Erkenntnis
https://doi.org/10.1007/s10670-022-00637-4

ORIGINAL RESEARCH

A Dilemma for the Russo–Williamson Thesis

Michael Wilde[1]

Received: 12 November 2021 / Accepted: 12 November 2022

Abstract
The Russo–Williamson thesis maintains that establishing a causal claim in medicine normally requires establishing both a correlation and a mechanism. In this paper, I present a dilemma for defenders of this thesis: a strong version of the thesis requires denying a plausible counterexample, but as the thesis is weakened, its defenders must give up their favoured account of the explanatory role of causal claims in medicine. I appeal to some recent work in epistemology on infallibilism to propose a way out of this dilemma, where this way out requires neither denying the plausible counterexample nor giving up the favoured account. I think this shows that even apparently abstract debates in epistemology can provide resources that may help to resolve debates in the philosophy of science and medicine.

1 Introduction

A thesis about the epistemology of causality has been put forward by Russo and Williamson (2007). It has become known as *the Russo–Williamson thesis* (Gillies, 2019). A recent statement of the thesis goes like this: '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' (Williamson, 2019: 33). I will call this *the strong thesis*, or sometimes simply *the thesis*.

The thesis has proved controversial (Broadbent, 2011; Campaner & Galavotti, 2012; Holman, 2019; Howick, 2011a, 2011b, 2019; Illari, 2011). On the one hand, it looks well-motivated, because it is arguably required to account for the limitations of statistical inference, as well as the explanatory role of causal claims in medicine (Clarke et al., 2014; Russo & Williamson, 2007; Williamson, 2019). On the other hand, there seem to be counterexamples to the thesis (Broadbent, 2011; Howick, 2011a, 2011b). In particular, tobacco smoking looks to have been established as

✉ Michael Wilde
m.e.wilde@kent.ac.uk

[1] University of Kent, Canterbury, UK

Published online: 24 November 2022

a cause of lung cancer on the basis of evidence that failed to establish a relevant mechanism (Gillies, 2019).

A critic may respond to this plausible counterexample by simply denying the thesis. *But what options are available to a defender of the thesis?* There are two main options discussed in the literature. The first option is for the defender of the thesis to deny the counterexample (cf. Williamson, 2019). However, I will argue against taking this option, by instead defending the plausibility of the counterexample.

The second option is to weaken the thesis, in light of the plausible counterexample. This option is taken by Donald Gillies (2019). Gillies proposes a *weak thesis*: in order to establish a causal claim in medicine, it is necessary that the existence of an appropriate correlation is established, and that the existence of a mechanism is at least *plausible*, rather than *established*. I will argue that any weakening of the strong thesis will require giving up a *mechanistic* account of the explanatory role of causal claims in medicine. And it is exactly this mechanistic account that is favoured by defenders of the thesis.

It therefore looks like the defender of the thesis is stuck between a rock and a hard place: a strong version of the thesis requires denying a plausible counterexample, but as the thesis is weakened, its defenders must give up their favoured account of the explanatory role of causal claims in medicine. In this paper, I appeal to some recent work on infallibilism in epistemology to propose a way out of this dilemma, where this way out requires neither denying the counterexample nor giving up the favoured account of the explanatory role of causal claims in medicine. I think this shows that even apparently abstract debates in epistemology can provide resources that may help to resolve debates in the philosophy of science and medicine.

2 Arguments in Favour of the Thesis

A distinction is sometimes made between *historical* and *theoretical* arguments for the thesis (Howick, 2011a: 130–136; Russo & Williamson, 2007: 163). A historical argument appeals to a description of some episode of causal discovery, and then claims that the truth of the thesis best explains that episode (cf. Clarke et al., 2014: 344–345). One example is the case of childbed fever, which is a type of infection following childbirth. Williamson maintains that a claim is *established* 'just when standards are met for treating the claim itself as evidence, to be used to help evaluate further claims' (2019: 35). According to one description then, the causal claim that handwashing prevents childbed fever was not established, despite an established correlation, until a relevant mechanism had also been established; only then was it possible to treat this causal claim as evidence for further claims, such as the claim that medical professionals ought to wash their hands before delivering babies. And it is argued that this example is best explained by the truth of the thesis, because the thesis maintains that establishing a causal claim requires both an established correlation and an established mechanism (Russo & Williamson, 2007: 163).

Of course, the difficulty with historical arguments is that descriptions of certain episodes of causal discovery are often contentious. Alex Broadbent says that '[i]t must be remarked that drawing morals from historical episodes is a delicate

business, because it is possible for different commentators to see different lessons in the same episode' (2011: 57). This is not to say that there is not a single correct description of a historical episode that can be used to support or undermine the thesis. Rather, the point is that it is often difficult to determine the correct description. Indeed, conflicting descriptions of the childbed fever case have been proposed (cf. Gillies, 2019: 168–174). For instance, Broadbent himself argues that the discovery of a relevant mechanism came *after* the causal claim about handwashing preventing childbed fever was established (2011: 55–63). In other words, even prior to a relevant mechanism having been established, it was possible to treat this causal claim as evidence for further claims, such as the claim that medical professionals ought to wash their hands before delivering babies.

Instead, I think *theoretical arguments* are the focus for Russo and Williamson. They say '[o]ur point, instead, is a *theoretical* but not an historical one' (2007: 163). A theoretical argument maintains that the truth of the thesis is the best explanation of certain claims about the metaphysics or epistemology of causality (Howick, 2011a: 134–136). And I take it that there were originally two main theoretical arguments put forward in favour of the thesis by Russo and Williamson. I will call the first *the argument from the limitations of statistical inference* (Russo & Williamson, 2007: 162–164). And I will call the second *the argument from the role of causal claims in medicine* (Russo & Williamson, 2007: 158–161). These arguments have been clarified and reiterated in more recent work (Clarke et al., 2014: 343–346; Illari, 2011; Williamson, 2019: 38–39).

Let us look first at *the argument from the limitations of statistical inference*. At least on one interpretation, evidence-based medicine acknowledges that in order to establish a causal claim in medicine, it is necessary to establish an appropriate correlation between the putative cause and the putative effect (cf. Auker-Howlett & Wilde, 2020: 459). The first theoretical argument emphasises the limitations of inferring causation solely on the basis of probabilistic evidence that establishes such a correlation. And its conclusion is 'that probabilistic evidence needs to be accounted for by an underlying mechanism before the causal claim can be established' (Russo & Williamson, 2007: 159).

But the original statement of this argument left a few matters unclear. In particular: *When exactly is a correlation an appropriate correlation? And what exactly is meant here by a mechanism?* In later work, Williamson clarifies that '[h]ere "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' (2019: 36). Williamson also distinguishes the complex-system mechanisms of Machamer et al. (2000), from the mechanistic processes of Salmon (1998). He clarifies that, in the statement of the thesis, '"mechanism" can be understood broadly as referring to a complex-system mechanism, a mechanistic process, or some combination of the two' (2019: 34).

