philosophy of Science*, 67, S163–S179.
- Woodward, F. (2011). Data and phenomena: A restatement and defense. *Synthese*, 182(1), 165–179. <https://doi.org/10.1007/s11229-009-9618-5>.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.