Note: This is a chapter from my 2025 dissertation "Epistemic Issues in Evidence Amalgamation," available at https://d-scholarship.pitt.edu/concern/etds/c024df94-a176-45e1-b34a-70f3a53d26ca

Explaining Multimodal Evidence

Michael W. Begun

1 Introduction

During the early stages of the Covid-19 pandemic, a critical question for understanding and controlling the disease was whether it spreads through aerosols. Epidemiologists have traditionally distinguished between three modes of transmission for respiratory viruses: contact, droplet, and aerosol transmission. Whether a virus transmits through aerosols has serious implications for limiting its spread. Facing a novel threat, health agencies needed to offer guidance to the public and policymakers based on a messy and incomplete body of evidence. In its first scientific brief on modes of Covid-19 transmission from March 2020, the WHO argued that Covid-19 transmits predominantly through contact and droplets (WHO 2020a). The brief held that aerosol transmission was likely limited to specific medical settings that employ aerosol-generating procedures. Consequently, the WHO generally recommended only droplet and contact precautions for people caring for Covid-19 patients. Noting that news outlets had raised the possibility of aerosol transmission based on preliminary studies, the WHO argued that the existing evidence that seemed to support aerosol transmission was weak and did not reflect real-world conditions (WHO 2020a). In the United States, the Centers for Disease Control and Prevention (CDC) expressed similar doubts about aerosol transmission (Jimenez et al. 2022).

With the benefit of hindsight, it is easy to criticize the early assessments of the WHO and CDC. The evidence supporting aerosol transmission is now overwhelming and there is a near-universal consensus among epidemiologists that aerosols are a major transmission mechanism. However, we ask scientists and public health agencies to make their best assessments given the evidence available, and in 2020 the evidence on aerosol transmission was not nearly so clear. The WHO and CDC had a particularly difficult problem. They needed to assess a methodologically diverse body of evidence in which studies sometimes appeared to support contradictory conclusions. To figure out whether Covid-19 transmits through aerosols, there is no single study design or type of experiment that can provide decisive evidence. Rather, understanding Covid-19 transmission requires integrating studies on the biology of SARS-CoV-2 infection, the physics of aerosol dispersion, viral load and shedding among infected individuals, real-world transmission events, and more. At best, each type of study can only supply a piece of a larger puzzle.

This sort of challenge is more common than it might appear. Scientists and policymakers often confront complex bodies of evidence when attempting to evaluate hypotheses and make informed decisions. Evidential complexity can come in many forms. Studies on the same topic may rely on different background assumptions, operationalize phenomena in different ways, and reach contradictory conclusions. While we may desire the evidence to be straightforward, this rarely occurs in many scientific fields and policy domains. A particularly challenging situation arises when evidence is produced using multiple methods or techniques. Studies on the same topic may employ experimental designs that differ in any number of ways. They may seek different kinds of data, rely on distinct background theories, and intervene in different physical systems. Following recent discussion in philosophy of science (Stegenga 2009; Hey 2015), I will use the term "multimodal evidence" to refer to this general idea. In Appendix A, I provide a detailed account of multimodal evidence that goes beyond the existing characterizations found in philosophy of science.

The question of how to properly amalgamate multimodal evidence is far from straightforward. A daunting variety of evidence amalgamation approaches exist, and well-established approaches have important limitations. Some techniques such as meta-analysis can only be applied to studies that report quantitative outcomes and use similar experimental designs, leaving out other relevant sources of evidence. Robustness analysis, an approach celebrated by philosophers of science, can overcome the narrow evidential scope of meta-analysis, yet fails to provide much insight when facing discordant bodies of evidence. Other plausible approaches similarly carry serious limitations. Considerations like these have led some philosophers of science to conclude that multimodal evidence cannot be amalgamated in any systematic way (Stegenga 2009).

In this chapter, I elucidate and defend a framework for amalgamating multimodal evidence. My account draws from scientific cases and from philosophy of science to put forward a unified framework for amalgamating multimodal evidence. I argue for an explanatory-eliminative approach for amalgamating multimodal evidence. According to this framework, researchers should systematically develop explanatory accounts of the evidence corpus and then attempt to rule out each account by showing the implausibility of one or more of its explanatory links. Among other things, this approach contrasts with robustness analysis and methodological triangulation, as well as certain Bayesian amalgamation methods.[1] A key feature of this approach is that it is able to incorporate highly indirect evidence as well as details about evidential context. To develop my account of the explanatory-eliminative approach, I examine two successful scientific cases of multimodal evidence amalgamation: how epidemiologists evaluated aerosol transmission of Covid-19, and how climate scientists applied a novel approach to estimate Earth's climate sensitivity.

This chapter is structured as follows. I start by examining well-known evidence amalgamation methods and explain their major weaknesses for dealing with multimodal evidence. In Section 3, I examine how scientists have successfully amalgamated evidence about Covid-

1. Some Bayesian amalgamation strategies compete with the explanatory-eliminative approach. However, Bayesian techniques can be used within the approach to attach probabilities to explanatory accounts, as I highlight later in the chapter. Bayesianism itself does not fully specify how to amalgamate complex sets of evidence in practice.

19 transmission routes and equilibrium climate sensitivity, highlighting common features in the two cases. Building off my discussion of these two cases, in Section 4 I outline a framework for amalgamating multimodal evidence that overcomes the weaknesses of other approaches.

2 The Evidence Amalgamation Challenge

Before examining several prominent evidence amalgamation techniques, I highlight two aspects of the evidence amalgamation challenge that have not received much attention in philosophical discussions. The first is that sources of evidence often bear highly indirectly on a hypothesis of interest. By indirectness, I have in mind the following idea: an empirical result may not be *about* the hypothesis in any typical sense, yet can be related to it through a chain of reasoning that appeals to auxiliary claims and possibly other empirical results. For example, a study that estimates the amount of SARS-CoV-2 in aerosols emitted from coughs of infected patients does not directly address the hypothesis that aerosols are the predominant mode of transmission for Covid-19. The evidential significance of this result will depend on other information like the dispersion of aerosols in various settings, the infectious dose of the virus, and the deposition of aerosols in the respiratory tract. A particular value for the amount of SARS-CoV-2 in aerosols from coughs could support the aerosol transmission hypothesis given certain findings about aerosol dispersion and infectious dose, or it could undermine the hypothesis given other findings. The evidential significance of a result may thus be highly dependent on other empirical evidence as well as theoretical or auxiliary assumptions. In contrast, philosophical discussions of robustness and methodological triangulation have often assumed a picture like this: there are methods or models M_1, $M_2 \ldots M_n$, each of which provides an estimate about a parameter value or some degree of confirmation to a hypothesis (e.g., Kuorikoski and Marchionni 2016; Heesen, Bright, and Zucker 2019). While this picture appropriately characterizes some evidence amalgamation problems, multimodal evidence amalgamation often works differently in that each M_i may estimate different parameters or address different sub-questions relevant for assessing the hypothesis of interest.