More generally, I think the argument from the limitations of statistical inference has been stated most clearly in the more recent work. Here is Brendan Clarke et al.:

> Evidence of an appropriate sort of correlation between A and B cannot be enough to establish a causal connection between A and B, because correlations

> can arise in a great variety of ways, only one of which is a causal connection between A and B. (Clarke et al., 2014: 343).

In other words, because some correlations have a non-causal explanation, establishing an appropriate correlation between a putative cause and effect is insufficient for establishing that those variables are causally related. In particular, an established correlation between two variables may be explained by chance, bias, or confounding (Williamson, 2019: 38–39). This shows at best that establishing a correlation is insufficient for establishing a causal claim; it provides little reason to think that it is an established *mechanism* that is required on top of an established correlation. But another idea here is that it is evidence of a mechanism that helps to rule out non-causal explanations of an established correlation: 'Evidence of mechanisms can often help us distinguish these sorts of [non-causal] correlations from causal correlations' (Clarke et al., 2014: 344).

Altogether these ideas are taken to clinch the strong thesis: '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' (Williamson, 2019: 33). *Why only normally?* Here again is Williamson: 'This is because there are certain cases in which causality is apparently not accompanied by a correlation and there are also cases in which causality is apparently not accompanied by an underlying mechanism. If this is so, one cannot expect to establish both correlation and mechanism in these cases' (2019: 37). (But I propose to ignore this qualification in the remainder of the paper by restricting attention to so-called *normal* cases throughout.)

Let us now turn to *the argument from the role of causal claims in medicine*. One important point of establishing a causal claim in medicine is *action-oriented*: causal claims are relied upon to predict the effects of medical interventions (Russo & Williamson, 2007: 157–161). For example, one of the points of establishing that tobacco smoking is a cause of lung cancer was to be able to predict the beneficial effects of smoking cessation, and thereby better control the prevalence of lung cancer in the community. Russo and Williamson argue that reliably predicting the effects of an intervention in this manner requires that the relevant causal claim logically implies the existence of an appropriate correlation between the effects and that intervention: 'in the health sciences, causal claims are used for prediction, diagnosis, and intervention; for these modes of inference to be possible, a cause needs to make a difference to its effects, i.e., there needs to be some appropriate probabilistic dependence' (2007: 159). It is supposed to follow from all this that a causal claim is established only if the existence of an appropriate correlation is established.

Another important point of establishing a causal claim in medicine is related to the *cognitive* rather than *action-oriented* goal of medicine: a causal claim may also be relied upon in order to help *explain* some effect by appealing to one of its causes, for example, when a patient's lung cancer may be explained in part by appealing to their tobacco smoking, at least insofar as they were in fact a tobacco smoker (Russo & Williamson, 2007: 157). Russo and Williamson favour a *mechanistic account* of explanation, according to which explanations in medicine require pointing towards the actual existence of a relevant mechanism (cf. Illari & Williamson, 2011). For

example, a patient's lung cancer may be partly explained by appealing to the relevant mechanisms of tobacco smoke carcinogenesis, if these mechanisms were in fact at least partly responsible for this patient's lung cancer. And it is supposed to follow from all this that a causal claim is established only if the existence of a relevant mechanism is also established. They say that 'the mechanistic aspect is required because mechanisms explain the dependencies, and in the health sciences causal relationships are also meant to be explanatory' (2007: 159).

In sum, in order to account for the role of causal claims in medicine for prediction, intervention, and explanation, a causal claim must logically imply the existence of both an appropriate correlation and a relevant mechanism. And this is taken to clinch the thesis that a causal claim is established only if it is also established that there exists both an appropriate correlation and a relevant mechanism (Clarke et al., 2014: 345).

I hope this provides a quick overview of the main theoretical arguments in favour of the thesis. *Are the arguments any good*? One worry is that such theoretical arguments cannot tell us anything about causality in medicine, since they need not be based specifically upon claims about the metaphysics and epistemology of causality *in medicine*. Now, I happen to think that causality in medicine is not so different from causality more generally; insofar as theoretical arguments can tell us something about causality in general, they can tell us something about causality in medicine in particular. Regardless, theoretical arguments can in fact be based specifically upon claims about the metaphysics and epistemology of causality in medicine, where such claims in turn may be supported by appealing to a description of some medical episode of causal discovery. A theoretical argument need not be an ahistorical argument.

A worry may remain that theoretical arguments can mislead by being based upon *inaccurate* claims about the metaphysics and epistemology of causality. Indeed, Bennett Holman (2019) makes the case that theoretical arguments may involve idealizations that lead to ignoring factors that are important in providing an accurate epistemology of medicine, including social factors such as industry influence (cf. Howick, 2019). A good theoretical argument will need to avoid making such idealizations. Indeed, Williamson (2021) has now begun to address the importance of accounting for such social factors when providing theoretical arguments.

Another reason to worry is that there are *two* theoretical arguments; if either argument was conclusive on its own, then presumably we would not need the other argument. A final reason to worry is that the thesis also seems to have a counterexample.

3 A Plausible Counterexample

Jeremy Howick argues that 'there are many … examples where treatments were widely accepted before any semblance of a mechanism was established' (2011a: 131–132). Among other examples, Howick gives the case of deep brain stimulation as an established treatment for Parkinson's disease. But it is not just causal claims about the effectiveness of medical interventions that can arguably be established in the absence of an established mechanism. Howick also thinks that there

are cases where causal claims about the harmful effects of environmental exposures were established before the relevant mechanisms had been established. He gives the example of soot being established as a cause of scrotal cancer before the mechanisms involving benzo[a]pyrene had been established (2011a: 132). But perhaps the most famous example here involves tobacco smoking being established as a cause of lung cancer on the basis of *epidemiological studies* (Broadbent, 2011; Doll & Peto, 1976). Indeed, Gillies provides a detailed case study to argue that 'Doll and Peto were quite justified in claiming in 1976 to have established that smoking causes lung cancer, even though the mechanism linking the two was, at that stage, only plausible rather than confirmed' (2019: 141). In other words, tobacco smoking looks to have been established as a cause of lung cancer on the basis of evidence that failed to establish the existence of a relevant mechanism. And this case has led some to think that the thesis must be mistaken. For instance, Broadbent discusses this case, concluding that '[t]he discovery of a mechanism *can* of course help to confirm a causal hypothesis, but a causal hypothesis can also be solidly confirmed well before the underlying mechanism is known. Therefore discovery of an underlying mechanism is not a necessary condition on warranted causal inference' (2011: 57).

A defender of the thesis will respond by denying the counterexample. They will maintain that tobacco smoking was not established as a cause of lung cancer in the absence of an established mechanism. For instance, contrary to the proposed counterexample, they might maintain that tobacco smoking had in fact not yet been established as a cause of lung cancer, perhaps by appealing to the possibility of a common cause, such as a genetic predisposition to both lung cancer and tobacco smoking (Russo & Williamson, 2007: 162–163). Otherwise, they might maintain that the existence of a mechanism linking lung cancer and tobacco smoking had in fact been established, again contrary to the proposed counterexample. Perhaps the proposed counterexample wrongly assumes that comparative studies cannot by themselves provide evidence of the existence of a mechanism (cf. Williamson, 2019: 43–45). Indeed, Howick seems to make this assumption by interpreting the thesis as saying that 'mechanistic reasoning is required alongside comparative clinical studies' (2011a: 136). In fact, the thesis maintains at most that a causal claim is established only if an appropriate correlation and a relevant mechanism have been established. It makes no claims about which methods might establish the existence of such a correlation or mechanism (Illari, 2011: 141–148).

