

★

## Establishing causal claims in medicine

★

Jon Williamson

Draft of December 14, 2016

### Abstract

Russo and Williamson (2007) maintain that in order to establish a causal claim in medicine, one normally needs to establish both that the putative cause and putative effect are appropriately correlated and that there is some underlying mechanism that can account for this correlation. I argue that, although this thesis conflicts with the tenets of contemporary evidence-based medicine (EBM), it offers a better causal epistemology than that provided by EBM because it better explains two key aspects of causal discovery. First, it better explains the role of clinical trials in establishing causal claims. Second, it provides a better account of extrapolation.

### §1

#### An epistemological thesis

Russo and Williamson (2007, §§1–4) put forward an epistemological thesis that can be phrased as follows:

In order to establish a causal claim in medicine one normally needs to establish two things: first, that the putative cause and effect are appropriately correlated; second, that there is some mechanism which explains instances of the putative effect in terms of the putative cause and which can account for this correlation.

This epistemological thesis, which has become known as the Russo-Williamson thesis or RWT, has generated some controversy—see, e.g., Weber (2007, 2009); Campaner (2011); Clarke (2011); Darby and Williamson (2011); Gillies (2011); Illari (2011); Howick (2011a,b); Russo and Williamson (2011a,b); Campaner and Galavotti (2012); Claveau (2012); Dragulinescu (2012); Clarke et al. (2013, 2014) and Fiorentino and Dammann (2015). The aim of this section is to explain what the thesis says and why it is controversial. In §2, I argue that an approach to medical methodology based on RWT fares better than present-day EBM in explaining three basic facts about how clinical studies can be used to establish causal claims in medicine. In §3, I argue that it gives a better account of extrapolation inferences too.

First, let us clarify some of the terms that occur within the statement of the thesis. ‘Mechanism’ here can be understood broadly as referring to a complex-systems mechanism, a mechanistic process, or some combination of the two. A

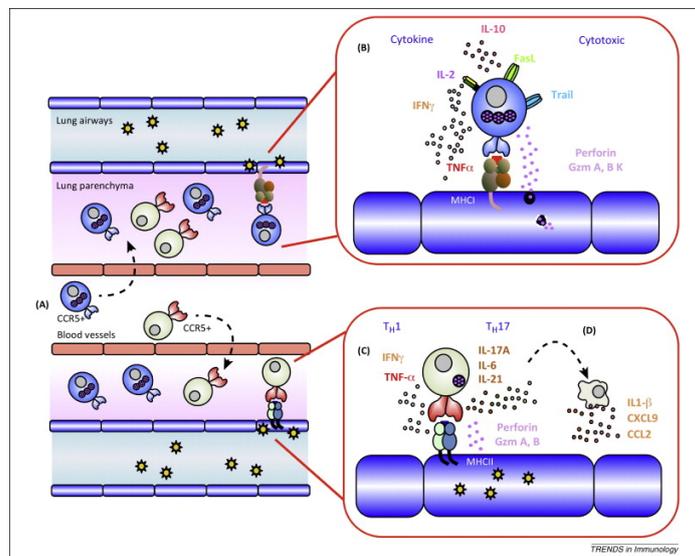


Figure 1: T cell effector mechanisms in a lung infected by influenza A virus (Gruta and Turner, 2014).

*complex-systems mechanism* consists of entities and activities organised in such a way that they are responsible for some phenomenon to be explained (Machamer et al., 2000; Illari and Williamson, 2012). An example is the mechanism by which the heart pumps blood. A *mechanistic process* is a spatio-temporally contiguous process along which a signal is propagated (Reichenbach, 1956; Salmon, 1998). An example is an artificial pacemaker's electrical signal being transmitted along a lead from the pacemaker itself to the appropriate part of the heart. A mechanism might also be a composed of both these sorts of mechanisms: for example, the complex-systems mechanism of the artificial pacemaker, the complex-systems mechanism by which the heart pumps the blood and the mechanistic process linking the two.

Note that a mechanism cannot in general be thought of simply as a causal network. A causal network can be represented by a directed graph whose nodes represent events or variables and where there is an arrow from one node to another if the former is a direct cause of the latter. On the other hand, a mechanism is typically represented by a richer diagram, such as that of Fig. 1, in which organisation tends to play a crucial explanatory role. Organisation includes both spatio-temporal structure and the hierarchical structure of the different levels of the mechanism. To take an extreme example of the importance of organisation, a chimney mechanism is responsible for the extraction of smoke purely in virtue of its spatial organisation. No activities constitute the chimney mechanism itself—although smoke actively passes through the mechanism—and the only relevant properties of the entities that constitute the mechanism (e.g., bricks and mortar) are structural properties to do with their impermeability and their ability to support the load of the chimney.<sup>1</sup>

<sup>1</sup>Kaiser (2016) provides further evidence for the claim that a mechanism cannot always be identified with a causal network.

Table 1: Examples of sources of evidence of mechanisms in medicine (Clarke et al., 2014).

---

Direct manipulation: e.g., in <i>vitro</i> experiments
Direct observation: e.g., biomedical imaging, autopsy, case reports
Clinical studies: e.g., RCTs
Confirmed theory
Analogy: e.g., animal experiments
Simulation: e.g., agent-based models

---

Note too that high-quality evidence of mechanism can be obtained by a wide variety of means. Table 1 provides some examples.

Let us turn to some other terms that occur in the epistemological thesis RWT. A causal claim is ‘established’ just when community-wide standards for granting the causal claim are met. This requires not only high confidence in the truth of the claim itself, but also high confidence in its stability, i.e., that further evidence will not call the claim into question. One theory of evidence holds that evidence consists of propositions that are rationally granted (Williamson, 2015); in which case, when a causal claim is established it can be treated as evidence for other claims. Establishing is thus crucial to our inferential practice. Establishing requires meeting a high epistemological standard. In particular, establishing a causal claim should be distinguished from acting in accord with a causal claim as a precautionary measure: in certain cases in which a proposed health action has a relatively low cost, or failing to treat has a high cost, it may be appropriate to initiate an action even when its effectiveness has not been established, so that benefits can be reaped in case the effectiveness claim turns out to be true. Despite being a high epistemological standard, establishing—unlike knowing, for instance—is not factive. One can say ‘They established that stress is the principal cause of stomach ulcers but further investigations showed they were mistaken,’ although one cannot substitute ‘knew’ for ‘established’ in this sentence; one would need ‘thought they knew’ instead.

The epistemological thesis says that one needs to establish that the putative cause and effect are ‘appropriately correlated’. Here ‘appropriately correlated’ just means *probabilistically dependent conditional on potential confounders*, where the probability distribution in question is relative to a specified population or reference class of individuals.<sup>2</sup> Thus, if  $A$  is the putative cause variable,  $B$  the putative effect variable and  $C$  is the set of potential confounder variables, one needs to establish that  $A$  and  $B$  are probabilistically dependent conditional on  $C$ , often written  $A \not\perp B | C$ . A confounder is a variable correlated with both  $A$  and  $B$ , e.g., a common cause of  $A$  and  $B$  (Fig. 2). The dependence needs to be established conditional on confounders because otherwise an observed correlation between  $A$  and  $B$  might be attributable to their correlation with  $C$ , rather than attributable to  $A$  being a cause of  $B$ . The set of *potential* confounders should include any variable that plausibly might be a confounder, given evidence of the area in question. Establishing

---

<sup>2</sup>‘Correlated’ is often used in a weaker sense, meaning unconditionally probabilistically independent, or even unconditionally linearly dependent. The arguments of this paper also go through under these weaker interpretations of ‘correlated’: if, under a strong reading of ‘correlation’, it is not enough simply to establish correlation to establish causation, then that is also true under a weak reading.

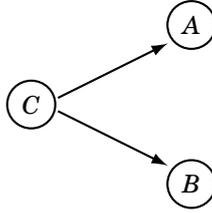


Figure 2:  $C$  is a confounder.

correlation is non-trivial for two reasons. First, because it requires establishing a probabilistic dependence in the data-generating distribution, rather than simply in the distribution of a sample of observed outcomes. The method of sampling and size of sample can conspire to render an observed sample correlation a poor estimate of a correlation in the population at large. Second, establishing correlation requires considering all potential confounders, and there can be very many of these.

To be clear, we shall use ‘observed correlation’ to refer to a correlation found in the data, ‘genuine correlation’ to refer to a correlation in the target population from which the data are drawn, and ‘established correlation’ to refer to claimed genuine correlation that has met the standards required for being considered established. Since establishing is fallible, that a correlation is established does not deductively imply that there is a genuine correlation, though it makes it very likely. Moreover, to establish a correlation it is not necessary that *every* relevant dataset yields an observed correlation, although some observed correlation would typically be required.

