Robert Northcott. *Science for a Fragile World*. Oxford University Press 2025. 224 pp. $100.00 USD (Hardcover 9780192849083).

Robert Northcott, at the beginning of his highly readable *Science for a Fragile World*, asks us to imagine a world "in which laws and causal relations are fragile, winking in and out like bubbles in a boiling soup" (1) or like the mud bubbles rising and popping on the cover image. Northcott has two aims in this philosophy of science book, written mainly for an academic audience. First, to defend the claim that many of the relations studied by and of interest to the sciences are indeed fragile. And second, to suggest that when relations are fragile, scientific investigations should be carried out in the manner of a 'Case-Worker' rather than a 'Stability-Theorist', methodological approaches the book sets out to define. Despite being relatively short, the book makes a convincing case for these two claims. It also opens avenues for further research, as well as inviting scientists from a range of disciplines and "science funders, planners, and educators" (15) to reflect on, and potentially change, some of their practices.

Structurally, the book is divided into four parts. The first four chapters introduce the main ideas and lay the theoretical groundwork. The next three chapters situate fragility in the landscape of recent work in philosophy of science. The following three chapters discuss fragility in the context of specific sciences and/or methodologies: economics, epidemiology, and big data approaches. The conclusion relates the foregoing discussion to debates about expertise. Although it is helpful to think of the book as having theoretical and applied parts, even the more theoretical chapters are richly illustrated with examples from a wide range of disciplines, including history, sociology and ecology. In this review I will focus on the theoretical chapters and mention only some of the examples.

What is fragility (chapter 2)? It is "a property of relations generally" (17), but those most frequently discussed in the book are causal and explanatory relations. Except for relations described by universal laws, "[a]ll relations hold in some circumstances but not in others"—it is only sometimes true that "a rise in national prosperity causes an incumbent to win an election, an open auction increases the price, a target being a small island causes a species invasion to succeed, and striking a match causes a flame" (17). The difference between these examples is the degree to which the causal relation holds *predictably*. It is easy to predict that, if oxygen is present, striking a match will cause a flame, whereas it is harder to predict whether and under what circumstances a rise in national prosperity will cause an incumbent to be re-elected. Whereas predictability comes in degrees, Northcott defines fragility as a binary property because "methodologically, there is a dichotomy between two strategies, roughly, between that of the theorist and that of the case worker" (19). Putting these two axes of variation—predictability and methodological strategy—together, we get the definition of fragility: "*A relation is **fragile** if and only if, in the salient range of circumstances, it does not hold predictably enough to favor the Stability-Theorist rather than the Case-Worker strategy*" (20). As Northcott points out, this definition can make fragility seem redundant, because what really matters is whether a relation is predictable enough to favour one research methodology over the other. However, fragility acts as a useful marker for the

appropriateness of the Case-Worker approach.

The difference between Stability-Theorist and Case-Worker is best illustrated by one of Northcott's examples. He describes two approaches to explaining why it was that during World War I spontaneous truces arose, and broke down, along the Western front (chapter 3). The Stability-Theorist uses a wide-scope repeated Prisoner's Dilemma model to explain these truces: it was in soldiers' self-interest not to shoot at the enemy (cooperation) rather than break these truces (defection), but also to retaliate when the enemy shot first. Northcott suggests that such a decision-theoretic account does not adequately explain the truces because it does not have empirical warrant. For example, the Prisoner's Dilemma model cannot explain why truces only arose at some times and places, why truces were sustained when they were, and why they broke down when they did. Instead, detailed historical analysis of "regimental archives, letters sent by soldiers, and … interviews of veterans" (38), revealed that truces tended to arise between non-elite rather than elite units, thus explaining why truces were not uniformly distributed. Sometimes "individual hotheads" (54) disrupted the truces, but "local truce norms … acquired a moral tinge" (36), making truces robust in a way even a repeated Prisoner's Dilemma model does not capture. The Case-Worker approach is appropriate here because the causal relations between various psychological and sociological mechanisms and the truces hold unpredictably, i.e. are fragile.