I am not convinced by this response. Firstly, it seems wrong to maintain that tobacco smoking was not established as a cause of lung cancer until after 1976. A report on tobacco smoking and health was published by the Royal College of Physicians as early as 1962. It considers the common cause explanation of the established correlation between lung cancer and tobacco smoking, but concludes that the explanation that best fits all the evidence is the hypothesis that tobacco smoking is simply a cause of lung cancer: 'We are therefore left with the hypothesis that habitual cigarette smoking over many years is a cause, in the ordinary sense, of lung cancer' (Royal College of Physicians, 1962: 25). Moreover, the report suggests possible interventions by the government to reduce the prevalence of lung cancer, including educational campaigns and restrictions to smoking in public places (1962: 53–55). I think that such interventions would seem out of place unless tobacco smoking was

an *established* cause of lung cancer. Now, it might be objected that such interventions could be justified as good precautionary measures, even if tobacco smoking had not yet been established to be a cause of lung cancer. However, there are costs associated with misplaced interventions, including financial costs to both individuals and governments, and the cost of losing public trust in future government health recommendations. Arguably, these costs outweigh any benefits that might justify the interventions as precautionary measures. Of course, there remained a controversy about the claim that tobacco smoking is a cause of lung cancer, but such a controversy is consistent with the relevant causal claim having been established, especially since such a controversy may simply be manufactured by industry (Proctor, 2012).

Secondly, it also seems wrong to maintain that the existence of a mechanism had in fact been established on the basis of the epidemiological studies, contrary to the proposed counterexample. Although an epidemiological study together with background knowledge may be enough to make *plausible* the existence of a mechanism, it is not clear that it is enough to *establish* the existence of a mechanism, since an epidemiological study necessarily provides very limited evidence of the details of any mechanism. Gillies points out that there may have been *independent* evidence sufficient to establish the existence of a mechanism that might explain the correlation between tobacco smoking and lung cancer (2019: 141). However, his point is that such independent evidence was not necessary to establish the claim that tobacco smoking is a cause of lung cancer; it was enough that the epidemiological studies made plausible the existence of a relevant mechanism.

It might be objected that these remarks do not provide a completely convincing argument in favour of accepting the counterexample. I agree. In effect, the proposed counterexample is a historical argument that remains open to conflicting descriptions: according to one description, tobacco smoking was established as a cause of lung cancer on the basis of evidence that failed to establish the existence of a relevant mechanism; according to another description, that evidence alone was not enough to establish tobacco smoking as a cause of lung cancer. However, I hope to have shown at least that the strong thesis involves denying a *plausible* counterexample. *How else then might a defender of the thesis respond to this plausible counterexample?*

4 The Weak Thesis

An alternative response to this plausible counterexample has been advocated by Gillies (2019). He thinks the example shows that establishing a causal claim in medicine requires only a *plausible mechanism* rather than an *established mechanism.* In effect, this results in *a weakened version of the thesis*: in order to establish a causal claim in medicine, it is necessary that the existence of an appropriate correlation is established, and that the existence of a mechanism is at least *plausible*, rather than *established* (Gillies, 2019: 141).

One immediate worry is that this response sets a low standard for establishing a causal claim. Indeed, it has been argued that it is almost always possible to propose a plausible mechanism, at least in the social sciences (Steel, 2004). However, such

an argument depends upon a certain notion of what constitutes a *plausible mechanism*, a notion according to which a mechanism that is merely supported by a psychologically compelling story counts as a plausible mechanism (cf. Clarke et al., 2014: 350). Gillies is clear that the notion of plausibility at play in the weak thesis is more demanding. He says that '[a] plausible mechanism is one, which is confirmed by our general background knowledge but not necessarily by particular investigations and experiments designed to test it out' (2019: 140). A plausible mechanism still requires a good level of confirmation; it is just that this level of confirmation falls short of *establishing* the mechanism.

Gillies thinks that this *weak thesis* avoids the counterexample involving tobacco smoking and lung cancer (2019: 140–141). I agree. Although the example is arguably a case in which a causal claim was established on the basis of evidence that failed to *establish* the existence of a mechanism, it is not a case in which a causal claim was established on the basis of evidence that failed to make the existence of a mechanism *plausible*. Indeed, even if an epidemiological study cannot establish the existence of a mechanism, it can at least make plausible the existence of a relevant mechanism, by ruling out certain non-causal explanations of an observed correlation, for example, if the effect size is sufficiently large (Howick, 2011a: 56–60; Williamson, 2019: 44). Moreover, in the smoking example there was at least *some* direct evidence of a mechanism:

> Cigarette smoke contains a large number of chemicals of various types and, in the case of those who smoke regularly, some of these are likely to find their way into the lungs. It seems likely that some of these chemicals will damage the tissue of the lungs, and some indeed might be carcinogens which initiate a cancerous tumour. It was part of the background knowledge of 1976 that there existed chemical carcinogens capable of initiating cancerous tumours. Indeed this had been established by experimental evidence. Consequently, it was known in 1976 that there was a plausible mechanism linking smoking to lung cancer. (Gillies, 2019: 140)

Although this evidence falls short of *establishing* the existence of a relevant mechanism, it is enough to make *plausible* the existence of a mechanism. Given this, the weak thesis does indeed accommodate the counterexample involving lung cancer and tobacco smoking.

However, I also think there is a problem with this way of responding to the proposed counterexample. In particular, any weakening of the thesis will involve giving up a mechanistic account of the explanatory role of causal claims in medicine.

Recall one of the two theoretical arguments for the thesis: the argument from the role of causal claims in medicine. One strand of this argument maintains that causal claims in medicine perform an explanatory role. The argument then appeals to a *mechanistic* account of this explanatory role. According to this account, a causal claim can be used for explanation only insofar as it logically implies the existence of a mechanism, since explanations are given by pointing towards existing mechanisms. Now, I take it that this strand of the argument clinches at best the *metaphysical* conclusion that a causal claim logically implies the existence of a mechanism. It does not clinch the needed *epistemological* conclusion that a causal claim is

established only if the existence of a relevant mechanism is *established*. But I take it that *establishing* in the relevant sense is then assumed to be closed under logical consequence, in the sense that if a proposition p logically implies a proposition q, then p is established only if q is established. It is together with this additional *epistemological* assumption that this strand of the argument clinches the conclusion that a causal claim is established only if the existence of a relevant mechanism is also established. However, this conclusion is exactly what any relevantly weakened version of the thesis is denying, because such a version of the thesis will maintain that it is possible to establish a causal claim without establishing the existence of a relevant mechanism. As a result, the defender of such a thesis is required to give up the present mechanistic account of the explanatory role of causal claims in medicine. In particular, the weak thesis maintains that a causal claim is established as long as the existence of that mechanism is at least plausible rather than established. So a defender of this weak thesis will be forced to give up the present mechanistic account of the explanatory role of causal claims in medicine. And it is precisely this account that is favoured by defenders of the thesis.