RWT says that one ‘normally’ needs to establish both correlation and mechanism. This is because there are certain cases in which causality is not accompanied by a correlation and there are cases in which causality is not accompanied by an underlying mechanism. In cases of overdetermination, where the cause does not raise the probability of the effect because the effect will happen anyway, there is no actual correlation between the cause and the effect. In many such cases, one can expect a *counterfactual* correlation: if things had been different in such a way that the effect would not have happened anyway—e.g., had a second, overdetermining cause been eliminated—then the cause and effect would indeed be correlated. One might think, then, that one ought to be able to establish a counterfactual correlation for any causal claim, if not an actual correlation. However, there are cases in which the cause of interest and a second, overdetermining cause are mutually exclusive, so that it is not possible both to eliminate the second cause and allow the first cause to vary so as to establish a correlation (see [Williamson, 2009, §10](#)). So, even the demand for a counterfactual correlation may be too strong. Let us turn next to causality without mechanisms. Where the cause and / or the effect is an absence, it cannot be connected by an actual mechanism. In many such cases, one can expect a *counterfactual* mechanism. Suppose cause and effect are both absences: e.g., failing to treat causes a lack of a heartbeat. If things had been different in such a way that what was absent in the cause were present (e.g., the treatment is administered), then one would expect a mechanism from this presence to a presence corresponding to the effect (e.g., a heartbeat). One might think, then, that one ought to be able to establish the existence of a counterfactual mechanism for any causal claim, if not an actual mechanism. However, there are cases where one of the cause and

effect is an absence and the other is a presence, and this strategy does not work. For example, suppose that failing to treat causes a blood clot. That the cause is an absence precludes a mechanism here, but the effect being an absence precludes a mechanism in the obverse case, namely, treating causes an absence of a blood clot. Now, establishing causality in these cases is not particularly problematic in practice. However, it is more subtle than simply establishing correlation and establishing mechanism, even where counterfactual correlations or mechanisms are admitted. The question as to how RWT needs to be modified to cover such cases will be not be considered here, because it is not central to the following arguments. The use of ‘normally’ is intended to leave open the possibility that in certain tricky cases one might not need to establish both correlation and mechanism.

RWT requires establishing the *existence* of a correlation and the *existence* of a mechanism, not the extent of the correlation, nor the details of the mechanism. Of course, in some cases establishing the extent of a correlation is a means to establishing its existence, and establishing the details of a mechanism is a means to establishing its existence, but these means are not the only means. We shall return to this point in §2.

RWT is a purely epistemological thesis, concerning the discovery of causal relationships. Russo and Williamson (2007) used the thesis to argue for a particular metaphysical account of causality, the epistemic theory of causality, but RWT itself does not say anything about the nature of causality. The thesis is intended to be both descriptive and normative: i.e., as capturing typical past cases of establishing causality in the biomedical sciences (e.g., Clarke, 2011; Gillies, 2011), as well as characterising the way in which one ought to establish causality.

To see why one ought to establish causality this way, consider that an observed correlation between two variables might be explained in a wide variety of ways, as depicted in Table 2. Some of these explanations provide reason to doubt that there is a genuine correlation in the underlying population. For example, one of the potential confounders might not have been adequately controlled for, or the sample may be rather small. On the other hand, some of these explanations provide reason to doubt that *A* is a cause of *B*, even where there is a genuine correlation between these variables. For example, there might be some variable that could not possibly be considered a potential confounder, given the evidence available, but nevertheless is a confounder, and has not been adequately controlled for. In such a case *A* and *B* can be appropriately correlated yet *A* may not be a cause of *B*—the correlation is attributable to a common cause. Or there may be a genuine correlation that is entirely non causal, explained by a semantic relationship, for instance. Thus there are two forms of error: error when inferring correlation in the data-generating distribution from an observed correlation and error when inferring that *A* is a cause of *B* from an established correlation. Evidence of mechanisms can help to eliminate both forms of error. For instance, it can help to determine the direction of causation, which variables are potential confounders, whether a treatment regime is likely to lead to performance bias, and whether measured variables are likely to exhibit temporal trends.<sup>3</sup>

The existence of the second kind of error—error when inferring that *A* is a cause of *B* from an established correlation—shows that it is not enough to simply establish correlation. If it is genuinely the case that *A* is a cause of *B*, then there is

---

<sup>3</sup>Evidence of mechanisms can help in other respects too. For example, evidence of mechanisms is often essential in order to properly design a clinical study or to interpret its results (Clarke et al., 2014).

Table 2: Possible explanations of an observed correlation between  $A$  and  $B$ .

Causation	$A$ is a cause of $B$ .
Reverse causation	$B$ is a cause of $A$ .
Confounding (selection bias)	There is some confounder $C$ that has not been adequately controlled for by the study.
Performance bias	Those in the $A$ -group are identified and treated differently to those in the $\neg A$ -group.
Detection bias	$B$ is measured differently in the $A$ -group in comparison to the $\neg A$ -group.
Chance	Sheer coincidence, attributable to too small a sample.
Fishing	Measuring so many outcomes that there is likely to be a chance correlation between $A$ and some such $B$ .
Temporal trends	$A$ and $B$ both increase over time for independent reasons. E.g., prevalence of coeliac disease & spread of HIV.
Semantic relationships	Overlapping meaning. E.g., phthisis, consumption, scrofula (all of which are TB).
Constitutive relationships	One variable is a part or component of the other.
Logical relationships	Measurable variables $A$ and $B$ are logically complex and logically overlapping. E.g., $A$ is $C \wedge D$ and $B$ is $D \vee E$ .
Physical laws	E.g., conservation of total energy can induce a correlation between two energy measurements.
Mathematical relationships	E.g., mean and variance variables from the same distribution will often be correlated.

some combination of mechanisms that explains instances of  $B$  by invoking instances of  $A$  and which can account for the correlation. Hence, in order to establish efficacy one needs to establish mechanism as well as correlation.<sup>4</sup>

Let us consider an example. The International Agency for Research on Cancer (IARC) Monographs evaluate the carcinogenicity of various substances and environmental exposures. When evaluating whether mobile phone use is a cause of cancer, IARC found that the largest study (the INTERPHONE study) showed a correlation between the highest levels of call time and certain cancers. This correlation was confirmed by another large study from Sweden. However, evidence of mechanisms was weak and certainly failed to establish the existence of an underlying mechanism. For this reason, chance or bias were considered to be the most likely explanations of the observed correlations, and while causality was not ruled out, neither was it established (IARC, 2013, §§5–6).

★

<sup>4</sup>These assertions hold ‘normally’, i.e., modulo the qualifications about causation between absences discussed above.



Figure 3: SUNY Downstate Medical Center EBM Tutorial (SUNY, 2004).

Further discussion of the descriptive and normative adequacy of RWT can be found in the above references. We will not revisit all these arguments here. Instead, I shall argue here that RWT provides a better account of the epistemology of causality than a rival theory which is motivated by evidence-based medicine (EBM).

Evidence-based medicine is concerned with making the evaluation of evidence explicit:

Evidence based medicine is the conscientious, explicit, and judicious use of current best evidence in making decisions about the care of individual patients. (Sackett et al., 1996)

Of course, this goal is hardly controversial. What characterises present-day EBM is not the goal itself but the means by which it attempts to achieve this goal. EBM employs hierarchies of evidence in order to evaluate evidence and these hierarchies of evidence tend to favour clinical studies and statistical analyses of these studies over other forms of evidence. Clinical studies (CSs) measure the putative cause and effect, together with potential confounders. CSs include controlled experiments such as randomised controlled trials (RCTs) as well as observational studies such as cohort studies, case control studies, case series and collections of case reports. In particular, non-CS evidence of mechanisms, i.e., evidence of mechanisms obtained from means other than the clinical studies that seek to establish a correlation between the putative cause and effect, tends to be either ignored or relegated to the bottom of the hierarchy. For example, Fig. 3 depicts an evidence hierarchy of SUNY (2004), used for EBM training. This places animal research and in vitro research, which in the right circumstances can provide high quality evidence of mechanisms, below ‘opinions’, and well below evidence obtained from clinical studies and statistical analyses of CSs. A similar point can be made about an evidence hierarchy used by the UK National Institute for Health and Care Excellence, depicted in Fig. 4. Fig. 5 depicts the current evidence hierarchy of the Oxford Centre for Evidence-Based Medicine, which places ‘mechanism-based reasoning’ at the lowest level. Other approaches, such as the GRADE system, tend to overlook evidence of mechanisms entirely (Guyatt et al., 2011, Fig. 2).

Levels of evidence	Type of evidence
Ia	Systematic review (with homogeneity) <sup>a</sup> of level-1 studies <sup>b</sup>
Ib	Level-1 studies <sup>b</sup>
II	Level-2 studies <sup>c</sup> Systematic reviews of level-2 studies
III	Level-3 studies <sup>d</sup> Systematic reviews of level-3 studies
IV	Consensus, expert committee reports or opinions and/or clinical experience without explicit critical appraisal; or based on physiology, bench research or 'first principles'

<sup>a</sup> Homogeneity means there are no or minor variations in the directions and degrees of results between individual studies that are included in the systematic review.

<sup>b</sup> Level-1 studies are studies:

- that use a blind comparison of the test with a validated reference standard (gold standard)
- in a sample of patients that reflects the population to whom the test would apply.

<sup>c</sup> Level-2 studies are studies that have only one of the following:

- narrow population (the sample does not reflect the population to whom the test would apply)
- use a poor reference standard (defined as that where the 'test' is included in the 'reference', or where the 'testing' affects the 'reference')
- the comparison between the test and reference standard is not blind
- case-control studies.

<sup>d</sup> Level-3 studies are studies that have at least two or three of the features listed for level-2 studies.

Figure 4: Hierarchy of evidence for diagnostic studies from NICE (2006, p.48).

The main feature of contemporary EBM of relevance to this paper, then, is that it views non-CS evidence of mechanisms as either irrelevant to the process of evidence evaluation or as strictly inferior to evidence obtained from clinical studies and analyses of CSs. In the latter case, opinions differ as to whether or not CSs trump non-CS evidence of mechanisms, i.e., whether or not one should ignore non-CS evidence of mechanisms when clinical studies are available. Either way, however, the sort of studies that provide high quality evidence of correlation are viewed as superior to other investigations which can provide high quality evidence of mechanism.