A second underexamined aspect of evidence amalgamation is the role that evidential context should play when we combine evidence. By "evidential context," I have in mind information about how data are produced, such as the techniques used for processing data and error characteristics of methods. Such contextual information presumably impacts the evidential bearing of results on hypotheses. As Woodward (2000) points out in a discussion on the evidential significance of data, whether an observation of black ravens supports the hypothesis that all ravens are black depends on how the observation was produced—we can imagine a machine that randomly samples black objects and determines if they are ravens, as well as a machine that randomly samples ravens and determines if they are black. Since the first machine could not have yielded any observations of non-black ravens, it cannot reliably be used to assess the hypothesis that all ravens are black. Woodward (2000) argues that characteristics of data production processes are evidentially significant because they "influence which data outcomes would have been produced had different phenomena claims obtained and hence which conclusions would have been accepted regarding those claims" (175). As an example pertinent to this chapter, the WHO's resistance to recognizing aerosol spread of Covid-19 was partly based on studies that tried and failed to detect viable SARS-CoV-2 from the air (Jimenez et al. 2022). A WHO scientific brief from July 2020 stated that "no studies have found viable virus in air samples" in the course of arguing that Covid-19 is primarily spread through contact and droplets (WHO 2020b). What the brief failed to mention is that it is very difficult to detect viable viruses from air samples due to various methodological issues, such as the fact that aerosol sampling techniques have a tendency to damage viruses and thereby weaken their infectivity in cell culture (Pan, Lednicky, and Wu 2019; Tellier 2022). Some scientists thus argued that the initial failure to detect viable SARS-CoV-2 from aerosols did not in fact undermine the aerosol hypothesis (Greenhalgh et al. 2021). In this case, it seems crucial to consider information about the data production processes to properly assess the evidential significance of the studies. As I will suggest shortly, a weakness of some evidence amalgamation techniques is that they fail to incorporate evidential context.

These features—the highly indirect nature of some evidence and the importance of evidential context—are two reasons why multimodal evidence amalgamation poses such a challenge. Other features can include the presence of discordant results and the large size of an evidence corpus. I now examine the prospects of some prominent amalgamation techniques for assessing multimodal evidence.

2.1 Meta-Analysis

One approach for combining large bodies of evidence is meta-analysis. Meta-analysis can be understood as a set of techniques for integrating results from individual studies to achieve a quantitative estimate of an outcome. Most meta-analyses compute a weighted average of effect estimates from a set of studies. In biomedical science, meta-analyses have often been regarded as providing the best or most reliable type of evidence (Murad et al. 2016). Common justifications are that meta-analyses can yield effect estimates that are both more accurate (closer to the true effect) and more precise (have narrower confidence intervals) than individual studies, resolve discrepancies between primary studies, and sometimes support generalizations across a range of populations or interventions.

One well-known challenge for meta-analysis occurs when dealing with heterogeneity in the primary studies. Researchers performing meta-analyses examine a number of studies on a topic that usually differ methodologically in significant respects. For instance, randomized controlled trials (RCTs) on the effectiveness of a heart medication may rely on populations with varying characteristics, apply different dosage regimens, and use different outcome measures. Some statistical techniques are commonly used for dealing with heterogeneity, most notably random-effects models. These models assume there is a set of effects to be estimated that follow some distribution, as opposed to one true underlying effect. While random-effects meta-analyses can be useful, it is widely believed that they become less reliable as heterogeneity increases (Berlin and Golub 2014; Imrey 2020). Thus, while it may be possible to amalgamate some sets of multimodal evidence using meta-analysis, there will likely be too much heterogeneity to warrant confidence in the results.

However, a more fundamental problem for meta-analysis in the context of multimodal evidence is that often the results we want to amalgamate are (a) not quantitative in character or (b) about different phenomena. For example, epidemiological investigations of Covid-19 transmission events that indicated specific patterns of spread were not quantitative (at least in the sense of reporting an overall effect or outcome estimate) yet proved influential in supporting the aerosol transmission hypothesis (e.g., Shen et al. 2020). Meta-analysis has no way of incorporating non-quantitative results like these. In other situations, we may have quantitative evidence about different phenomena where it does not make sense to summarize the results in a single estimate. For example, estimates of the amount aerosols generated by coughing and the length of time that SARS-CoV-2 remains viable in aerosols are pertinent to the aerosol transmission hypothesis but cannot be pooled through meta-analysis. Because multimodal evidence often consists of results that are about different phenomena and which do not arise from an underlying distribution (as in random-effects meta-analysis), meta-analysis often cannot incorporate relevant evidence.

2.2 Robustness and Methodological Triangulation

Philosophers of science have often promoted robustness analysis and methodological triangulation for assessing diverse evidence. For this discussion, I will assume that robustness analysis is a strategy that uses a set of methods (or means of detection) and examines whether they each support a specific claim about a phenomenon (e.g., as in Schupbach 2018). If the phenomenon claim is supported by all or perhaps most methods, then the claim is usually thought to have particularly strong epistemic support. I will understand methodological triangulation as a related inferential technique that advocates the following: when there are multiple methods that yield conclusions regarding some phenomenon, we should believe the conclusion that is supported by the greatest number of methods (e.g., as in Heesen, Bright, and Zucker 2019).

A major problem for the usefulness of robustness analysis in practice is that multimodal evidence often appears non-robust. While convergent evidence from multiple methods is valuable when it occurs, Douglas (2012) notes that "such clean convergence rarely happens in practice. More common is evidence that diverges, with some evidence indicating one thing and other evidence something else" (140). Stegenga (2009) argues that robustness analysis provides us with little direction in these situations. I agree that robustness analysis is not very informative when evidence diverges. It may tell us to seek to resolve evidential discordance, but this strategy is far from unique to robustness analysis. It may also tell us to suspend belief, but I do not think this is very helpful: scientific advisory bodies need to make recommendations based on evidence currently available, and sometimes non-robust evidence will nevertheless lend significant support for a particular conclusion.