A defender of the weak thesis may try to avoid this problem by denying the assumption that establishing is closed under logical consequence. However, I take this assumption as basically stipulating the sense of *establishing* under discussion. Instead, the defender will have to deny the metaphysical assumption that a causal claim logically implies the existence of a mechanism. But the problem with this response is that the metaphysical assumption is more than just an assumption, precisely because it is motivated by their favoured mechanistic account of the explanatory role of causal claims in medicine: explanations in medicine are given by pointing to the existence of a mechanism. And giving up this metaphysical assumption is tantamount to giving up the present mechanistic account of explanation.

A defender of the weak thesis may instead want to give up the present mechanistic account. However, there is a reason why this account should be favoured even by defenders of the weak thesis. In particular, if a causal claim can be explanatory without logically implying the existence of a mechanism, then there seems little reason to think that even a *plausible* mechanism is required in order to establish a causal claim. In that case, this response scuppers the motivation for even the weak version of the thesis. Now, it might be argued that a plausible mechanism is still required in order to overcome the limitations of statistical inference. In other words, the weak thesis may still be motivated by the other theoretical argument. But if it is possible for a causal claim to be explanatory even though it does not logically imply the existence of a mechanism, then there seems little reason to think that it is evidence specifically of a *mechanism* that is required to rule out non-causal explanations of an established correlation.

In sum, there is a dilemma: the strong thesis involves denying a plausible counterexample, but as the thesis is weakened, defenders of such a thesis are required to give up their favoured account of the explanatory role of causal claims in medicine. One proposal for resolving this dilemma is to allow that the strong thesis holds only in some contexts, and that the weak thesis holds in the remaining contexts. However, this proposal still involves giving up the mechanistic account of the explanatory role of causal claims in medicine, at least in certain contexts. It therefore really looks like

the defender of the thesis is stuck between a rock and a hard place. I will now appeal to some recent work on infallibilism in epistemology to propose a way out of this dilemma.

5 Infallibilism About Knowledge

I take infallibilism to be a thesis that relates the notion of *knowledge* to the notion of a *body of evidence*. *What exactly is a body of evidence?* It is simply one way of cashing out the idea that for a proposition to be evidence for a hypothesis for some subject, that subject needs to hold that proposition in a good enough epistemic standing (Williamson, 2000: 186–187). To see this, note that even if a proposition raises the probability of a hypothesis for some subject, that proposition may still not count as evidence for the hypothesis for that subject, since they may not hold that proposition in good enough epistemic standing, for instance, the subject may know the proposition to be false (Williamson, 2000: 187). A subject's body of evidence is simply the set of propositions that the subject does hold in good enough epistemic standing. It is the set of propositions that the subject can properly use as evidence for or against a hypothesis.

Infallibilism is the thesis that a subject knows a proposition p only if their body of evidence entails p. *Why would anyone advocate this sort of infallibilism?* Here is David Lewis: 'To speak of fallible knowledge, of knowledge despite uneliminated possibilities of error, just *sounds* contradictory' (1996: 549). In particular, utterances of the following form are infelicitous: "A subject S knows that p but it is epistemically possible for S that not-p." These utterances have become known as *concessive knowledge attributions* (Dodd, 2010; Dougherty & Rysiew, 2011; Littlejohn, 2011). One so-called *simple* account of the infelicity of concessive knowledge attributions is that they are false, because a subject's knowing that p is simply inconsistent with the epistemic possibility of not-p for that subject (cf. Dodd, 2011: 672). This is the account favoured by the defender of infallibilism. It follows from this account that a subject knows p only if not-p is epistemically impossible for that subject. But this consequence does not obviously lead to infallibilism, because it might be that not-p is epistemically impossible for a subject without the subject's body of evidence entailing p (cf. Stanley, 2005). However, a defender of infallibilism might then assume a link between evidence and epistemic impossibility: not-p is epistemically impossible for a subject if and only if p is entailed by that subject's body of evidence (see, for example, Williamson, 2000: 186–228). It then follows that a subject knows p only if p is entailed by that subject's body of evidence.

A defender of infallibilism will therefore think it is well-motivated, since it follows from their favoured account of the infelicity of concessive knowledge attributions, at least given the assumed link between evidence and epistemic impossibility. However, there is also something clearly problematic about infallibilism: it seems to be susceptible to counterexamples, namely, cases in which a subject looks to know p even though their body of evidence fails to entail p (Brown, 2013; Dodd, 2007). Indeed, any case of knowledge by non-deductive inference would show that it is possible to know p on the basis of evidence that fails to entail p; for example, it looks

possible to know that all emeralds are green on the basis of evidence of only a much more limited sample of emeralds (Brown, 2013: 627).

One response to such a counterexample is to provide a weaker alternative to infallibilism: for a subject to know p, the subject's evidence need not entail p, but it must at least make p *plausible*. Essentially, this is just a version of *fallibilism*, because it allows for the possibility of a subject knowing p with a body of evidence that fails to entail p (Cohen, 1988; Reed, 2002, 2012). This fallibilism can therefore accommodate the counterexamples involving knowledge by non-deductive inference. However, a defender of this fallibilism cannot give the simple account of the infelicity of concessive knowledge attributions. This is because, according to fallibilism, the epistemic possibility of not-p for a subject is not inconsistent with the subject's knowing that p. The point has been made by Dylan Dodd: '[T]he simplest explanation of the infelicity of [concessive knowledge attributions] … is that they're false. They strike us as infelicitous because we grasp that they're obviously false. This simple explanation can be given only by the infallibilist' (2011: 672). So the fallibilist must give up this simple account. But the problem is that this is the favoured account, since it is the simplest way to explain the infelicity of concessive knowledge attributions.

This should seem a familiar dilemma. Earlier, I argued that defenders of the strong thesis were motivated by their favoured mechanistic account of the explanatory role of causal claims in medicine. However, the strong thesis cannot accommodate the plausible counterexample involving lung cancer and tobacco smoking. On the other hand, I argued that the weak thesis can accommodate this plausible counterexample. But a defender of this weak thesis must give up their favoured account of the explanatory role of causal claims in medicine. Here there is a similar situation. A defender of infallibilism is motivated by their favoured account of the infelicity of concessive knowledge attributions, but infallibilism cannot accommodate the counterexamples involving knowledge by non-deductive inference. A version of fallibilism can accommodate these counterexamples, but only by giving up the favoured simple account of the infelicity of concessive knowledge attributions. It is a similar dilemma.

How should we resolve the dilemma between fallibilism and infallibilism? One approach is to defend infallibilism by denying the counterexamples involving knowledge by non-deductive inference. Another approach is to defend fallibilism by denying the simple account of the infelicity of concessive knowledge attributions. However, Clayton Littlejohn (2008) has explained that an alternative approach has opened up. Importantly, this alternative involves neither denying the counterexamples nor giving up the simple account of the infelicity of concessive knowledge attributions. It requires first disambiguating *standard* and *non-standard* versions of both fallibilism and infallibilism by appealing to some work by Stewart Cohen (1988). I will now present the details of this disambiguation.