As a consequence, contemporary EBM stands in conflict with RWT. EBM prioritises evidence of correlation over other evidence of mechanism, whereas RWT treats evidence of mechanism alongside evidence of correlation. An advocate of RWT might interpret the EBM hierarchy of evidence as a way of evaluating evidence of correlation, rather than causation. Under this interpretation, Fig. 6 portrays the epistemological picture motivated by EBM. On the other hand, Fig. 7 represents the approach motivated by RWT (Clarke et al., 2014).

Given this conflict and the fact that EBM is now very well entrenched in medicine, it is no wonder that RWT is controversial. However, we shall see that there are good reasons to prefer the RWT-motivated causal epistemology to the EBM-motivated view. Next, in §2, I shall argue that RWT better explains the role of

Question	Step 1 (Level 1*) Local and current random sample surveys (or consensus)	Step 2 (Level 2*) Systematic review of surveys that allow matching to local circumstances**	Step 3 (Level 3*) Local non-random sample**	Step 4 (Level 4*) Case-series**	Step 5 (Level 5)
<b>How common is the problem?</b>	Local and current random sample surveys (or consensus)	Systematic review of surveys that allow matching to local circumstances**	Local non-random sample**	Case-series**	n/a
<b>Is this diagnostic or monitoring test accurate?</b> (Diagnosis)	Systematic review of cross sectional studies with consistently applied reference standard and blinding	Individual cross sectional studies with consistently applied reference standard and blinding	Non- consecutive studies, or studies without consistently applied reference standards**	Case-control studies, or "poor or non-independent reference standard**	Mechanism-based reasoning
<b>What will happen if we do not add a therapy?</b> (Prognosis)	Systematic review of inception cohort studies	Inception cohort studies	Cohort study or control arm of randomized trial*	Case-series or case-control studies, or poor quality prognostic cohort study**	n/a
<b>Does this intervention help?</b> (Treatment Benefits)	Systematic review of randomized trials or n-of-1 trials	Randomized trial or observational study with dramatic effect	Non-randomized controlled cohort/follow-up study**	Case-series, case-control studies, or historically controlled studies**	Mechanism-based reasoning
<b>What are the COMMON harms?</b> (Treatment Harms)	Systematic review of randomized trials, systematic review or nested case-control studies, n-of-1 trial with the patient you are raising the question about, or observational study with dramatic effect	Individual randomized trial or (exceptionally) observational study with dramatic effect	Non-randomized controlled cohort/follow-up study (post-marketing surveillance) provided there are sufficient numbers to rule out a common harm. (For long-term harms the duration of follow-up must be sufficient.)**	Case-series, case-control, or historically controlled studies**	Mechanism-based reasoning
<b>What are the RARE harms?</b> (Treatment Harms)	Systematic review of randomized trials or n-of-1 trial	Randomized trial or (exceptionally) observational study with dramatic effect			
<b>Is this (early detection) test worthwhile?</b> (Screening)	Systematic review of randomized trials	Randomized trial	Non-randomized controlled cohort/follow-up study**	Case-series, case-control, or historically controlled studies**	Mechanism-based reasoning

Figure 5: Oxford Centre for Evidence-Based Medicine 2011 Levels of Evidence (OCEBM Levels of Evidence Working Group, 2011).

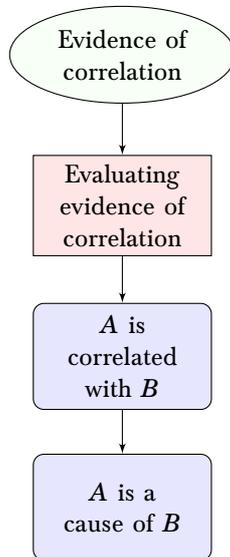


Figure 6: The causal epistemology of contemporary EBM.

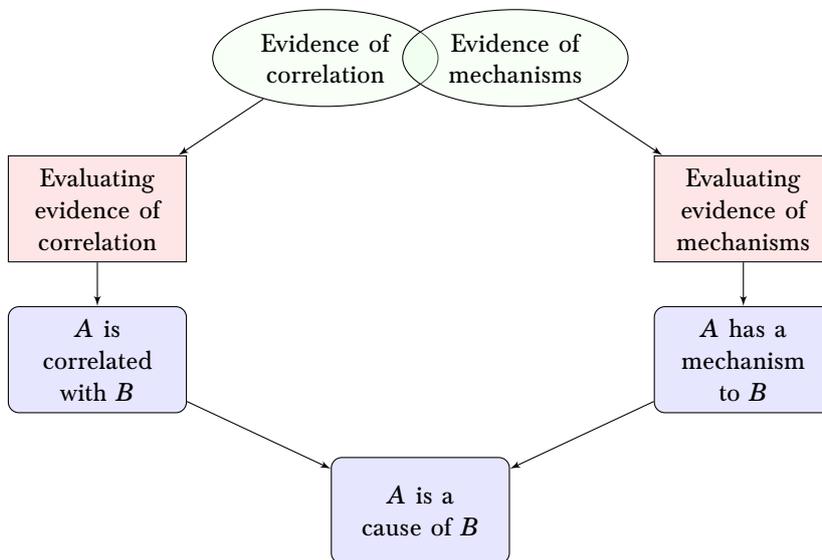


Figure 7: Treating evidence of mechanisms alongside evidence of correlation, as motivated by RWT.

clinical studies in establishing a causal claim. In §3 I shall argue that RWT better explains the process of extrapolating a causal claim from the study population to a target population.

That present-day EBM fails to provide an adequate epistemology of causality does not imply that the whole enterprise of evidence-based medicine is doomed. Current EBM provides a first approximation to the correct epistemology and has led to numerous advances in patient care. The claim made here is that improvements to contemporary EBM can be made, and that the picture of Fig. 7 provides a better approximation. This picture can thus be viewed as a way to develop ‘EBM+’, i.e., as a proposal to advance the methodology of EBM by taking better account of evidence of mechanisms (c.f., [ebmplus.org](http://ebmplus.org)). The main ideas behind EBM+ are that all kinds of evidence of mechanisms need to be explicitly scrutinised and evaluated, not just evidence arising from clinical studies (Table 1), and that this evidence needs to be considered alongside evidence of correlation—rather than as inferior to it—in order to establish effectiveness in medicine, as per Fig. 7. No claim is made that Fig. 7 is the end of the story—further improvements can be made, no doubt.

While present-day EBM advances an essentially monistic account of causal discovery in terms of clinical studies that are evaluated according to how well they establish correlation, the RWT-motivated EBM+ approach is dualistic, treating evidence of mechanisms and evidence of correlation on a par. In this sense, EBM+ has a close affinity to the approach of Austin Bradford Hill, where causal claims are established by means of evidence of mechanisms in addition to evidence of correlation (Bradford Hill, 1965; Russo and Williamson, 2007, §2; Clarke et al., 2014, §2.2). This sort of dualist approach can perhaps be traced back another century to Claude Bernard, who applied it to medicine in general:

Scientific, experimental medicine goes as far as possible in the study of vital phenomena; it cannot limit itself to observing diseases or content itself with expectancy or stop at remedies empirically given, but in addition it must study experimentally the mechanism of diseases and the action of remedies (Bernard, 1865, p. 207).

## §2

### Explananda concerning clinical studies

In this section, I shall argue that RWT can successfully explain three fundamental facts about the role of clinical studies in establishing a causal claim, and that the view motivated by present-day EBM cannot account for all of these facts (although it can account for the first fact). The three facts are these: (i) in some cases, clinical studies suffice to establish a causal claim; (ii) in some cases, randomised studies are not required to establish a causal claim; (iii) in some cases, randomised studies are trumped by other evidence of mechanisms. We shall examine each of these facts in turn.

#### §2.1. In some cases, clinical studies suffice to establish a causal claim

Howick (2011a) suggests that in a number of cases, medical interventions have been accepted on the basis of comparative clinical studies alone. He cites the following cases: the use of aspirin as an analgesic; the use of general anaesthesia; and the use of deep brain stimulation in treating patients with advanced Parkinson’s disease or

Tourette's syndrome. He argues that these cases are a problem for the epistemological thesis RWT, because the mechanisms of action were not—in some cases, still are not—known. Howick points out that these cases are quite compatible with contemporary EBM, which focuses overwhelmingly on clinical studies.

In response to this objection, one might question whether, in these examples, the causal claims really were established on the basis of comparative clinical studies alone. Some examples, such as aspirin and general anaesthesia, pre-date EBM and their effectiveness was arguably established before they were tested in a systematic comparative clinical study. In all cases, background knowledge was important and it is far from obvious that the causal claims were established on the basis of comparative clinical studies *alone*.

However, I do not want to question particular examples here, because I want to accept the general principle that it is possible that clinical studies alone can be used to establish a causal claim in medicine. The point I want to make is that this general principle is quite compatible with RWT.

Consider the RWT-motivated picture of Fig. 7. Some of the total available evidence can be considered to provide evidence of correlation, in the sense that these items of evidence contribute to support or undermine the claim that the putative cause and effect are appropriately correlated. (An item of evidence *contributes* to support a claim if, when taken together with other items, it supports the claim, and the other items do not on their own support the claim to the same degree.) Some of the total available evidence can be considered to provide evidence of mechanisms, in the sense that these items of evidence contribute to support or undermine a claim that there is some mechanism which explains instances of the putative effect in terms of the putative cause and which can account for the extent of the correlation. There is no suggestion that an item of evidence cannot provide both evidence of correlation and evidence of mechanisms.