Methodological triangulation in theory provides a way around this problem because it says we should believe the conclusion supported by the greatest number of methods. According to the theorems proven by Heesen, Bright, and Zucker (2019), a triangulation strategy is more likely to yield correct conclusions than simply selecting one method and believing its results. While there may be many scenarios in which the formal model of Heesen, Bright, and Zucker (2019) is appropriate, this strategy encounters several serious problems in practice. First, their formal model assumes that the methods are independent in the sense each method has an independent probability of arriving at the correct answer. In many real-world contexts, there are strong reasons for thinking that this sort of independence does not hold. For example, a well-known issue for climate model ensembles is that the models are often constructed using shared parameterizations and sections of code, and are tuned using the same data sets (Tebaldi and Knutti 2007). As a consequence, the errors or biases of models are expected to be correlated. In an analysis of robustness arguments about climate ensembles, Parker (2011) gives strong reasons to doubt that the kind of independence often assumed by voting-style theorems holds in practice. Schupbach (2018) similarly argues that various probabilistic notions of independence are unlikely to hold in scientific practice. A second issue is that applying the triangulation strategy of Heesen, Bright, and Zucker (2019) requires us to individuate between methods, which as the authors note, may be difficult in practice. Stegenga (2012) argues that it is very difficult to individuate methods in a principled way. Issues like these are why vote-counting methods (e.g., "believe the hypothesis that is supported by the highest number of studies") are usually seen as unreliable in biomedical literature on evidence amalgamation (e.g., Deeks, Higgins, and Altman 2008).

A more conceptual limitation for robustness analysis and methodological triangulation has to do with the indirectness of many relevant sources of evidence. Accounts of robustness analysis and methodological triangulation present situations in which there are multiple methods, models, or means of detection that each provide an indication of the truth of a hypothesis or the value of a parameter. In Heesen, Bright, and Zucker's (2019) account of methodological triangulation, for example, each method endorses one hypothesis from a set of competing hypotheses. In Schupbach's (2018) account of robustness analysis, there are multiple means of detection that each give an indication regarding the truth of a hypothesis. A limitation of these accounts is that they do not incorporate various kinds of indirect evidence. For example, one way that information can be relevant for assessing a hypothesis is by undermining interpretations of data that would support a competing hypothesis. If we have evidence indicating that common aerosol sampling techniques tend to damage viruses and their ability to replicate in cell culture, this seems to suggest that certain experimental findings—namely some failures to detect viable SARS-CoV-2 from aerosols—do not support competing modes of transmission. Accounts of robustness analysis and methodological triangulation do not provide a clear way to incorporate indirect evidence like this. The evidence about the reliability of aerosol sampling techniques cannot serve as a means of detection for the aerosol hypothesis because it does not say anything about the hypothesis. Nevertheless, it is still relevant for assessing the truth of the aerosol hypothesis because it undermines potential support between empirical results and a competing hypothesis. Because robustness and triangulation strategies fail to incorporate indirect evidence, their ability to handle multimodal evidence is limited.

2.3 Bayesian Approaches

Because Bayesian evidence amalgamation methods are highly flexible, my discussion cannot do justice to the full variety of Bayesian techniques and will instead focus on some issues commonly associated with some Bayesian amalgamation techniques. Here I will use the example of estimating equilibrium climate sensitivity, where a variety of Bayesian techniques have been employed. First, the well-known issue of choosing priors has serious practical consequences. Unlike in the ideal scenario in which priors "wash out" in the long run with enough data, the choice of priors in Bayesian assessments of evidence on climate sensitivity has been repeatedly shown to have a strong impact (Frame et al. 2005; Annan and Hargreaves 2011). This has led to considerable discussion among climate scientists about how to select priors, with objections raised against seemingly every proposed strategy for assigning priors. In their discussion of Bayesian techniques for amalgamating evidence about climate sensitivity, Knutti and Hegerl (2008) note that "[s]ubjective choices are part of Bayesian methods, but because the data constraint is weak here, the implications are profound" (740). As a practical matter, Bayesian attempts to amalgamate evidence are often strongly sensitive to prior assumptions about which researchers often disagree.

A second practical problem for Bayesian analyses is that it can often be very unclear how to calculate the relevant probabilities, in particular the likelihood $P(E|H)$. For example, we could ask about the probability of obtaining a particular spatial distribution of people infected with Covid-19 on a bus ride conditional on the hypothesis that aerosol transmission is the predominant mode of transmission. My claim here is not that such probabilities do not exist or that they do not make sense, but rather that researchers often lack a basis for ascribing the precise probabilities that Bayesianism requires. Stegenga (2012) makes a similar point in the service of arguing that the usefulness of Bayesian approaches for amalgamating evidence is highly constrained in practice. If we want our evidence amalgamation approaches to be able to face up to the challenge of non-ideal evidential scenarios, it seems that Bayesian techniques have serious limitations.

One further issue for Bayesian techniques has particular importance in this chapter. Some Bayesian approaches do not incorporate contextual information about data production processes; Woodward (2000) points out that some accounts of Bayesian inference do not attempt to model or incorporate error characteristics of data production processes. If such contextual information holds evidential significance as I have suggested, then viable Bayesian evidence amalgamation techniques will need to incorporate it. To the extent that a technique fails to incorporate contextual information about data production processes, we should be wary of relying on it.

While this discussion of evidence amalgamation approaches is far from exhaustive, the major challenges discussed here tend to be shared by lesser-known approaches. This discussion has helped to identify several desiderata we should look for in a multimodal evidence amalgamation approach. The approach should be able to (1) include various kinds of indirect evidence; (2) incorporate qualitative results and work in contexts where we cannot compute precise likelihoods or effect size estimates; (3) handle discordant empirical results; (4) incorporate information about data generation processes and other aspects of evidential context.

3 Two Cases of Multimodal Evidence Amalgamation

Standard approaches struggle to meet these four desiderata. However, we should not therefore conclude that there are no principled or systematic approaches for amalgamating multimodal evidence. In the two cases I discuss in this section, researchers have been able to successfully amalgamate complex bodies of evidence. My primary aim in this section is to understand what made them successful in overcoming the challenges of multimodal evidence amalgamation.