Infallibilism puts forward a necessary condition on knowing a proposition p: a subject knows p only if their body of evidence entails p. But there are different ways of thinking about such a necessary condition. According to *the standard interpretation*, infallibilism is a thesis about what it takes *to come to know* p. In other words, infallibilism says that a subject can come to know p only by

believing p on the basis of a body of evidence e that entails p. On this standard interpretation, infallibilism is of course inconsistent with the counterexamples involving knowledge by non-deductive inference, because those are examples where it is possible to come to know p by believing p on the basis of a body of evidence e that fails to entail p. But there is another way of thinking about infallibilism. According to *the non-standard interpretation*, infallibilism is not a thesis about what it takes to come to know p. Instead, it is simply a thesis about the *consequences* of knowing p. On this non-standard interpretation, infallibilism is the thesis that once a subject knows p, this results in the subject having a body of evidence e' that now entails p.

Importantly, the standard interpretation of infallibilism is not required in order to give the simple account of the infelicity of concessive knowledge attributions. It is enough to have the non-standard interpretation, at least given the assumption linking evidence and epistemic impossibility. In particular, if one of the consequences of a subject coming to know p is that the subject then gets a body of evidence e' that entails p, then it remains inconsistent to say "S knows that p but it is epistemically possible for S that not-p." Moreover, the non-standard interpretation avoids the counterexamples involving knowledge by non-deductive inference. On the non-standard interpretation, infallibilism says *nothing* about what it takes *to come to know* p, so in particular it does not rule out the possibility of coming to know p on the basis of a body of evidence e that fails to entail p (Littlejohn, 2008: 682–684).

The upshot is that it is possible for the infallibilist to avoid the counterexamples, while also accommodating the simple account of the infelicity of concessive knowledge attributions, namely, by advocating the non-standard rather than the standard interpretation of infallibilism. The main worry about the non-standard interpretation is precisely that it is *non-standard*. Infallibilism now only tells us about the *consequences* of knowing p. It no longer says anything about what it takes *to come to know* p. However, some help in this respect is provided by giving a similar disambiguation of fallibilism.

The present version of fallibilism maintains that a subject knows p only if the subject's body of evidence makes p at least plausible. The problem was that this fallibilism looks to involve giving up the simple account of the infelicity of concessive knowledge attributions. In fact, it is not clear that this version of fallibilism has any such problem; again, it depends on the way it is interpreted. According to *the non-standard interpretation*, this fallibilism is not a thesis about what it takes to come to know p. Instead, it is simply a thesis about the consequences of knowing p. In other words, it says that once a subject comes to know p, the subject then gets a body of evidence e' that makes p at least plausible. On this non-standard interpretation, this fallibilism does indeed involve giving up the simple account of the infelicity of concessive knowledge attributions, because the epistemic possibility of not-p for a subject is now no longer taken to be inconsistent with the subject's knowing p. However, there is an alternative way of thinking about this fallibilism (cf. Cohen, 1988; Littlejohn, 2008: 682–683). And this alternative interpretation does say something about what it takes to come to know p. In particular, according to *the standard interpretation*, this fallibilism says that a subject can *come to know* p on the basis of a body of evidence e only if e makes p at least plausible.

Importantly, the standard interpretation of fallibilism is all that is required to accommodate the counterexamples involving knowledge by non-deductive inference, because by itself it allows for the possibility of coming to know p on the basis of a body of evidence e that fails to entail p. Moreover, this fallibilism no longer gives up the simple account of the infelicity of concessive knowledge attributions. It now says *nothing* about the consequences of a subject knowing p. In particular, it may still be the case that a subject's coming to know p thereby results in the subject getting a body of evidence e' that entails p. The upshot is that it is possible to avoid ruling out the simple account of the infelicity of concessive knowledge attributions, while accommodating the counterexamples to infallibilism, namely, by advocating only the standard interpretation of this version of fallibilism. But now the worry is that this interpretation comes at the cost of it saying nothing about the *consequences* of knowing p.

Where does this leave us? A non-standard interpretation of infallibilism avoids the counterexamples, while also accommodating the simple account of the infelicity of concessive knowledge attributions. But this interpretation comes at the cost of saying nothing about what it takes to come to know p. A standard interpretation of the fallibilist alternative does say something about what it takes to come to know p. It also avoids giving up the simple account of the infelicity of concessive knowledge attributions, while accommodating the counterexamples to infallibilism. But this interpretation now comes at the cost of saying nothing about the consequences of knowing p.

Littlejohn (2008) notes that these two interpretations are entirely compatible, because one talks solely about what it takes to come to know p, whereas the other talks solely about the consequences of knowing p. And the combination results in a *conciliatory approach* that: (1) says something about the consequences of knowing p; (2) says something about what it takes to come to know p; (3) accommodates the counterexamples of knowledge by non-deductive inference; (4) gives the simple account of the infelicity of concessive knowledge attributions. One way out of the dilemma concerning fallibilism and infallibilism thus involves being a fallibilist about what it takes to come to know p, while at the same time being an infallibilist about the consequences of knowing p. In other words, one maintains that a subject can come to know p on the basis of a body of evidence e that fails to entail p, but as soon as they come to know p, this results in a body of evidence e' that now does entail p (Littlejohn, 2008: 682–683).

It might be objected that this conciliatory approach is in fact incoherent (cf. Littlejohn, 2008: 683). The approach implies that a subject can get evidence e' that entails p simply by believing p on the basis of evidence e that itself fails to entail p. And it looks hard to explain where this additional evidence has come from (Dodd, 2007: 642). However, there are some possible explanations available. Littlejohn mentions the explanation that appeals to *the knowledge theory of evidence* advocated by Timothy Williamson (2000). According to this theory, p is entailed by a subject's body of evidence if and only if the subject knows p (cf. Williamson, 2000: 184–208). Given the counterexample, it is possible to come to know p by non-deductive inference, that is, by believing the p on the basis of evidence e that fails to entail p. In that case, some instances of believing p on the basis of non-entailing

evidence e will also be instances of coming to know p. A proponent of the knowledge theory of evidence can therefore say that it is this additional knowledge that provides the explanation of the additional evidence: the subject now has more evidence because they now have more knowledge (Littlejohn, 2008: 683–684). It has not been shown that the conciliatory approach is incoherent.

It might still be objected that a conclusive argument has not yet been provided in favour of the conciliatory approach. Firstly, no argument has been provided in favour of the knowledge theory of evidence. Secondly, the alternative non-conciliatory approaches to resolving the dilemma have not been ruled out: a standard infallibilist may deny the counterexamples; a standard fallibilist may deny the simple account of the infelicity of concessive knowledge attributions. However, Littlejohn's aim here is not to provide a conclusive argument for the conciliatory approach. He is in effect merely showing that there is an alternative way to resolve this dilemma, where this alternative way involves neither denying the putative counterexamples nor giving up the simple account of the infelicity of concessive knowledge attributions.