In particular, clinical studies not only provide evidence of correlation, they can also—in the right circumstances—provide high-quality evidence of mechanisms (Table 1). Suppose:

- There are sufficiently many independent clinical studies,
- They are of sufficient quality (e.g., they are sufficiently large, well-conducted RCTs),
- Sufficiently many studies point in the same direction,
- They observe a large enough correlation ('effect size'),
- It is clear that the variables in question are such that there are no temporal trends and they are not semantically, constitutively, logically, physically or mathematically related,
- There is no other evidence to suggest lack of a suitable mechanism.

Then one can infer both:

- That there is a correlation,
- That there must be some underlying mechanism that explains this correlation.

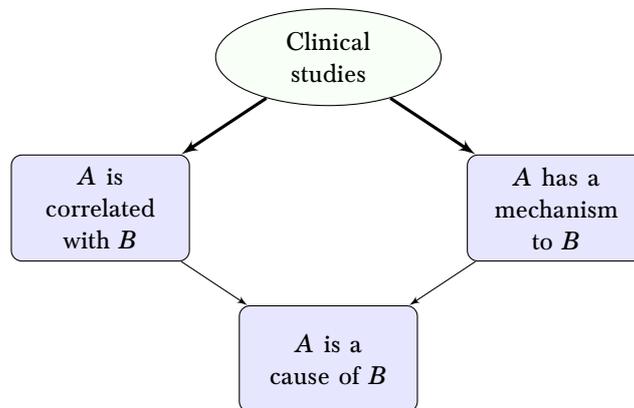


Figure 8: Clinical studies can, in the right circumstances, establish a causal claim.

This is because alternative explanations of the correlation are ruled out (c.f. Table 2). In such cases, clinical studies can provide evidence of the existence of a mechanism even though they may fail to shed light on the details of the mechanism. Note that this inference scheme is non-deductive; there is no suggestion that the premisses guarantee the truth of the conclusions. Fig. 8 depicts this kind of inference, from the perspective of RWT. (In this diagram, an arrow from node  $X$  to node  $Y$  is thick if  $X$  on its own would suffice to establish  $Y$ ; an arrow is thin if  $X$  is insufficient on its own to establish  $Y$ , but nevertheless supports  $Y$ .)

In sum, then, while Howick cites as counterexamples to RWT cases in which clinical studies have sufficed to establish causality, these cases are in fact quite compatible with RWT. There are two separate distinctions at play here. The first is the distinction between evidence of correlation and evidence of mechanisms, which is invoked by RWT. The second is the distinction between clinical studies and non-CS evidence, which is central to present-day EBM. These distinctions do not coincide, and it is only by erroneously conflating the two distinctions that one might think that instances of the above inference scheme refute RWT.

While the above inference scheme is compatible with RWT, it is important to observe that the conditions of the inference are very rarely met in practice. Indeed, instances of this form of inference are very hard to find in IARC evaluations; establishing the carcinogenicity of mists from strong inorganic acids may offer one rare example (IARC, 2012a, pp. 487–495). Thus, although non-CS evidence of mechanisms is not always essential to establishing causality, it is typically an important part of the inferential story.

Confusingly, Howick also cites as evidence against RWT a range of cases in which evidence of mechanisms alone led to erroneous causal inferences; see also Howick (2011b, Chapter 10). These cases clearly confirm—rather than disconfirm—RWT, which says that causal claims cannot be established just by establishing mechanisms, since one needs to establish correlation as well. Moreover, these cases also support EBM+, which holds that all evidence of mechanisms needs to be made explicit and its quality scrutinised, not just evidence arising from clinical studies. This is because in many of these cases the evidence of mechanisms was rather weak.

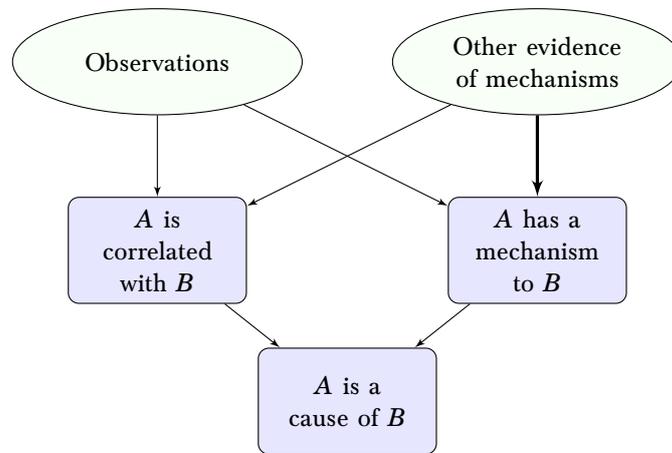


Figure 9: One way to establish a causal claim without RCTs.

**§2.2. In some cases, randomised studies are not required to establish a causal claim**

To support the view that in some cases there is no need for RCTs when establishing causality, consider three examples.

First consider the tongue-in-cheek conclusions of [Smith and Pell \(2003\)](#), who study ‘parachute use to prevent death and major trauma related to gravitational challenge’:

As with many interventions intended to prevent ill health, the effectiveness of parachutes has not been subjected to rigorous evaluation by using randomised controlled trials. Advocates of evidence based medicine have criticised the adoption of interventions evaluated by using only observational data. We think that everyone might benefit if the most radical protagonists of evidence based medicine organised and participated in a double blind, randomised, placebo controlled, crossover trial of the parachute. ([Smith and Pell, 2003](#), p. 1459.)

From the point of view of EBM, the evidence for the effectiveness of parachutes is very weak: no systematic studies, let alone RCTs, and some mechanistic evidence which sits at the bottom of the evidence hierarchy, if it occurs at all. It is hard to see how causality could be established on the basis of this evidence, if EBM is right. From the point of view of EBM+, however, the evidence is strong: excellent evidence of mechanisms, and, although unsystematic, plenty of observational evidence relating to instances where parachutes were and were not used, and a very large observed effect size. From the point of view of EBM+, the evidence of mechanisms on its own suffices to establish the existence of a suitable mechanism, and, when combined with the unsystematic observations, the total evidence suffices to establish correlation too. Hence causality is established. This inference is depicted in Fig. 9.

Having clarified the structure of this inference, let us consider a second example, also considered by [Worrall \(2007\)](#). The question here is how to establish

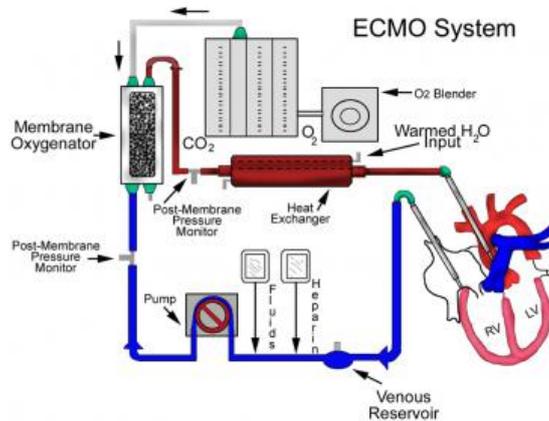


Figure 10: The ECMO mechanism.

the effectiveness of extracorporeal membraneous oxygenation (ECMO) for treating persistent pulmonary hypertension (PPHS). With PPHS, immaturity of the lungs in certain newborn babies leads to poor oxygenation of the blood. ECMO oxygenates the blood outside the body (Fig. 10). Observational studies suggested that ECMO increases survival rate from about 20% to about 80% (Bartlett et al., 1982). However, under standard EBM procedures for evaluating evidence, the available evidence was viewed as insufficient to establish causality, and it was felt necessary to conduct an RCT (Bartlett et al., 1985). At least five subsequent RCTs were carried out, leading to significant loss of life in the control groups.

Conducting RCTs in such a case is considered standard EBM procedure. That non-RCT evidence is viewed as insufficient by EBM was confirmed by a recent Cochrane Review of ECMO, which explicitly disregarded any evidence that did not take the form of an RCT (Mugford et al., 2010).

On the other hand, Worrall (2007) suggests that RCTs were unnecessary in the ECMO case. This conclusion is supported by the RWT-motivated EBM+ approach. This case is analogous to the parachute case: there was strong observational evidence which indicated a large effect size, as well as excellent evidence of mechanisms. Indeed, as in the parachute case, the details of the mechanism of action were very well established. Thus Fig. 9 captures the pre-RCT inference in the ECMO case. There is little doubt that conducting RCTs led to yet greater surety; however, despite being mandated by EBM, RCTs were arguably unnecessary to establish causality.

As a third example, consider the case of establishing the carcinogenicity of aristolochic acid. When IARC originally investigated aristolochic acid in 2002, it found that, while there was observational evidence that Chinese herbs that contain aristolochic acid cause cancer, there was 'limited' evidence in humans concerning the carcinogenicity of aristolochic acid itself as an active ingredient, so carcinogenicity could not be established (IARC, 2002, pp. 69–128). IARC re-examined the question some years later and found that there was little in the way of further observational evidence in humans, so the study evidence involving humans was still 'limited'. However, there was much more evidence of the underlying mechanisms available,

to the extent that the mechanistic evidence could now be described as ‘strong’ and causality could be considered established (IARC, 2012b, pp. 347–361). The key point here is that the change in evidence that warranted establishing causality was a change in mechanistic evidence.

These three cases thus instantiate the following form of inference. Suppose:

- The mechanisms involved are established,
- Observational studies suggest a sufficiently large effect size,
- Sufficiently many studies point in the same direction,
- It is clear that the mechanisms involved can account for the effect size,
- It is clear that the variables in question are such that there are no temporal trends and they are not semantically, constitutively, logically, physically or mathematically related,
- There is no other evidence to suggest that the correlation is best explained in another way,

Then one can infer both:

- That there is a correlation,
- That there must be some underlying mechanism that explains this correlation.