3.1 Understanding Covid-19 Transmission

Researchers often point out that it is very difficult to acquire direct evidence of how respiratory viruses spread under real-world conditions (e.g., Greenhalgh et al. 2021). The thought here is that it is not feasible to design any single type of experimental or observational study that can reliably assess certain hypotheses about transmission. While in some contexts we may be able to assess hypotheses using a high-quality RCT, when it comes it respiratory virus transmission, it is hard to imagine any experimental procedure that could sufficiently control for all the possible sources of error to reliably discriminate between hypotheses about transmission modes. Establishing knowledge about viral transmission often requires the assessment of multiple indirect lines of evidence.

Epidemiologists generally recognize three major transmission routes for respiratory viruses: contact, droplet, and aerosol spread (Tellier 2009). Following the emergence of Covid-19, a major uncertainty about transmission was whether it spreads primarily through droplets or aerosols. Roughly, droplets are larger bodies that fall to the ground relatively quickly and infect by depositing on exposed mucous membranes, while aerosols are smaller particles that remain aloft and infect through inhalation. A traditional view in epidemiology has been that the boundary between aerosols and droplets occurs at a diameter of 5 microns, but this view has been challenged in recent years, with some researchers proposing a 100-micron boundary (Wang et al. 2021), and others suggesting that a useful size-based threshold does not exist due to the continuous nature of the underlying phenomena and the strong impact of environmental factors on aerodynamic behavior (Drossinos, Weber, and Stilianakis 2021). For the purposes of this discussion, I will assume that aerosols and droplets can be reasonably well distinguished by their aerodynamic behavior.

In 2020, the WHO and CDC initially judged droplet and contact transmission to be the dominant modes of Covid-19 spread. Early reports gave the following reasons as justification: (1) an absence of experimental studies demonstrating aerosol transmission under normal human cough conditions; (2) the inability of epidemiological investigations of transmission events to rule out droplet and contact spread; (3) an inability to detect viable SARS-CoV-2 in aerosols from air samples; (4) the lower basic reproduction number of Covid-19 compared to known airborne viruses such as measles (WHO 2020a; 2020b; Conly et al. 2020). However, some researchers objected to the WHO's and CDC's analyses and argued that aerosols could form a major transmission route (Morawska and Milton 2020; Greenhalgh et al. 2021). With respect to the four reasons above, researchers argued that (1) experimental demonstration of aerosol transmission would be very difficult or impossible; (2) epidemiological investigations indicated patterns of spread that would be much more likely to occur by aerosol transmission than by droplet or contact transmission; (3) the failure to detect viable SARS-CoV-2 from aerosols was not evidence against aerosol transmission; (4) the reproduction number of a virus is not a reliable indicator of its mode of transmission. There was thus substantial expert disagreement about the evidence regarding aerosol transmission during the early stages of the Covid-19 pandemic.

By 2021, however, a robust consensus had developed around the view that Covid-19 spreads predominantly through aerosols. The WHO and CDC officially revised their stances on aerosol transmission in the spring of 2021 (WHO 2021; CDC 2021), a step that researchers widely judged as too late (e.g., Jimenez et al. 2022). For the rest of this section, my aim is to explain how researchers amalgamated the complex evidence regarding aerosol transmission. This analysis focuses on review articles that evaluated evidence regarding aerosol transmission (Wang et al. 2021; Randall et al. 2021; Tellier 2022), as well as some of the major experimental evidence that supports aerosol transmission (Jiang et al. 2020; Shen et al. 2020; Azimi et al. 2021).

First, researchers believed that one of two general hypotheses had to be correct: Covid-19 spreads predominantly through aerosol transmission or predominantly through droplet and contact transmission. This set of hypotheses was thought to be exhaustive in a pragmatic sense: while it is possible to conceive of other hypotheses, background knowledge about respiratory viruses suggested that one of the two had to be correct. While the hypotheses are clearly not exhaustive in a logical sense, this space of hypotheses seems to have been universally accepted among experts, no matter which of the hypotheses they may have been inclined to favor. Reviews and other articles on Covid-19 transmission generally take it for granted that there are only two possible hypotheses at the coarse-grained level of what transmission mode is predominant.

Next, it is worth noting that researchers did not evaluate these two hypotheses in isolation from certain other explanatory claims about empirical results. These other explanatory claims were used to determine the evidential significance of specific empirical results. For example, as I highlighted earlier, the failure of several studies to detect viable SARS-CoV-2 from air samples was sometimes viewed as disconfirming the aerosol transmission hypothesis. One possible explanation for these results is that SARS-CoV-2 transmits only through droplet and contact transmission. However, some researchers argued these results could be readily explained by methodological issues such as the fact that the aerosol sampling techniques tend to damage virions. Thus, the two primary hypotheses about transmission were not the only explanatory claims being assessed. It was necessary to clarify the evidential bearing of empirical results on the two primary hypotheses through additional claims about why those results were produced. As a consequence, larger sets of explanatory claims rather than just the two primary hypotheses were at issue in debates about Covid-19 transmission.

When presenting evidence for aerosol transmission, researchers have sometimes divided the evidence into two categories: evidence that helps establish a plausible mechanistic pathway for aerosol transmission to occur, and evidence that favors the aerosol hypothesis over the droplet-and-contact hypothesis. Since I am focusing on the hypothesis that aerosols are the predominant transmission pathway, I will focus on evidence of the latter sort. A major review by Wang et al. (2021) argued that a "growing body of research on COVID-19 provides abundant evidence for the predominance of airborne transmission of SARS-CoV-2" (2). This review pointed to several phenomena as crucial pieces of evidence: the occurrence of long-range Covid-19 transmission, a large difference between transmission risk outdoors versus indoors, and a strong effect of ventilation on transmission indoors. Other reviews also tend to highlight these phenomena as providing the strongest evidence for the hypothesis that aerosols are the main transmission route (Tellier 2022).