6 A Way Out of the Dilemma

What has all this got to do with the Russo–Williamson thesis? I think a similar conciliatory approach provides a way out of the dilemma concerning the weak and strong versions of the thesis.

Infallibilism puts forward a necessary condition on a particular epistemic standing with respect to some claim p: a subject knows p only if their body of evidence entails p. As we have seen, there are two different interpretations of infallibilism, because there are two different ways of thinking about such a necessary condition. It might be a condition on what it takes *to come to occupy* that epistemic standing, or it could be a condition merely detailing the *consequences* of that epistemic standing. In this respect, infallibilism is similar to the strong thesis, because the strong thesis also puts forward a necessary condition on a particular epistemic standing with respect to some claim: a causal claim in medicine is established only if the existence of both an appropriate correlation and a relevant mechanism have been established. And so there are similar standard and non-standard ways of thinking about this necessary condition. It is therefore possible to disambiguate the strong thesis in the same way it was possible to disambiguate infallibilism.

According to *the standard interpretation*, the strong thesis says something about what it takes *to come to establish* a causal claim: in order to come to establish a causal claim in medicine, one needs to believe that causal claim on the basis of evidence that establishes both an appropriate correlation and a relevant mechanism. The problem with the standard interpretation is that it makes the strong thesis inconsistent with the plausible counterexample involving tobacco smoking and lung cancer, because this is plausibly a case where a causal claim came to be established on the basis of evidence that made the existence of a mechanism at most plausible rather than established (Gillies, 2019: 140–141).

On the other hand, the non-standard interpretation is not inconsistent with the counterexample. According to *the non-standard interpretation*, the strong thesis

says nothing about what it takes *to come to establish* a causal claim in medicine. It talks only about the *consequences* of establishing a causal claim: if a causal claim is established, then this results in both an established correlation and an established mechanism. Note that the non-standard interpretation is enough to provide the mechanistic account of the explanatory role of causal claims in medicine, because it does not deny the metaphysical assumption that a causal claim logically implies the existence of a relevant mechanism. I take it that something like this non-standard interpretation of the strong thesis was first described by Broadbent (2011). He grants the metaphysical assumption that a causal claim implies the existence of a relevant mechanism, but does not think that this clinches the epistemological conclusion that a causal claim can come to be established only on the basis of an established mechanism: 'Even if the mechanistic *interpretation* is plausible, and general causal claims are to be interpreted as existence claims about underlying mechanisms, it does not follow that a general causal claim is only warranted when the underlying mechanism is identified' (2011: 57).

Now, the problem is that the non-standard interpretation of the strong thesis comes at the cost of saying very little about what it takes *to come to establish* a causal claim in medicine (cf. Broadbent, 2011: 66). One might therefore object that this interpretation wrongly trivializes the thesis. Arguably, the thesis is most naturally interpreted as providing practical guidance about how exactly to go about coming to establish causal claims in medicine. Indeed, it was proposed as an alternative to a crude version of evidence-based medicine, which attempts to come to establish causal claims on the basis of established correlations alone (Russo & Williamson, 2011). Moreover, the thesis has also motivated a set of action-guiding principles and procedures for coming to establish causal claims in practice (Parkkinen et al., 2018). But the non-standard interpretation seems to rid the thesis of these implications for practice, because the thesis then talks only about the consequences of establishing a causal claim in medicine, rather than what it takes to come to establish such a causal claim. However, some help in this respect is provided by disambiguating also the *weak thesis*.

According to a non-standard interpretation, the weak thesis concerns the *consequences* of establishing a causal claim in medicine: if a causal claim is established in medicine, then this results in a body of evidence that makes the existence of a relevant mechanism at least plausible. The problem with this interpretation is that it allows for the possibility of establishing a causal claim without establishing the existence of a relevant mechanism. Given that establishing is closed under logical consequence, it therefore involves giving up the mechanistic account of the explanatory role of causal claims in medicine, because this interpretation no longer maintains the metaphysical assumption that a causal claim logically implies the existence of a relevant mechanism.

On the other hand, the standard interpretation of the weak thesis does not involve giving up this mechanistic account. According to the standard interpretation, the weak thesis says nothing about the *consequences* of establishing a causal claim in medicine. It talks only about what it takes *to come to establish* a causal claim: in order to come to establish a causal claim in medicine, it is necessary to believe that claim on the basis of a body of evidence that makes the existence of a mechanism

at least plausible, rather than established. In particular, it may still be the case that establishing a causal claim thereby results in an established mechanism. But note that the standard interpretation of the weak thesis also accommodates the plausible counterexample involving lung cancer and tobacco smoking, because it allows for the possibility of coming to establish a causal claim on the basis of a body of evidence that has not established the existence of a mechanism.

Unfortunately, the standard interpretation of the weak thesis comes at the cost of the thesis now saying nothing about the *consequences* of establishing a causal claim in medicine. And we are left in a familiar position. A non-standard interpretation of the strong thesis avoids the counterexample, while also accommodating the mechanistic account of the explanatory role of causal claims in medicine. But this interpretation comes at the cost of saying nothing about what it takes *to come to establish* a causal claim in medicine. A standard interpretation of the weak thesis does say something about what it takes to come to establish a causal claim. Moreover, it avoids giving up the mechanistic account of the explanatory role of causal claims, while still accommodating the counterexample. But this interpretation now comes at the cost of saying nothing about the *consequences* of establishing a causal claim in medicine.

Again, it can be noted that it is entirely consistent to maintain both the standard interpretation of the weak thesis and the non-standard interpretation of the strong thesis. This is because the former talks solely about what it takes to come to establish a causal claim, whereas the latter talks solely about the consequences of establishing a causal claim. And the combination results in a conciliatory approach that: (1) says something about the consequences of establishing a causal claim; (2) provides practical guidance on what it takes to come to establish a causal claim in medicine; (3) accommodates the counterexample; (4) provides the favoured mechanistic account of the explanatory role of causal claims in medicine. Given this, one way out of the dilemma is to advocate the weak thesis concerning what it takes *to come to establish* a causal claim in medicine, while at the same time advocating the strong thesis concerning the *consequences* of establishing a causal claim in medicine. In other words, the key is to maintain that a causal claim can come to be established on the basis of a body of evidence that makes the existence of a mechanism plausible rather than established, but as soon as the causal claim is thus established, this thereby establishes the existence of the relevant mechanism.

Now, it might be similarly objected that this conciliatory approach is in fact incoherent, or at least in need of further explanation. It implies the following possibility: that a body of evidence that fails to establish the existence of a mechanism can serve as a basis for providing a body of evidence that does establish the existence of that mechanism. And it looks hard to explain where exactly this additional evidence of the existence of a mechanism has come from. However, philosophers of science have long recognized such a possibility; an established scientific theory often goes beyond the evidence upon which it was established. And once again there are possible explanations available. One proposal is to adopt a corresponding *knowledge theory of establishing*: a claim p is established if and only if p follows from what is known (cf. Williamson, 2000: 228). According to this proposal, evidence that fails to establish the existence of a mechanism can still serve as a basis for providing

further evidence that does establish the existence that mechanism. This is because evidence that fails to establish the existence of a mechanism can serve as a basis for coming to *know* the relevant causal claim, and it follows from this causal knowledge that there exists a relevant mechanism, at least given the metaphysical assumption that a causal claim logically implies the existence of the relevant mechanism.