In these cases, evidence of mechanisms obtained by means other than clinical studies provides evidence of correlation. When taken in conjunction with the observational studies, this can be sufficient to establish a correlation (Fig. 9). Note that the observational studies do not need to be very systematic: this is so in the parachute case, it may also be true when establishing some adverse drug reactions (Aronson and Hauben, 2006; Hauben and Aronson, 2007), and it is also true of many interventions that pre-date EBM, such as the use of ileostomy surgery.

While this mode of inference clearly fits the EBM+ approach, which is motivated by RWT, it is harder for contemporary EBM to explain, because, as we saw in the ECMO case, much of the practice of present-day EBM seems to demand randomised studies in order to establish causality. To be sure, some deny that randomised trials are required. For example, Glasziou et al. (2007) argue that in cases where there is a large effect size, RCTs may be unnecessary. However, they struggle to explain from within the EBM paradigm how evidence of mechanisms can be treated on a par with observational studies to help establish causality. Instead they evoke the Bradford Hill indicators of causality, and the Bradford Hill approach is much more in line with RWT and EBM+ than with standard EBM (§1).

### **§2.3. In some cases, randomised studies are trumped by other evidence of mechanisms.**

So far, we have seen that while present-day EBM can account for situations in which RCTs are sufficient to establish causality, it is doubtful whether EBM adequately handles cases in which RCTs are unnecessary. As we shall now see, it is clear that EBM cannot capture cases in which randomised studies are trumped by other evidence of mechanisms. This is because evidence of mechanisms obtained by

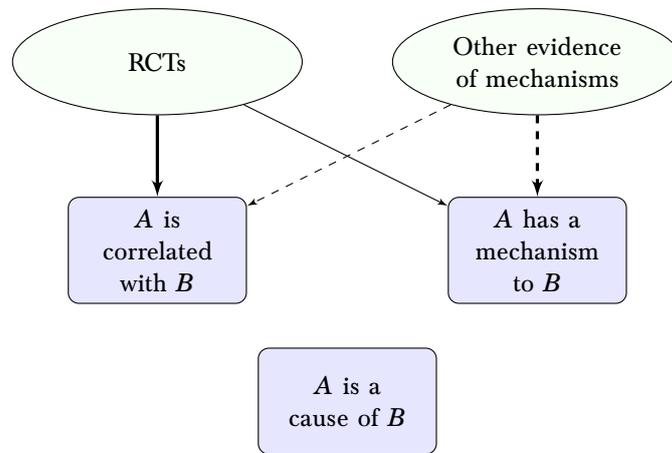


Figure 11: Evidence of a lack of mechanism can trump RCTs.

means other than randomised studies is viewed—when it is considered at all—as strictly inferior to evidence arising from randomised studies (§1).

There are two kinds of example here. One sort of example involves positive evidence of causality from randomised studies; this evidence is trumped by evidence that there is no mechanism by which causality can operate. To start with another tongue-in-cheek example, [Leibovici \(2001\)](#) presented an RCT which observed a correlation between remote, retroactive intercessory prayer and length of stay of patients in hospital. The patients in question had bloodstream infections in Israel during the period 1990–6; the intervention involved saying ‘a short prayer for the well being and full recovery of the group as a whole’ in the year 2000 in the USA, long after recovery or otherwise actually took place. The study also found a correlation between the intervention and duration of fever. The author concludes:

No mechanism known today can account for the effects of remote, retroactive intercessory prayer said for a group of patients with a bloodstream infection. However, the significant results and the flawless design prove that an effect was achieved. ([Leibovici, 2001](#), p. 1451.)

Present-day EBM clearly accords with this inference to an effect, because it views considerations to do with mechanisms as strictly inferior to evidence produced by clinical studies. However, the implicit inference is that this conclusion is ridiculous: no effect should be inferred. This contrary conclusion goes against EBM. It is not possible for present-day EBM to account for the possibility that a large, well-conducted RCT can be trumped by the fact that current science has no place for a mechanism between remote, retroactive intercessory prayer and length of stay of in hospital. On the other hand, this is quite compatible with EBM+. Fig. 11 depicts the inference here, from the perspective of RWT. Undermining evidence is represented by dashed arrows. The thick dashed arrow depicts an inferential connection that is enough on its own to rule out mechanism. As before, the thick solid arrow depicts a connection that would normally be enough on its own to establish the conclusion (correlation): a significant result from a large well-conducted RCT.

However, there is evidence which undermines this conclusion: well-confirmed scientific theory. The presence of this undermining evidence blocks any inference to either correlation or mechanism, and thereby blocks an inference to causation.

Other inferences follow the same pattern. Some comparative studies for precognition have observed a significant correlation (see, e.g., Bem, 2011), as have others in the case of homeopathy (e.g., Chuchrat et al., 2000; Faculty of Homeopathy, 2016). What are the options for resisting an inference to causality in such cases? EBM will point to the fact that the evidence base shows mixed results and is thus inconclusive. However, while this may be so for precognition and homeopathy in general, it is not the case for certain specific interventions which are instances of precognition or homeopathy; as the above references show, there are specific interventions for which only positive studies are available. An alternative way to resist an inference to causality in such cases is to invoke the machinery of Bayesianism: to argue that the prior probability of effectiveness is so low that the posterior probability remains low, despite confirmatory trials. This strategy is open to the charge of subjectivity, or relativity to the way the problem is framed. This charge is clearly correct in the case of a subjective Bayesian analysis, but even under an objective Bayesian reading this charge is on the mark, as an objective Bayesian analysis typically requires high prior probability in deception or experimental error (Jaynes, 2003, §§5.1–2). A third alternative is to apply the RWT-motivated EBM+ approach. The inference in these cases follows the pattern of Fig. 11, and it is clear that causality has not been established, even in specific cases where trials would sufficient in the absence of other evidence to establish correlation. Arguably, then, the RWT-motivated approach is the most promising of these three strategies.

In the kind of example considered above, positive evidence from randomised studies is trumped by evidence of absence of mechanism. But there is another sort of example, in which there is observational evidence, evidence from RCTs and other evidence of mechanisms, and in which the other evidence of mechanisms plays more of a role in establishing causality than do the RCTs. The ECMO case takes this form at the point after the first randomised trial. The first randomised trial provided weak evidence, because after the first baby was randomly assigned to the control arm of the trial and subsequently died, no more individuals were assigned to this arm. Thus the size of the trial was not sufficient to draw any strong conclusions. Arguably, at that point in time the evidence of mechanisms was stronger than that arising from RCTs and it played more of a role in establishing causality. Indeed, if the analysis of §2.2 is correct then the RCT evidence was redundant. The evidence of mechanisms trumps the RCT evidence in such a case.

★

To conclude, the causal epistemology motivated by RWT can validate all three facts about the role of clinical studies in establishing a causal claim. The EBM approach certainly captures the first fact (in some cases, clinical studies suffice to establish a causal claim). However, the practice of EBM goes against the second fact (in some cases, randomised studies are not required to establish a causal claim) and EBM certainly fails to explain the third fact (in some cases, randomised studies are trumped by other evidence of mechanisms).

The proponent of present-day EBM might object that one should not infer a normative thesis about appropriate methodology from a description of actual practice—i.e., from the three facts about the role of clinical studies in actual in-

stances of causal discovery. It is no doubt true that some actual instances of causal discovery were methodologically flawed, and that in some cases researchers thought that they had established a causal claim when in fact they had failed to establish it. Thus one must be cautious when generalising from actual instances to normative claims. However, it is also beyond doubt that—in recent times—medicine has successfully discovered a great number of causal claims. Methods employed in actual medical examples work, by-and-large, and so they tell us something about appropriate methodology. Given this, the three facts do indeed admit a normative interpretation. It is thus incumbent upon the proponent of EBM who denies the normative interpretation of one (or more) of these facts to explain away all the apparent instances of causal discovery which seem to support it. Each fact, under a normative reading, says only that in *some* cases certain methods are appropriate, so in order to deny one of these facts the onus is on the proponent of present-day EBM to show that in *all* cases the corresponding methods are inappropriate.

### §3

#### Extrapolation

We now turn to the question of how a causal claim can be extrapolated from a study population to a target population of interest. This mode of inference is ubiquitous, because the population within which clinical studies establish a correlation (e.g., hospital patients in a particular region who are not too young, not too old, not too ill and not pregnant) rarely coincides with the population within which a treatment is intended to be used. It is also very common—and particularly challenging—to extrapolate causal claims from animals to humans. Any adequate causal epistemology needs to explain how extrapolation is possible and needs to clarify the logic of extrapolation.

Here is a first approximation to the logic of extrapolation:

The causal relationship holds in the study population
The study and target populations are similar in causally relevant respects
The causal relationship holds in the target population

As Steel (2008) points out, this explication faces two immediate problems. The first, which Steel calls the *extrapolator's circle*, is that 'it needs to be explained how we could know that the model and the target are similar in causally relevant respects without already knowing the causal relationship in the target' (p. 78). The worry is that extrapolation seems redundant, since the conclusion of the above rule of inference is apparently needed to establish the second premiss. The second problem, which we shall call the *extrapolator's block*, is that 'any adequate account of extrapolation in heterogeneous populations must explain how extrapolation can be possible even when [causally relevant differences between the model and the target] are present' (pp. 78–9). That is, the study and target population are rarely entirely similar in causally relevant respects—particularly when extrapolating from animals to humans—and it needs to be made clear what sort of differences are permissible in order to prevent the second premiss of the above argument from failing and the inference thereby being blocked.