What exactly is the inferential connection between these phenomena and the aerosol hypothesis? While the review articles are not fully explicit, here is my best reconstruction. The occurrence of long-range airborne transmission establishes that aerosol transmission occurs because droplets can only travel a few meters in the air, though it does not entail that aerosols are the predominant transmission mode. The two other phenomena imply that aerosol transmission is predominant in real-world settings. If the droplet-and-contact hypothesis were correct, we would not see a large difference in transmission risk outdoors versus indoors because the aerodynamics of droplets are essentially the same indoors and outdoors. At most, we would see a minor difference in transmission risk. Thus, the large difference in outdoor versus indoor transmission entails that aerosol transmission is predominant. Similarly, under the droplet-and-contact hypothesis, we would not see a strong effect of ventilation on indoor transmission risk because droplet behavior is not affected by ventilation. Thus, the strong effect of ventilation is only compatible with the aerosol hypothesis.

The reasoning sketched above is an example of phenomena (or phenomena claims) being used as evidence for a physical hypothesis. For a complete evidential story, we would need to consider how data justified these phenomena claims. Here, I will just note that the three key phenomena claims were widely judged to have been well-established by the end of 2020. While the WHO correctly noted that early studies reporting long-range transmission could not rule out droplet and contact transmission, the accumulation of carefully investigated long-range transmission events was widely viewed as establishing long-range transmission by aerosols. Similarly, the large differences between indoor and outdoor transmission and the strong effect of ventilation came to be seen as well-established by the end of 2020 (Morawska and Milton 2020; Bhagat et al. 2020; Jimenez et al. 2022).

To summarize, the successful amalgamation of evidence about Covid-19 transmission can be divided into several phases. In the first phase, researchers demonstrated that the aerosol hypothesis was not ruled out by early results such as failures to detect viable SARS-CoV-2 in the air. This contradicted early assessments by the WHO and others that concluded that most transmission was attributable to droplets and contact. In the next phase, researchers attempted to eliminate the droplet-and-contact hypothesis by appealing to phenomena inconsistent with that hypothesis. Since these phenomena came to be seen as empirically well-established, it was no longer possible to reconcile the droplet-and-contact hypothesis with the totality of evidence.

3.2 Estimating Earth's Equilibrium Climate Sensitivity

The equilibrium climate sensitivity (ECS) of Earth is an important quantity in climate science. It is roughly defined as the amount of global warming that will occur at Earth's surface in response to a doubling of atmospheric CO_2 from pre-industrial levels, once the climate system has reached a new equilibrium or stable temperature (Knutti 2008). In other words, ECS is a measure of how sensitive Earth's surface temperature is to increases in CO_2 in the long run. It has been a key source of uncertainty in long-term temperature projections, and narrowing uncertainty about ECS is a critical task for understanding long-term climate change. Sherwood et al. (2020) explain that this is a daunting task:

> Quantifying ECS is challenging because the available evidence consists of diverse strands, none of which is conclusive by itself. This requires that the strands be combined in some way. Yet, because the underlying science spans many disciplines within the Earth Sciences, individual scientists generally only fully

> understand one or a few of the strands. Moreover, the interpretation of each strand requires structural assumptions that cannot be proven, and sometimes ECS measures have been estimated from each strand that are not fully equivalent. This complexity and uncertainty thwarts rigorous, definitive calculations and gives expert judgment and assumptions a potentially large role. (Sherwood et al. 2020, 2)

This passage helps illustrate why the estimation of ECS is a compelling case study for multimodal evidence amalgamation. Determining ECS has proven particularly difficult, with climate scientists reporting "disturbingly little progress in narrowing the large uncertainty range" over recent decades (Knutti 2008).

Three main lines of evidence can be used to derive estimates of ECS: the historical climate record, the paleoclimate record, and feedback process understanding. The first line of evidence uses the trajectory of warming as tracked by the instrumental record (going back to approximately 1850) in conjunction with an energy balance model and data about CO_2 emissions to infer ECS. The paleoclimate record can be used to estimate ECS by using measurements of global mean temperature and other climate variables during certain periods in Earth's history in conjunction with modified energy balance models. In contrast to these two methods for estimating ECS, feedback process understanding attempts to directly estimate the impact of climate feedbacks on warming, such as the positive feedbacks of increased water vapor and decreased surface ice. These estimates can be combined through an energy balance model to estimate ECS. The evidence for estimating feedbacks is diverse. It includes information from global climate models (GCMs), process-resolving models, and theoretical understanding (Sherwood et al. 2020).

Published estimates of ECS from each line of evidence are broadly concordant, tending to fall between 1.5°C and 4.5°C. The IPCC's Fifth Assessment Report, released in 2013, determined that ECS was "likely" to fall in this range using a somewhat opaque process of expert judgment (Collins et al. 2013, 1111). However, climate scientists have frequently hoped to narrow uncertainty about ECS by systematically combining various lines of evidence. By the time of the Fifth Assessment Report, several studies had attempted to do this using Bayesian techniques or expert elicitations. However, the IPCC judged these studies to be unreliable for several reasons. First, there is no agreed upon statistical method for combining estimates of ECS from the various lines of evidence (Collins et al. 2013, 1111). This is partly because Bayesian techniques are highly sensitive to prior assumptions and to the particular technique for selecting priors (Frame et al. 2005; Knutti and Hegerl 2008). Additionally, many of these studies assumed that the lines of evidence are independent from one another, which can yield overconfident assessments (Knutti and Hegerl 2008; Knutti et al. 2010; Collins et al. 2013).

More recently, a team of climate scientists proposed another approach for amalgamating evidence about ECS (Stevens et al. 2016). The core of this approach is "the development and eventual refutation of the physically based—hence testable—'storylines' that would enable a given bound for ECS to be violated" (Stevens et al. 2016, 514). In a nutshell, the goal is to consider potential physical explanations of why the evidence appears as it does under hypothetical values for ECS, and determine which of those potential explanations are incompatible with the evidence. If we can rule out all the potential explanations tied to a given ECS value, then we can rule out that ECS value. For example, we can try to determine whether the evidence rules out ECS values below 1.5°C by first identifying what would have to be the case physically for those values to be true, such as clouds having a net negative feedback on global warming (Stevens et al. 2016). Then we can examine whether those physical scenarios are compatible with the evidence corpus by trying to find specific pieces of evidence that bear on the physical scenarios in question. Stevens et al. (2016) argue that achieving progress on ECS requires a shift in focus "away from arguments in support of particular ECS values, and a move toward efforts to rule values out through a careful process of hypothesis testing" (512).