It might still be objected that a conclusive argument has not yet been provided in favour of the present conciliatory approach. Firstly, no argument has been provided in favour of the knowledge theory of establishing. Secondly, the alternative non-conciliatory approaches to resolving the dilemma have not been ruled out: a defender of the standard version of the strong thesis may deny the counterexample; a defender of the standard version of the weak thesis may deny the mechanistic account of the explanatory role of causal claims in medicine.

However, the knowledge theory of establishing is not crucial to the conciliatory approach, since such an approach is by itself consistent with any theory of establishing that allows for the possibility of gaining evidence e' that entails p by believing p on the basis of a body of evidence e that does not entail p. I have appealed to the knowledge theory of establishing here for dialectical purposes, since the present conciliatory approach was proposed by way of an analogy with a corresponding conciliatory approach in the debate concerning infallibilism about *knowledge*. Regardless, I am not here attempting a conclusive argument in favour of the conciliatory approach. I am merely setting up the dilemma, and then proposing one way of potentially resolving the dilemma. I agree that the best way out of the dilemma has not yet been established. It may be that the benefits of simply advocating the strong thesis ultimately outweigh the costs of denying the plausible counterexample; it may be that the benefits of simply advocating the weak thesis ultimately outweigh the costs of denying the mechanistic account of the explanatory role of causal claims in medicine. However, a comparative cost–benefit analysis will be conclusive only if all the relevant options on the table. And I hope to have shown that there is a novel alternative option: the conciliatory approach provides an alternative way out of the present dilemma, where this way out involves neither denying the plausible counterexample nor giving up the favoured mechanistic account of the explanatory role of causal claims in medicine. This is all in line with the method of inference to the best explanation: firstly, all the relevant options need to be shortlisted; secondly, these options can then be ranked in terms of their explanatory power; thirdly, the highest ranked option is inferred (Lipton, 2004).

Nevertheless, I do think that the knowledge theory of establishing is a natural proposal, especially given the knowledge theory of evidence, since the proposal then has the plausible consequence that a proposition p is established if and only if p follows from the evidence. And arguments in favour of the knowledge theory of evidence have been provided elsewhere by Alexander Bird (2004, 2018, 2022) and Timothy Williamson (2000). Moreover, Julian Reiss (2015: 344) maintains that a good theory of evidence: (1) provides an account of evidential *support*; (2) provides a distinct account of evidential *warrant* that allows for warrant to come in degrees; (3) applies even in non-idealized circumstances of medical science where certain contextual factors are important, such as 'background knowledge about how the world works, the nature and purpose of the inquiry, and certain normative

commitments' (Reiss, 2015: 349). It seems to me that the knowledge theory of evidence meets these desiderata. According to this theory, a proposition *supports* a hypothesis if that proposition raises the probability of the hypothesis, and a hypothesis is instead *warranted* to the extent that it is probable conditional upon knowledge (Williamson, 2000: 184–237). Now, it might be objected that this results in a non-contextual theory of evidence that therefore cannot apply in the non-idealized circumstances of medical science. However, the knowledge theory of evidence in fact allows that contextual factors may affect the possibility of coming to know a proposition, in which case evidence is appropriately context-sensitive (Williamson, 2000: 188). And it also maintains that a proposition can support a hypothesis in some contexts but not others, depending upon the background information (Williamson, 2000: 186–187).

In addition, the knowledge theory of establishing would seem to apply in the case of tobacco smoking and lung cancer. In the case of lung cancer and tobacco smoking, the initial evidence was enough to establish a correlation between tobacco smoking and lung cancer, but it was not enough to establish the existence of a mechanism to explain the extent of this correlation. And yet arguably the example shows that this body of evidence still managed to serve as a basis for establishing the claim that tobacco smoking is a cause of lung cancer. According to the present proposal, this is because it was possible to *come to know* that tobacco smoking is a cause of lung cancer on the basis of that limited body of evidence, perhaps in part thanks to the relevant contextual factors. But this is not to say that the causal claim was established even though the existence of a relevant mechanism linking lung cancer and tobacco smoking had not been established, since on the present proposal *knowing* a causal claim thereby establishes the existence of a relevant mechanism.

More generally, I also think that the conciliatory approach is a plausible option here. Its benefits are not just that it provides a potential way out of a dilemma in the philosophy of medicine; it manages to provide this way out while making good sense of the arguments on both sides of this dilemma.

Firstly, the conciliatory approach makes sense of the counterexample. Arguably, this counterexample shows that it is possible to come to establish a causal claim on the basis of a body of evidence that makes the existence of a relevant mechanism merely plausible. In other words, the counterexample concerns what it takes to come to establish a causal claim, rather than the consequences of establishing a causal claim. It therefore arguably shows that *the standard interpretation* of the strong thesis is incorrect. But it is a mistake to think that the counterexample also shows that *the non-standard interpretation* of the strong thesis is incorrect. In fact, this non-standard interpretation is needed to provide the favoured mechanistic account of the explanatory role of causal claims in medicine.

Secondly, there is also a sense in which *the argument from the role of causal claims in medicine* is a good argument. It is a good argument insofar as it is an argument concerning the consequences of establishing a causal claim, rather than what it takes to come to establish a causal claim. In effect, the argument says that in order to account for the explanatory and predictive uses of causal claims in medicine, an established causal claim must logically imply an established correlation and an established mechanism, at least in normal cases. As a result, the argument clinches

the non-standard interpretation of the strong thesis. But it is a mistake to think that this argument shows also that *the standard interpretation* of the strong thesis is correct. Arguably, it cannot show this thanks to the present counterexample.

What about the argument from the limitations of statistical inference? I noted above that it was a little odd that there were two theoretical arguments in favour of the thesis: the argument from the limitations of statistical inference, and the argument from the role of causal claims in medicine. If either argument was conclusive on its own, then presumably we would not need the other argument. However, the conciliatory approach makes good sense of there being these two arguments. One argument concerns the *consequences* of establishing a causal claim in medicine. And the other argument concerns the conditions on what it takes *to come to establish* a causal claim. As we have seen, the argument from the role of causal claims in medicine is an argument concerning the consequences of establishing a causal claim. So perhaps the argument from the limitations of statistical inference concerns what it takes to come to establish a causal claim in medicine. And this seems exactly correct. This argument says that in order to come to establish a causal claim, it is not enough to believe that claim solely on the basis of evidence that establishes only the appropriate correlation. Instead, some evidence of the existence of a mechanism is also required in order to help rule out alternative non-causal explanations of the established correlation. That is, the existence of a mechanism must be at least plausible. According to the conciliatory approach, then, there is also a sense in which the argument from the limitations of statistical inference is a good argument. It shows that the standard interpretation of the weak thesis is correct. But it is a mistake to think that this shows that the non-standard interpretation of the weak thesis is correct. Arguably, it cannot show this because the argument from the role of causal claims in medicine shows that a non-standard interpretation of the weak thesis cannot be correct.