On Northcott's view, wide-scope theories and models play a toolbox role for the Case-Worker (chapter 4). The application of a theory/model has to have warrant for the particular explanation at hand. For example, although there may be wide-scope psychological theories about how "individual hotheads" are disruptive, this theory has to be applied locally to ascertain whether, when and how hothead behaviour led to the breakdown of truces. The Case-Worker approach also has implications for theory/model development: detailed, local investigation can suggest new theory, refine existing theory, and discover "diagnostic indicators" (68), i.e. markers that aid investigators in predicting whether a theory/model will apply in new cases. Incidentally, by drawing on multiple case studies from lots of different sciences for his own philosophical analysis, Northcott himself exemplifies the Case-Worker approach—a satisfying instance of form following function.

Having explained the core divide, it is worth further clarifying the relation between predictability, the dichotomous methodological choice (Stability-Theorist vs. Case-Worker), and fragility. Predictability not only comes in degrees, but also has "two faces" (23): objective and subjective. The holding (or not) of some relations is easier to predict than that of others in an objective—agent-independent—manner. For example, linear systems are objectively easier to predict than nonlinear dynamical systems. But predictability also has an agent-relative side: how predictable even a linear system is will depend on the agent's knowledge of it. Fragility, being defined in terms of predictability, also has these two faces. This suggests that fragility has both metaphysical and epistemic dimensions. Since, however, the key decision that scientists face is the methodological one between Stability-Theorist and Case-Worker, the salient aspect of fragility seems to boil down to its epistemic "face". Northcott sometimes equivocates on this point. In other words, when a relation holds unpredictably—whether the reason for this unpredictability be a

feature of the world, the agent's ignorance, or both—Case-Worker is appropriate. Upon suitable investigation it may turn out that the fragile relation is a stable relation in the world; we just didn't know enough about it. For example, the discovery that certain chemotherapy combinations work better than using the drug cisplatin on its own was unpredictable in the 1960s (chapter 5). But once it became clear that each drug in the combination has a complementary mechanism of action this provided the basis for stable prediction and further development of new drugs. On the other hand, insights gained about the causes of political violence in Germany and Italy in the 1960s and 1970s from a detailed case study could not be extrapolated to other contexts, such as the US civil rights movement (chapter 4). This extrapolation failed because, objectively, the "causal relations that govern political violence hold unpredictably" (75).

One overarching argument of the book is that scientists ought not assume as often as they do that the relations they are studying are stable, i.e. that they license the Stability-Theorist approach. When scientists make this assumption, but their wide-scope theories turn out not to be empirically accurate and/or do not make accurate predictions, the temptation is to tweak those theories and try again. The case studies in the book demonstrate that this "one more heave" (6) approach is rarely successful when relations are (objectively) fragile. In other words, although Northcott doesn't frame it in these terms, he is advocating for greater epistemic humility among scientists. This has implications for research resource allocation. Regarding research on the World War I truces, "[i]f our currency is explanations of real-world phenomena, resources directed at the Prisoner's Dilemma are not well spent. They are better directed to the history department" (44).

In sum, Northcott succeeds at showing that many relations worthy of scientific inquiry are fragile, and that this fragility warrants a Case-Worker approach to theory application as well as development. Along the way he touches on lots of topics, such as ontology and scientific modelling, that philosophers of science will want to know more about. As a first foray into systematising the concept of fragility, there are necessarily gaps to be filled in, but this raises interesting new questions. For example, how exactly do case studies help furnish the theory toolbox and help discover markers for future investigations? Does the fragility approach presuppose that the relata—explanandum and explanans—themselves are well-defined or stable entities? What happens when the phenomena to be explained are not well-defined? Does the Case-Worker approach apply to philosophy? Does philosophy also make progress by having more tools in the theoretical toolbox?

Victoria Min-Yi Wang, University of Edinburgh