In 2020, this approach was put into practice in an exhaustive and detailed analysis by

Sherwood et al. (2020). This analysis, along with the "storyline" reasoning discussed by Stevens et al. (2016), subsequently figured heavily in the IPCC's analysis of ECS in its Sixth Assessment Report (P. Forster et al. 2021). The Sherwood et al. (2020) study was able to rule out ECS values below 2°C with a high degree of confidence, improving upon prior determinations of the plausible ECS range. The Sixth Assessment Report narrowed its previous estimate of the "likely" range for ECS to between 2.5 and 4.0°C. While the Sixth Assessment Report attributed this to improvements in physical understanding and data sources, it also noted that in the Fifth Assessment Report "these lines of evidence were not explicitly combined in the assessment of climate sensitivity, but as demonstrated by Sherwood et al. (2020) their combination narrows the uncertainty ranges of ECS" (P. Forster et al. 2021, 993). Thus, the Sherwood et al. (2020) analysis appears to be a successful case of evidence amalgamation that helped drive consensus around a narrower range of ECS. In the rest of this section, my aim is to explain the "storyline" approach and suggest why it has proven more influential than other evidence amalgamation techniques for assessing ECS.

The "storyline" approach as characterized by Stevens et al. (2016) can be broken down into two basic stages. In the first stage, researchers develop explanations of what would have to physically be the case for a hypothetical value of ECS to hold. For example, we can ask what Earth's climate feedback processes would have to be like for global mean temperature to be relatively insensitive to CO_2, corresponding to an ECS value of 1.5°C or lower. Among other things, Stevens et al. (2016) identify that there would have to be negative cloud feedbacks from warming. We can also ask about information associated with other lines of evidence, such as what the climate would have to have been like during the last glacial maximum period if ECS is below 1.5°C. The overall goal of this stage in the approach is to identify a set of physical circumstances needed to hold for various values or ranges of ECS.

The next stage in the storyline approach is to use the lines of evidence to rule out as many of the physical circumstances as possible for each ECS interval. Ruling out a physical storyline involves showing that evidence renders it implausible. The evidence used to rule out a physical storyline can come from one or multiple lines of evidence. An apparent advantage of this approach compared to some other techniques for amalgamating evidence about ECS is that it allows researchers to rule out certain values of ECS even if those values are plausible on some lines of evidence. For example, if an ECS value of 5°C is consistent with lines of evidence A and B, but ruled out by C, the storyline approach says we can treat 5°C as ruled out by evidence corpus. Sherwood et al. (2020) write: "Whatever the true value of S is, it must be reconcilable with all pieces of evidence; if any one piece of evidence effectively rules out a particular value of S, that value does not become likely again just because it is consistent with some other, weaker, piece of evidence as long as there are other S values consistent with all the evidence" (74). Of course, evidence can be misleading or misinterpreted, such that we could wrongly rule out a value of ECS based on one line of evidence. Sherwood et al. (2020) do not seem to see this as a serious problem for their analysis because they perform sensitivity tests for structural uncertainties associated with each line of evidence. For them, demonstrating that evidence rules out a physical storyline requires taking into account structural uncertainties.

Stevens et al. (2016) and Sherwood et al. (2020) recognize that we might wish to quantify our confidence that evidence refutes a physical storyline. Stevens et al. (2016) suggest using Bayesian inference as a useful framework here. They view the storyline approach as distinct from Bayesian inference and one that in principle could be carried out qualitatively or with other statistical techniques. As part of their study, Sherwood et al. (2020) include a Bayesian analysis to calculate a posterior probability distribution for ECS and also perform sensitivity tests for different choices of priors and structural uncertainties. They view this analysis as a useful quantitative summary of the evidence that is secondary to the overall storyline approach:

> It must be remembered that every step of this process (choosing priors, computing likelihoods, etc.) involves judgments or models, and results will depend on

> assumptions and assessments of structural uncertainties that are hard to quantify. Thus, we emphasize that a solid *qualitative* understanding of how the evidence stacks up is at least as important as any probabilities we assign. (Sherwood et al. 2020, 73)

The storyline approach helps to structure the Bayesian analysis and make explicit the evidential reasoning for the plausible range of ECS.

With respect to its conclusions, the main advance of the Sherwood et al. (2020) analysis was to effectively rule out ECS values below 2°C. Sherwood et al. (2020) attribute their ability to constrain ECS more than previous assessments to their methodology and recent refinements in physical understanding. In their analysis, one of the physical circumstances required for an ECS value below 2°C is that clouds would have to have net negative feedback on global warming. Feedbacks from clouds are widely judged to be less well constrained than other feedback processes. Sherwood et al. (2020) inferred from the small uncertainties in other feedback processes that cloud feedback would have to be negative for ECS to be below 2°C. Next, they argued that multiple sources of evidence effectively rule out negative cloud feedback. The evidence here included numerical models of clouds, emergent-constraint studies, and observations of cloud behavior in response to historical warming. Their key argument is that reconciling "all of the above evidence with an overall *negative* feedback from clouds... would require multiple lines of evidence to have failed significantly for at least one cloud type" (Sherwood et al. 2020, 37).

In summary, Sherwood et al.'s (2020) use of the "storyline" approach for amalgamating climate evidence described in Stevens et al. (2016) represents a major advance in estimating ECS. This approach differs from previous attempts to amalgamate evidence about ECS through its systematic use of physical storylines and refutational reasoning.

4 An Explanatory-Eliminative Approach

While the cases discussed in this chapter come from very different areas of science, they share important features. Before moving to a more detailed discussion, I highlight several of these features. First, researchers developed and assessed explanatory accounts of the evidence—accounts that explain in detail why the evidence appears as it does—instead of pursuing more narrow goals such as identifying the correct physical hypothesis or estimating a certain quantity. Second, eliminative or refutational reasoning played a crucial role in how the researchers combined evidence. In both cases, the attempt to rule out hypotheses, as opposed to focusing more narrowly on what the evidence supports, contributed to empirical advances. Third, while researchers used quantitative analyses at times, the overall approaches were qualitative in that (a) the relevance of the diverse modes of evidence to hypotheses was established through qualitative arguments; (b) any quantitative analyses of evidence were used as part of the broader eliminative strategy. At an abstract level, the same approach to amalgamating evidence is at work in both cases. In the next section I articulate that approach.