Note that the study population is chosen for investigation precisely because it is easier to conduct clinical studies on this population than on the target population. Thus the clinical studies that one can perform on the study population tend to

be of a higher standard than those directly obtained on the target population. In the light of this fact, one can sketch an approach to extrapolation motivated by contemporary EBM as follows:

RCTs establish a causal relationship in the study population  
Observational studies in the target population are consistent with this relationship  
The causal relationship holds in the target population

This approach to extrapolation circumvents the aforementioned problems very nicely. There is no extrapolator's circle because one does not need to know that the causal relationship holds in the target population to obtain observational studies in the target population. There is no extrapolator's block because this theory of extrapolation makes extrapolation possible even when there are substantial differences between the study and target populations.

That there may be substantial differences between the study and target populations points to two problems that face the EBM-motivated approach. First we have what we might call the *extrapolator's fallacy*: it needs to be explained how extrapolation is a reliable form of inference, rather than simply fallacious. The worry is that the EBM-motivated account will lead to lots of mistaken conclusions, because observational studies in the target population typically provide weak evidence that the target population is similar to the study population in causally relevant respects. This problem may explain some recent scepticism about extrapolation amongst those interested in medical methodology (see, e.g., Ioannidis, 2012). However, since almost every causal claim of interest has to be extrapolated from some study population, scepticism is hardly a viable option.

The second, related problem is that the *extrapolator's standards are slipping*. In the EBM-motivated approach we have a high standard for internal validity but a low standard for external validity: evidence deemed to be of high quality by EBM (such as that obtained from RCTs) is used to establish causality in the study population, while low quality evidence (such as that obtained from observational studies) is used to establish causality in the target population. In general, an account of extrapolation should not have double standards—the burden of proof for causality should be similar in the study and target populations.

As Steel (2008, Chapter 5) suggests, in order to extrapolate a causal claim from a study population to a target population, one needs evidence that similar mechanisms operate in the two populations. This is particularly important in contexts where mechanisms are likely to differ, such as with extrapolations from animals to humans or interventions involving long causal pathways. It turns out that this feature of extrapolation can be captured by the following RWT-motivated account.

Fig. 12 depicts an account of the logic of extrapolation that is motivated by RWT. In the study population, one can carry out clinical studies that normally cannot be carried out in the target population; these studies are often enough on their own to establish correlation. By also establishing mechanism, one can then establish causality in the study population. Let us turn to the the target population. Clinical studies conducted on the target population, even when augmented with other evidence of the mechanisms of the target population, are insufficient to establish both correlation and mechanism—otherwise there would be no need for extrapolation. Extrapolation is possible when evidence of mechanisms in the target population is strong enough not only to establish the *existence* of a suitable mechanism  $M'$  in the target population, but also to establish that this mechanism

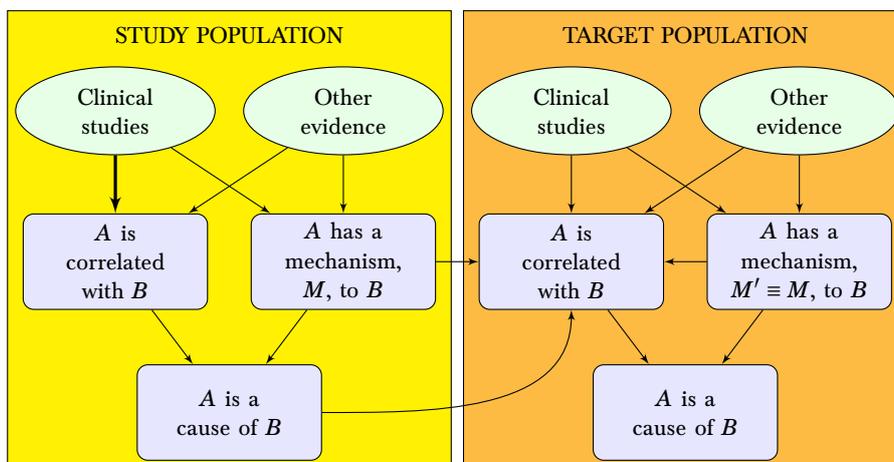


Figure 12: The logic of extrapolation as motivated by RWT.

is *similar* in key respects to the mechanism  $M$  inferred in the study population. The expression  $M' \equiv M$  in Fig. 12 denotes this similarity claim. By means of this similarity of mechanisms, one can use the claim that  $A$  is a cause of  $B$  established in the study population to further support the correlation claim in the target population. In sum, where clinical studies and other mechanistic investigations in the target population are not jointly sufficient to establish correlation in the target, if the corresponding causal claim is established in the study population and it is also established that the mechanisms in the target population are sufficiently similar to those which underpin causation in the study population then this combination of evidence may be enough to establish correlation in the target population. If so, since mechanism in the target is also established, causality can be inferred.

As an extreme case, there may be no clinical studies in the target population; this in itself does not preclude extrapolation under the RWT-motivated account. For example, when IARC evaluated benzo[a]pyrene, they found no human studies measuring exposure to benzo[a]pyrene together with relevant cancer outcomes. However, there were excellent animal studies and enough evidence of mechanisms in animals to establish carcinogenicity in the relevant animal models and to determine the details of the mechanism of action there. Furthermore, there was excellent evidence that the human mechanisms were similar to the mechanisms found in animals. This was considered enough to establish carcinogenicity in humans (IARC, 2012a, pp. 111–144). Note that this example is not validated by the EBM-motivated account of extrapolation provided above, because there were no relevant clinical studies in humans. Thus the example favours the RWT-motivated account of extrapolation.

To take another case where there were no clinical studies in the target population, consider the IARC evaluation of  $d$ -Limonene as a cause of cancer. In this case, there were no studies available in humans. Carcinogenicity of  $d$ -Limonene was established in male rats, so this seemed to be a candidate for extrapolation. However, there were crucial dissimilarities between the mechanism of action in rats and the corresponding human mechanisms: in particular, a protein responsi-

ble for nephrotoxicity in male rats is specific to male rats. Thus no extrapolation was possible and carcinogenicity was not established (IARC, 1999b, pp. 3017–327). This example, which is also in accord with the RWT-motivated account, shows how crucial it is to establish similarity of mechanisms.

Determining similarity of mechanisms can be rather tortuous. With regard to the question of the carcinogenicity of Di(2-ethylhexyl)phthalate (DEHP), causality was established in animals by 1982. In 2000, however, IARC downgraded its carcinogenicity rating in humans—to some controversy (Huff, 2003)—because new evidence suggested that ‘DEHP caused liver tumours in rats and mice by a non-DNA-reactive mechanism involving peroxisome proliferation, which was considered not relevant to humans’ (Grosse et al., 2011, p. 329). In 2011, a third IARC working group had substantially more mechanistic evidence available, and this evidence suggested that there are other pathways in the cancer mechanism, some of which are relevant to humans. This led to the carcinogenicity rating to be upgraded again (Grosse et al., 2011). That the evaluation of carcinogenicity tracks evidence of mechanistic similarity simply cannot be explained by the EBM-motivated account outlined above.

In some cases, new clinical studies in the target population can lead to a re-evaluation of a mechanistic similarity claim. IARC first examined acrylonitrile in 1979 (IARC, 1979, pp. 73–86), and in 1987 decided that carcinogenicity in rats was established and carcinogenicity in humans was likely (IARC, 1987, pp. 79–80). Carcinogenicity was not established in humans because studies in humans provided limited evidence of correlation and other evidence of similarity of mechanisms between rats and humans was also limited. Nevertheless, similarity of mechanisms was credible enough for carcinogenicity in humans to be considered likely. By 1999, later studies in humans had suggested that earlier observed correlations were probably due to confounding by smoking (IARC, 1999a, pp. 43–108). These studies cast doubt both on correlation and on similarity of mechanisms and led to a downgrading of the likelihood of carcinogenicity.

It is important to note that demonstrating mechanistic similarity requires showing that the whole *structure* of relevant mechanisms is sufficiently similar, not just that the mechanism  $M$  by which causality operates in the study population has an analogue in the target population. Thus, one needs to establish that any new counteracting mechanisms in the target population are not so significant that they can cancel out (‘mask’) the action of the analogue of  $M$ . This masking problem was a stumbling block for Anitschkow when he tried to establish that dietary cholesterol causes atherosclerosis by appealing to animal experiments (Anitschkow, 1933). He provided compelling evidence that the causal relationship holds in rabbits and that the mechanism responsible for this relationship also occurs in humans. However, various non-herbivorous animals, including rats, did not exhibit the correlation between dietary cholesterol and atherosclerosis that was found in rabbits. This lack of robustness suggests the presence of a counteracting mechanism in certain non-herbivorous species which masks the action of the positive mechanism of action that was found in rabbits. The presence of such a masking mechanism in humans would count as an important difference between the relevant mechanistic structures in rabbits and humans. Thus, similarity of mechanisms was not established, and causation in humans was rightly not considered established by Anitschkow’s work (see Parkkinen, 2016).

★

We shall now see that this RWT-motivated account of extrapolation survives the four problems of extrapolation identified above.

First let us consider the extrapolator’s circle. That there is no circle should be apparent from the fact that Fig. 12 is acyclic: one does not need to have already established causality in the target in order to meet any of the requirements of establishing causality. Of course, once these requirements are all met, causality in the target is thereby established, but there is no inferential circle here. See Steel (2008, §5.4.2) for further discussion of how mechanism-based approaches can avoid the extrapolator’s circle.