In sum, I have argued that the present conciliatory approach provides an alternative way to resolve the dilemma between the weak and strong thesis. Importantly, this way out of the dilemma requires neither denying the plausible counterexample nor giving up the mechanistic account of the explanatory role of causal claims in medicine. It thus presents a third way for the Russo–Williamson thesis. A next step would be to better determine the relative merits of the different ways of resolving the dilemma concerning the weak and strong thesis: denying the counterexample; denying the mechanistic account of explanation; or instead taking the present conciliatory approach. However, I hope to have shown that the conciliatory approach is at least a *prima facie* plausible way out of the dilemma. In particular, it seems to make good sense of the arguments on both sides of the dilemma. Moreover, this conciliatory approach was arrived at by considering a debate about infallibilism in epistemology. I think this shows that philosophers of science and medicine can benefit by paying closer to attention even to apparently abstract debates in epistemology. And some of this work in epistemology has real potential to steer debates in the philosophy of science and medicine in new directions. In particular, one version of so-called *knowledge-first epistemology* maintains that it is impossible to provide a non-circular explanation of knowledge in terms of necessary and sufficient conditions (Williamson, 2000). Given the knowledge theory of establishing, it would then be

pointless for philosophers of science and medicine to carry on attempting to provide a non-circular explanation for how causal and other claims come to be established in science and medicine (cf. Wilde, 2021).

Acknowledgements I am very grateful for feedback from Daniel Auker-Howlett, Raffaella Campaner, Lorenzo Casini, Iain Chalmers, Brendan Clarke, Virginia Ghiara, Donald Gillies, Bennett Holman, Jeremy Howick, Phyllis Illari, Elselijn Kingma, Adam La Caze, Veli-Pekka Parkkinen, Federica Russo, Christian Wallmann, and Jon Williamson.

References

Auker-Howlett, D., & Wilde, M. (2020). Reinforced reasoning in medicine. *Journal of Evaluation in Clinical Practice, 26*(2), 458–464.
Bird, A. (2004). Is evidence non-inferential? *The Philosophical Quarterly, 54*(215), 252–265.
Bird, A. (2018). Evidence and inference. *Philosophy and Phenomenological Research, 96*(2), 299–317.
Bird, A. (2022). *Knowing science*. Oxford University Press.
Broadbent, A. (2011). Inferring causation in epidemiology: Mechanisms, black boxes, and contrasts. In P. McKay Illari, F. Russo, & J. Williamson (Eds.), *Causality in the sciences* (pp. 45–69). Oxford University Press.
Brown, J. (2013). Infallibilism, evidence, and pragmatics. *Analysis, 73*(4), 626–635.
Campaner, R., & 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., Gillies, D., Illari, P., Russo, F., & Williamson, J. (2014). Mechanisms and the evidence hierarchy. *Topoi, 33*, 339–360.
Cohen, S. (1988). How to be a fallibilist. *Philosophical Perspectives, 2*, 91–123.
Dodd, D. (2007). Why Williamson should be a sceptic. *The Philosophical Quarterly, 57*(229), 635–649.
Dodd, D. (2010). Confusion about concessive knowledge attributions. *Synthese, 172*, 381–396.
Dodd, D. (2011). Against fallibilism. *Australasian Journal of Philosophy, 89*(4), 665–685.
Doll, R., & Peto, R. (1976). Mortality in relation to smoking: 20 years' observations on male British doctors. *British Medical Journal, 2*, 1525–1536.
Dougherty, T., & Rysiew, P. (2011). Clarity about concessive knowledge attributions: Reply to Dodd. *Synthese, 181*, 395–403.
Gillies, D. (2019). *Causality, probability, and medicine*. Routledge.
Holman, B. (2019). Philosophers on drugs. *Synthese, 196*, 4363–4390.
Howick, J. (2011a). *The philosophy of evidence-based medicine*. BMJ Books.
Howick, J. (2011b). Exposing the vanities—And a qualified defense-of mechanistic reasoning in health care decision making. *Philosophy of Science, 78*, 926–940.
Howick, J. (2019). Exploring the asymmetrical relationship between the power of finance bias and evidence. *Perspectives in Biology and Medicine, 62*(1), 159–188.
Illari, P. (2011). Mechanistic evidence: Disambiguating the Russo–Williamson thesis. *International Studies in the Philosophy of Science, 25*, 139–157.
Illari, P., & Williamson, J. (2011). Mechanisms are real and local. In P. McKay Illari, F. Russo, & J. Williamson (Eds.), *Causality in the sciences* (pp. 818–844). Oxford University Press.
Lewis, D. (1996). Elusive knowledge. *Australasian Journal of Philosophy, 74*(4), 549–567.
Lipton, P. (2004). *Inference to the best explanation* (2nd ed.). Routledge.

Littlejohn, C. (2008). From E = K to scepticism? *The Philosophical Quarterly, 58*(233), 679–684.
Littlejohn, C. (2011). Concessive knowledge attributions and fallibilism. *Philosophy and Phenomenological Research, 83*(3), 603–619.
Machamer, P., Darden, L., & Craver, C. (2000). Thinking about mechanisms. *Philosophy of Science, 67*(1), 1–25.
Parkkinen, V.-P., Wallmann, C., Wilde, M., Clarke, B., Illari, P., Kelly, M., Norrell, C., Russo, F., Shaw, B., & Williamson, J. (2018). *Evaluating evidence of mechanisms in medicine*. Springer.
Proctor, R. N. (2012). The history of the discovery of the cigarette-lung cancer link: Evidentiary traditions, corporate denial, global toll. *Tobacco Control, 21*, 87–91.
Reed, B. (2002). How to think about fallibilism. *Philosophical Studies, 107*, 143–157.
Reed, B. (2012). Fallibilism. *Philosophical Compass, 7*(9), 585–596.
Reiss, J. (2015). A pragmatist theory of evidence. *Philosophy of Science, 82*(3), 341–362.
Royal College of Physicians. (1962). *Smoking and health*. Pitman Medical Publishing.
Russo, F., & Williamson, J. (2007). Interpreting causality in the health sciences. *International Studies in the Philosophy of Science, 21*(2), 157–170.
Russo, F., & Williamson, J. (2011). Epistemic causality and evidence-based medicine. *History and Philosophy of the Life Sciences, 33*, 563–582.
Salmon, W. (1998). *Causality and explanation*. Oxford University Press.
Stanley, J. (2005). Fallibilism and concessive knowledge attributions. *Analysis, 65*(2), 126–131.
Steel, D. (2004). Social mechanisms and causal inference. *Philosophy of the Social Sciences, 34*(1), 55–78.
Wilde, M. (2021). Mechanistic reasoning and the problem of masking. *Synthese, 199*, 6103–6118.
Williamson, J. (2019). Establishing causal claims in medicine. *International Studies in the Philosophy of Science, 32*(1), 33–61.
Williamson, J. (2021). The feasibility and malleability of EBM+. *Theoria: an International Journal for Theory, History and Foundations of Science, 36*(2), 191–209.
Williamson, T. (2000). *Knowledge and it limits*. Oxford University Press.