4.1 Explanation in Evidence Amalgamation

Philosophers of science have long viewed explanation as a central goal of scientific inquiry. However, they usually recognize that other goals such as prediction and intervention can take priority in some contexts. In what sense is explanation a goal of evidence amalgamation? Some amalgamation methods do not seem to have explanation as a goal. The goal of a meta-analysis is to estimate an effect size as accurately as possible from a set of studies. Whether or not a meta-analytic estimate explains something about the evidence or phenomena is not intrinsic to the task of meta-analysis. While a meta-analytic estimate may be used within a broader explanatory process of explanatory reasoning, a meta-analysis does not have an intrinsically explanatory aim. Similar claims could be made about amalgamation techniques like Bayesian model averaging and multi-model ensemble forecasting. Thus, explanation is not always a central aim of evidence amalgamation, and some methods may be better suited for explanatory tasks than others.

There are at least two distinct ways in which an evidence amalgamation process could be explanatory. First, an amalgamation process could rely on explanatory power or other explanatory criteria in assessing hypotheses. Such a process would involve some form of inference to the best explanation (IBE). This is the usual way that "explanatory" is understood in the context of scientific inference; I will refer to this notion as *explanatory₁*. Douglas (2012), for example, outlines what she describes as an "explanatory approach" to weighing complex evidence, and Douglas explicitly frames the approach as a version of IBE. The association between explanation and IBE has become so strong that the term "explanationist" has come to label someone who advocates for IBE or gives it a central role in their thinking (e.g., Saatsi 2017). However, there is another substantive sense of "explanatory" that is easily missed, and that distinguishes certain amalgamation techniques that do not use IBE from non-explanatory methods such as meta-analysis.

The second way an amalgamation process could be explanatory is this: it constructs and considers explanations of why the full body of evidence appears as it does in order to arrive at a conclusion (*explanatory₂*). Let me clarify what I have in mind by starting with a schematic example. Suppose there are three experimental methods (M_1, M_2, M_3) that each gives an indication about which of two competing hypotheses (H_1, H_2) is true. Methods M_1 and M_2 support H_1, while M_3 supports H_2. Here is a simple procedure to amalgamate the evidence: accept the hypothesis supported by the greatest number of methods. In this case, the vote-counting procedure tells us to accept H_1. However, this procedure does not explain why the body of evidence appears as it does. Most notably, it does not explain why the evidence is discordant or why we should believe that method M_3 is mistaken. While we could develop explanations for these after the vote-counting procedure, the explanations do not play a role in the process of determining what hypothesis to accept. An explanatory$_2$

amalgamation method, on the other hand, would develop and consider these explanations in arriving at a result. The central idea is that developing and considering explanations of the evidence is internal to the process of determining what to conclude from the evidence.

To illustrate this using a real-world example, consider how researchers dealt with discordant evidence on how Covid-19 is transmitted. As I mentioned earlier, several studies tried and failed to detect viable SARS-CoV-2 in aerosols emitted by infected patients; these results appeared to conflict with other findings that supported aerosol transmission. One strategy would have been to "weigh" all the results together along similar lines as the vote-counting procedure I just mentioned. Instead, some researchers considered explanations for the discordance and incorporated those explanations into the amalgamation process. One possible explanation, in brief, was that the studies that failed to detect viable SARS-CoV-2 in aerosols relied on aerosol sampling techniques that tend to damage viruses and reduce their infectivity in cell culture. Another was that the studies that appeared to support aerosol transmission—such as investigations of some super-spreading events—failed to adequately rule out droplet and contact transmission. Researchers considered and assessed these explanations in the process of amalgamating evidence about Covid-19 transmission. In doing so, they assessed full explanations of the evidence rather than only assessing physical hypotheses about Covid-19 transmission that, on their own, are not capable of explaining the full set of empirical results.

So far I have only used examples with discordant evidence to illustrate how an amalgamation process could be explanatory without involving IBE. In these examples, the contrast with non-explanatory methods such as meta-analysis and vote-counting algorithms is more noticeable than when the evidence is concordant. However, an explanatory$_2$ amalgamation process can also be applied to a concordant body of evidence. The "storyline" approach used by Sherwood et al. (2020) to estimate the Earth's climate sensitivity worked on a broadly concordant evidence base. What made this an explanatory$_2$ process was that it constructed and assessed explanations for the full body of empirical results under hypothetical values of

ECS. Many previous estimates of ECS that combined multiple lines of evidence did not do this. For example, the IPCC's Fifth Assessment Report applied a somewhat vague process of expert judgment to estimate the likely range of ECS using probability density functions produced by climate models and several other sources of evidence (Collins et al. 2013, 1110–1112). The IPCC's process was effectively to delineate a region of consensus on estimates of ECS, as opposed to assessing explanations of the evidence under different hypothetical values of ECS (Sherwood and Forest 2024).

An evidence amalgamation process can be explanatory$_2$ without being explanatory$_1$. While the role of explanations in amalgamating evidence in the Covid-19 and climate sensitivity cases might tempt us to see them as examples of IBE, that would be a mistake. A non-trivial notion of IBE requires explanatory quality to be used as a guide to the likeliness or truth of an explanation, as Lipton (2004) notes. This did not occur in either of the cases. We can see this clearly in the Covid-19 transmission case. Researchers did not make a judgment about the superior explanatory qualities of the aerosol transmission hypothesis and then infer that it is likely correct. Rather, they attempted to eliminate the droplet-and-contact transmission hypothesis by looking for features in the evidence corpus that are incompatible with what we would expect to see if that hypothesis were correct. Similarly, climate researchers did not assess ECS by using explanatory quality as a guide to truth. In neither case did researchers make any appeals to explanatory criteria associated with IBE to justify their conclusions. While it is possible for an amalgamation process to be explanatory in both of the senses I have described, an explanatory$_2$ process need not rely on any judgments about explanatory quality to guide inference.

Explaining why a body of evidence appears the way it does requires more than a single physical hypothesis. In particular, it requires contextual information about data production processes. Because there may be genuine uncertainty about aspects of a data production process, potential explanations of evidence often differ in their claims about data production processes in addition to their primary physical hypotheses. Explanations may also involve secondary physical hypotheses. For these reasons, I use the term *explanatory accounts* to label a set of claims used for explaining an evidence corpus, following Douglas's (2012) usage. Explanatory accounts typically involve multiple explanations linked under a primary physical hypothesis. For example, the explanatory account for the aerosol transmission hypothesis of Covid-19 needed to explain, in addition to results that appeared to support it, why (1) several studies failed to detect viable SARS-CoV-2 from aerosols; (2) SARS-CoV-2 has a lower basic reproduction number than most known airborne viruses such as measles; (3) surgical masks are somewhat effective in reducing transmission even though they are not designed to stop aerosols. Often, numerous features of a body of evidence stand in need of explanation, and the basic task of an explanatory account is to coherently say why the evidence has these features.