Turning next to the extrapolator’s block, one might worry that we are lacking an account of how extrapolation is possible when mechanisms in the study and target populations are not identical. Similarity of mechanisms is a matter of degree, and the more similar the mechanisms, the more causation in the study population confirms correlation in the target population. Steel (2008, §5.3.2) discusses this question and presents *comparative process tracing* as a method for establishing similarity:

First, learn the mechanism in the model organism, by means of process tracing or other experimental means. For example, a description of a carcinogenic mechanism would indicate such things as the product of the phase I metabolism and the enzymes involved; whether the metabolite is a mutagen, an indication of how it alters DNA; and so on. Second, compare stages of the mechanism in the model organism with that of the target organism in which the two are most likely to differ significantly. For example, one would want to know whether the chemical is metabolized by the same enzymes in the two species, and whether the same metabolite results, and so forth. In general, the greater the similarity of configuration and behavior of entities involved in the mechanism at these key stages, the stronger the basis for the extrapolation. (Steel, 2008, p. 89.)

It is important to note that comparative process tracing is but one of several methods for establishing similarity of mechanisms. One can also establish similarity of mechanisms without determining the details of the mechanisms  $M$  and  $M'$  by employing phylogenetic reasoning, robustness analysis or even enumerative induction (Parkkinen and Williamson, 2017, §4). Thus there is a portfolio of methods for overcoming the extrapolator’s block.

Let us consider the extrapolator’s fallacy next. Unlike the EBM-motivated approach, the RWT-motivated analysis of extrapolation requires evidence that ensures that the study and target populations are similar in causally relevant respects. Mechanistic evidence is playing a key role here, in ensuring that  $M' \equiv M$ . By being more demanding in terms of the evidence required in the target population, extrapolation promises to be more reliable under the RWT account than under the EBM account.

Finally, we can ask whether the extrapolator’s standards are slipping. That this is not so is apparent from Fig. 12: the inferential requirements—establishing correlation and mechanism—are the same in both the study and target populations. If anything, one might one worry that the standards of evidence are higher in the target population than in the study, since Fig. 12 includes the extra requirement of establishing similarity of mechanism there. However, this is just an artefact of the

diagram. Similarity of mechanisms concerns the relation between the study and target populations, not just the target population. Therefore, there is a genuine symmetry between what is required of the study and target populations.

That the RWT account of extrapolation overcomes the latter two problems while the EBM approach does not, speaks in favour of the RWT approach.

★

Having developed the RWT-motivated theory of extrapolation, we shall now consider some criticisms of mechanistic accounts of extrapolation in the light of this theory.

Guala (2010, §6) suggests that there are cases of extrapolation that do not proceed via comparative process tracing. Guala develops an example involving outer continental shelf auctions, which are used to sell oil leases in the Gulf of Mexico, to show that it is not always necessary to determine the details of the relevant mechanisms, as would be required by comparative process tracing. As noted above, the RWT-motivated account sees comparative process tracing as but one of several strategies for establishing similarity of mechanisms, and Guala's case is perfectly in accord with this. What is important to the RWT account is the inferential step  $M' \equiv M$ : strategies for extrapolation seek to demonstrate *similarity of mechanisms*. As Guala notes,

This clearly falls short of a proper articulation of the mechanism ... And yet, it is perfectly adequate for extrapolation purposes. Large parts of the mechanism can be “black boxed” as long as there are good reasons to believe that they are analogously instantiated in the laboratory and target system.’ (Guala, 2010, p. 1080.)

One of the advantages of the RWT-motivated approach, then, is that by situating extrapolation in the inference scheme depicted by Fig. 12 it covers much a broader range of scenarios than comparative process tracing.

Howick et al. (2013a,b) are broadly sceptical of mechanism-based extrapolation. They identify several problems for basing extrapolations on mechanistic evidence. First, our understanding of mechanisms is often incomplete. This is of course true, but insufficient knowledge of the details of  $M$  and  $M'$  for comparative process tracing does not always preclude establishing that  $M' \equiv M$ : one can often employ the other strategies discussed above. Second, knowledge of mechanisms is not always applicable outside the tightly controlled laboratory conditions in which is gained. This is also true, but it is symptomatic of science in general: whatever approach one takes, one must make sure that one's conclusions are robust enough to extend to the application of interest. In particular, an EBM-motivated approach has to ensure that conclusions based on trials with strict exclusion criteria are transportable to the population to be treated. The third problem that they identify is that mechanisms can behave ‘paradoxically’, e.g., a drug can have opposite effects in different contexts. However, it is only by understanding the underlying mechanisms that one can explain these paradoxical effects and improve treatment. On the other hand, clinical studies are crucial for identifying such effects, and one advantage of the RWT-motivated account is that it is not exclusively mechanism-based: it treats evidence of correlation and evidence of mechanisms on a par. The fourth problem that they pick out is the extrapolator's circle. Their worry is that the evidence of the target population required to establish that  $M' \equiv M$  makes the evidence on the

study population redundant. As Fig. 12 makes clear, this need not be the case: one can establish that  $M' \equiv M$  in the absence of evidence from clinical studies in the target population that would be sufficient to establish causality. Howick et al. might respond by noting that under the EBM-motivated account of extrapolation presented above, only weak evidence of the target population is required to establish causality in the target population. However, as discussed above, this is a problem for the EBM-motivated account: it makes extrapolations too easy to be entirely credible—it is subject to the extrapolator’s fallacy. That the RWT-motivated theory of extrapolation is more demanding in terms of the evidence required for extrapolation is an advantage over the EBM-motivated account.

#### §4

#### Conclusion

We have seen that the epistemological thesis RWT motivates a view of medical methodology that stands in conflict with contemporary EBM. Although there is a tension between RWT and EBM, I have argued that RWT can better explain three key features of the use of clinical studies to establish causality, and that it yields a better account of extrapolation. Thus, I conclude that EBM+ is a promising way forward in the controversy as to how best to improve evidence based medicine.

The EBM approach to causal inference has in recent years extended well beyond medicine, to public policy making and various areas of the social sciences, for example. While this paper has focussed on medicine, RWT can be interpreted as having a broader range of application, and similar conclusions to those drawn in this paper may apply beyond medicine. The broader scope of these conclusions is left as a question for further research.

#### Acknowledgements

I am very grateful to the Leverhulme Trust and to the UK Arts and Humanities Research Council for supporting this research, to Julian Reiss, Kurt Straif and Michael Wilde for pointing me to useful IARC examples, and to Brendan Clarke, Donald Gillies, Phyllis Illari, Mike Kelly, Veli-Pekka Parkkinen, Christian Wallmann and Michael Wilde for helpful comments.

---

## Bibliography

---

- Anitschkow, N. (1933). Experimental atherosclerosis in animals. In Cowdry, E., editor, *Arteriosclerosis*. Macmillan, New York.
- Aronson, J. K. and Hauben, M. (2006). Anecdotes that provide definitive evidence. *British Medical Journal*, 333(7581):1267–1269.
- Bartlett, R., Andrews, A., Toomasian, J., Haiduc, N., and Gazzaniga, A. (1982). Extracorporeal membrane oxygenation for newborn respiratory failure: forty-five cases. *Surgery*, 92(2):425–433.
- Bartlett, R. H., Roloff, D. W., Cornell, R. G., Andrews, A. F., Dillon, P. W., and Zwischenberger, J. B. (1985). Extracorporeal circulation in neonatal respiratory failure: A prospective randomized study. *Pediatrics*, 76(4):479–487.
- Bem, D. J. (2011). Feeling the future: Experimental evidence for anomalous retroactive influences on cognition and affect. *Journal of Personality and Social Psychology*, 100:407–425.
- Bernard, C. (1865). *An introduction to the study of experimental medicine*. Henry Schuman, 1949 edition.
- Bradford Hill, A. (1965). The environment and disease: association or causation? *Proceedings of the Royal Society of Medicine*, 58:295–300.
- Campaner, R. (2011). Understanding mechanisms in the health sciences. *Theoretical Medicine and Bioethics*, 32:5–17.
- Campaner, R. and Galavotti, M. C. (2012). Evidence and the assessment of causal relations in the health sciences. *International Studies in the Philosophy of Science*, 26(1):27–45.
- Clarke, B. (2011). *Causality in medicine with particular reference to the viral causation of cancers*. PhD thesis, Department of Science and Technology Studies, University College London, London.
- Clarke, B., Gillies, D., Illari, P., Russo, F., and Williamson, J. (2013). The evidence that evidence-based medicine omits. *Preventative Medicine*, 57(6):745–747.
- Clarke, B., Gillies, D., Illari, P., Russo, F., and Williamson, J. (2014). Mechanisms and the evidence hierarchy. *Topoi*, 33(2):339–360.
- Claveau, F. (2012). The Russo-Williamson theses in the social sciences: Causal inference drawing on two types of evidence. *Studies in History and Philosophy of Biological and Biomedical Sciences*, 43(4):806–813.
- Cucherat, M., Haugh, M. C., Gooch, M., and Boissel, J.-P. (2000). Evidence of clinical efficacy of homeopathy: A meta-analysis of clinical trials. *European Journal of Clinical Pharmacology*, 56:27–33.
- Darby, G. and Williamson, J. (2011). Imaging technology and the philosophy of causality. *Philosophy & Technology*, 24(2):115–136.
- Dragulinescu, S. (2012). On ‘stabilising’ medical mechanisms, truth-makers and epistemic causality: a critique to Williamson and Russo’s approach. *Synthese*, 187(2):785–800.
- Faculty of Homeopathy (2016). The research evidence base for homeopathy. <http://facultyofhomeopathy.org/wp-content/uploads/2016/03/2-page-evidence-summary-for-homeopathy.pdf>.
- Fiorentino, A. R. and Dammann, O. (2015). Evidence, illness, and causation: An epidemiological perspective on the Russo-Williamson thesis. *Studies in History*

- and *Philosophy of Biological and Biomedical Sciences*, 54:1–9.
- Gillies, D. A. (2011). The Russo-Williamson thesis and the question of whether smoking causes heart disease. In Illari, P. M., Russo, F., and Williamson, J., editors, *Causality in the Sciences*, pages 110–125. Oxford University Press, Oxford.
- Glasziou, P., Chalmers, I., Rawlins, M., and McCulloch, P. (2007). When are randomised trials unnecessary? Picking signal from noise. *British Medical Journal*, 334(7589):349–351.
- Grosse, Y., Baan, R., Secretan-Lauby, B., El Ghissassi, F., Bouvard, V., Benbrahim-Tallaa, L., Guha, N., Islami, F., Galichet, L., and Straif, K. (2011). Carcinogenicity of chemicals in industrial and consumer products, food contaminants and flavourings, and water chlorination byproducts. *The Lancet Oncology*, 12(4):328–329.
- Gruta, N. L. L. and Turner, S. J. (2014). T cell mediated immunity to influenza: mechanisms of viral control. *Trends in Immunology*, 35(8):396–402.
- Guala, F. (2010). Extrapolation, analogy, and comparative process tracing. *Philosophy of Science*, 77:1070–1082.
- Guyatt, G., Oxman, A. D., Akl, E. A., Kunz, R., Vist, G., Brozek, J., Norris, S., Falck-Ytter, Y., Glasziou, P., deBeer, H., Jaeschke, R., Rind, D., Meerpohl, J., Dahm, P., and Schünemann, H. J. (2011). {GRADE} guidelines: 1. introduction of grade evidence profiles and summary of findings tables. *Journal of Clinical Epidemiology*, 64(4):383–394.
- Hauben, M. and Aronson, J. K. (2007). Gold standards in pharmacovigilance: The use of definitive anecdotal reports of adverse drug reactions as pure gold and high-grade ore. *Drug Safety*, 30(8):645–655.
- Howick, J. (2011a). Exposing the vanities—and a qualified defence—of mechanistic evidence in clinical decision-making. *Philosophy of Science*, 78(5):926–940. Proceedings of the Biennial PSA 2010.
- Howick, J. (2011b). *The philosophy of evidence-based medicine*. Wiley-Blackwell, Chichester.
- Howick, J., Glasziou, P., and Aronson, J. (2013a). Can understanding mechanisms solve the problem of extrapolating from study to target populations (the problem of ‘external validity’)? *Journal of the Royal Society of Medicine*, 106(3):81–86.
- Howick, J., Glasziou, P., and Aronson, J. K. (2013b). Problems with using mechanisms to solve the problem of extrapolation. *Theoretical Medicine and Bioethics*, 34(4):275–291.
- Huff, J. (2003). IARC and the DEHP quagmire. *International Journal of Occupational and Environmental Health*.
- IARC (1979). *Some monomers, plastics and synthetic elastomers, and acrolein*, volume 19 of *IARC Monographs Series*. International Agency for Research on Cancer, Lyon. <http://monographs.iarc.fr/ENG/Monographs/vol11-42/mono19.pdf>.
- IARC (1987). *Overall evaluation of carcinogenicity: an updating of IARC Monographs volumes 1 to 42*, volume Supplement 7 of *IARC Monographs Series*. International Agency for Research on Cancer, Lyon. <http://monographs.iarc.fr/ENG/Monographs/suppl17/Suppl17.pdf>.
- IARC (1999a). *Re-evaluation of some organic chemicals, hydrazine and hydrogen peroxide*, volume 71 of *IARC Monographs Series*. International Agency for Research on Cancer, Lyon. <http://monographs.iarc.fr/ENG/Monographs/vol71/mono71.pdf>.
- IARC (1999b). *Some chemicals that cause tumours of the kidney or urinary bladder in rodents and some other substances*, volume 73 of *IARC Monographs Series*. Interna-

- tional Agency for Research on Cancer, Lyon. <http://monographs.iarc.fr/ENG/Monographs/vol173/mono73.pdf>.
- IARC (2002). *Some traditional herbal medicines, some mycotoxins, naphthalene and styrene*, volume 82 of *IARC Monographs Series*. International Agency for Research on Cancer, Lyon. <http://monographs.iarc.fr/ENG/Monographs/vol100A/mono100A.pdf>.
- IARC (2012a). *Chemical agents and related occupations*, volume 100F of *IARC Monographs Series*. International Agency for Research on Cancer, Lyon. <http://monographs.iarc.fr/ENG/Monographs/vol100F/mono100F.pdf>.
- IARC (2012b). *Pharmaceuticals*, volume 100A of *IARC Monographs Series*. International Agency for Research on Cancer, Lyon. <http://monographs.iarc.fr/ENG/Monographs/vol100A/mono100A.pdf>.
- IARC (2013). *Non-ionizing radiation, part 2: radiofrequency electromagnetic fields*, volume 102 of *IARC Monographs Series*. International Agency for Research on Cancer, Lyon. <http://monographs.iarc.fr/ENG/Monographs/vol102/mono102.pdf>.
- Illari, P. M. (2011). Disambiguating the Russo-Williamson thesis. *International Studies in the Philosophy of Science*, 25(2):139–157.
- Illari, P. M. and Williamson, J. (2012). What is a mechanism? thinking about mechanisms across the sciences. *European Journal for Philosophy of Science*, 2:119–135.
- Ioannidis, J. P. A. (2012). Extrapolating from animals to humans. *Science Translational Medicine*, 4(151):151ps15.
- Jaynes, E. T. (2003). *Probability theory: the logic of science*. Cambridge University Press, Cambridge.
- Kaiser, M. I. (2016). On the limits of causal modelling: spatially-structurally complex biological phenomena. *Philosophy of Science*, 83:921–933.
- Leibovici, L. (2001). Effects of remote, retroactive intercessory prayer on outcomes in patients with bloodstream infection: randomised controlled trial. *British Medical Journal*, 323:1450–1451.
- Machamer, P., Darden, L., and Craver, C. (2000). Thinking about mechanisms. *Philosophy of Science*, 67:1–25.
- Mugford, M., Elbourne, D., and Field, D. (2010). Cochrane review: Extracorporeal membrane oxygenation for severe respiratory failure in newborn infants. *Evidence-Based Child Health: A Cochrane Review Journal*, 5(1):241–298.
- NICE (2006). *The guidelines manual*. National Institute for Health and Clinical Excellence, London. Available from: [www.nice.org.uk](http://www.nice.org.uk).
- OCEBM Levels of Evidence Working Group (2011). The Oxford 2011 levels of evidence. Oxford Centre for Evidence-Based Medicine, <http://www.cebm.net/index.aspx?o=5653>.
- Parkkinen, V.-P. (2016). Robustness and evidence of mechanisms in early experimental atherosclerosis research. *Studies in History and Philosophy of Science Part C: Studies in History and Philosophy of Biological and Biomedical Sciences*, 60:44–55.
- Parkkinen, V.-P. and Williamson, J. (2017). Extrapolating from model organisms in pharmacology. In Osimani, B., editor, *Uncertainty in pharmacology: epistemology, methods, and decisions*. Springer, Dordrecht.
- Reichenbach, H. (1956). *The direction of time*. University of California Press, Berkeley and Los Angeles, 1971 edition.
- Russo, F. and Williamson, J. (2007). Interpreting causality in the health sciences. *International Studies in the Philosophy of Science*, 21(2):157–170.

- Russo, F. and Williamson, J. (2011a). Epistemic causality and evidence-based medicine. *History and Philosophy of the Life Sciences*, 33(4):563–582.
- Russo, F. and Williamson, J. (2011b). Generic versus single-case causality: the case of autopsy. *European Journal for Philosophy of Science*, 1(1):47–69.
- Sackett, D. L., Rosenberg, W. M. C., Gray, J. A. M., Haynes, R. B., and Richardson, W. S. (1996). Evidence based medicine: what it is and what it isn't. *British Medical Journal*, 312(7023):71–72.
- Salmon, W. C. (1998). *Causality and explanation*. Oxford University Press, Oxford.
- Smith, G. C. S. and Pell, J. P. (2003). Parachute use to prevent death and major trauma related to gravitational challenge: systematic review of randomised controlled trials. *British Medical Journal*, 327:1459–1460.
- Steel, D. (2008). *Across the Boundaries, Extrapolation in Biology and Social Science*. Oxford University Press, Oxford.
- SUNY (2004). SUNY Downstate Medical Center evidence based medicine tutorial. <http://library.downstate.edu/EBM2/contents.htm>. The Medical Research Library of Brooklyn.
- Weber, E. (2007). Social mechanisms, causal inference, and the policy relevance of social science. *Philosophy of the Social Sciences*, 30(3):348–359.
- Weber, E. (2009). How probabilistic causation can account for the use of mechanistic evidence. *International Studies in the Philosophy of Science*, 23(3):277–295.
- Williamson, J. (2009). Probabilistic theories. In Beebe, H., Hitchcock, C., and Menzies, P., editors, *The Oxford Handbook of Causation*, pages 185–212. Oxford University Press, Oxford.
- Williamson, J. (2015). Deliberation, judgement and the nature of evidence. *Economics and Philosophy*, 31(1):27–65.
- Worrall, J. (2007). Why there's no cause to randomize. *British Journal for the Philosophy of Science*, 58:451–488.