When evidence is highly diverse or multimodal, explanatory accounts will tend to be complex, since they need to explain results from various modes as well as any apparent discordances between modes. The complexity of explanatory accounts may seem disadvantageous from a practical perspective. However, while explanatory accounts themselves can be highly complex, the need to explain a diverse body of evidence provides a substantial source of constraint on the space of plausible explanatory accounts. Constructing a coherent and empirically competent account of the evidence tends to become more difficult with more evidential diversity. This is because additional modes of evidence can produce results that an existing explanatory account cannot explain while staying coherent and empirically competent. The constraining power of additional modes of evidence is an interesting feature of an explanatory$_2$ amalgamation approach. On some other approaches, such as methodological triangulation and vote-counting procedures, evidential diversity does not have this constraining effect. For example, in the vote-counting algorithm described earlier, an additional mode of evidence will be helpful in the sense that it provides another vote, but it does nothing to constrain the space of possible hypotheses. Thus, an explanatory$_2$ amalgamation process will often be simpler than other amalgamation methods insofar as it has fewer plausible hypotheses to evaluate.

Finally, an explanatory$_2$ process is particularly well suited for amalgamating multimodal evidence. I have already mentioned why meta-analysis does not fare well with multimodal evidence. Philosophers of science have turned to robustness analysis and methodological triangulation as methods for amalgamating multimodal evidence. As I noted in Section 2, an assumption of these methods is that each evidential mode estimates a common parameter or provides an indication regarding the truth of the hypothesis; however, in scientific cases involving multimodal evidence, evidential modes may estimate different parameters or address different sub-questions relevant for assessing the hypothesis. Robustness analysis and methodological triangulation are not capable of incorporating this kind of indirect evidence. By constructing explanatory accounts of the full body of evidence, an explanatory$_2$ amalgamation process can incorporate this information. Explanatory accounts can incorporate indirect evidence through the use of auxiliaries such as background theoretical claims and models. Thus, an explanatory$_2$ amalgamation process can overcome a major limitation of prominent amalgamation methods when faced with multimodal evidence.

4.2 Eliminative Reasoning

On its own, the development of explanatory accounts of evidence cannot offer what we want from an evidence amalgamation process. There needs to be a way to reliably infer accurate conclusions about what the evidence indicates. Eliminative reasoning promises one such strategy. Let me first address some terminology. I use "eliminative reasoning" to refer to any form of reasoning that proceeds by attempting to render implausible potential hypotheses or explanations. This is a broad notion that can describe many kinds of attempts to rule out empirical possibilities. A special case of eliminative reasoning is an "eliminative induction." A successful eliminative induction requires an exhaustive space of hypotheses, at least in the pragmatic sense that all plausible hypotheses are included. Eliminative induction is typically thought of as eliminating all but one of the hypotheses, though I will understand it in a weaker sense as eliminating *at least* one hypothesis. (If we think of the non-eliminated hypotheses as a single disjunctive hypothesis, the traditional conception is equivalent to the weaker sense.) Thus, one can understand the Sherwood et al. (2020) analysis of climate sensitivity as an eliminative induction because they ruled out some ECS values. Finally, I understand a "Holmesian inference" is a special case of eliminative induction in which all hypotheses except for one are eliminated. Of course, these distinctions are sensitive to the partitioning of the hypothesis space, but they are useful for describing inferences when there is a partition.

Philosophers of science have offered differing accounts of the nature of the elimination step in eliminative reasoning. Some have described it as a deductive inference. For example, Bird (2005; 2010) argues that Holmesian inference involves the deductive elimination of explanatory possibilities. The evidence needed to effect such an inference must be logically inconsistent with all but one of the hypotheses. In response to this sort of deductive picture, Earman (1992) suggests that "in scientific cases the elimination may not consist of the simple one-two knockout deduction of a prediction and falsification of the prediction via direct observation," but that some kind of induction is usually required to confront hypotheses and data (170). I share Earman's skepticism about accounts that require the elimination of explanatory possibilities to be deductive, and I think they do not accurately describe the eliminative reasoning in the Covid-19 and climate sensitivity cases. Earman suggests that hypotheses can be "probabilistically eliminable" in the sense that evidence can drive their probability low enough that they are ruled out for all practical purposes. Norton (1995) similarly describes an eliminative process that can be driven by inductive arguments. While I cannot argue it here, I think that any realistic account of eliminative reasoning in science needs to allow for the inductive elimination of hypotheses.

In evidence amalgamation, an eliminative process cannot work by generating predictions and gathering further data with the aim of falsifying predictions, as it is sometimes described. Instead, any elimination must rely on the existing body of evidence. This is a substantial limitation that threatens to make eliminative reasoning rather useless for evidence amalgamation. Things are not this grim, however. For one thing, any evidence amalgamation process is limited to existing results. On the positive side, though, an existing set of results can often enable the elimination of some hypotheses or explanatory accounts that are coherent and empirically competent. The explicit construction of explanatory accounts can help researchers identify previously unrecognized arguments for the elimination of hypotheses. In other words, when coupled with an explanatory$_2$ process, eliminative reasoning is capable of narrowing the space of explanatory possibilities.

An eliminative process is particularly useful for amalgamating multimodal evidence because it can produce stronger or more informative conclusions than some other approaches capable of handling multimodal evidence. For example, suppose there are three methods (M_1, M_2, M_3) for estimating a parameter P that yield the following ranges of possible values for P, respectively: $[0, 3]$, $[0, 3]$, and $[0, 2]$. Accounts of robustness analysis and methodological triangulation imply that the proper conclusion from these results is that P has a value between 0 and 3. This conclusion is consistent with all three methods, and it is also endorsed by the majority of the methods. For an eliminative process, though, there may be sufficient evidence to conclude that the value of P is between 0 and 2. This is because whatever evidence underlies M_3's estimate that P is between 0 and 2 may be sufficient for ruling out values of P greater than 2.[2] More precisely, in an explanatory$_2$ framework, the evidence that underlies M_3's estimate may be sufficient to rule out all explanatory accounts on which the value of P is greater than 2. An eliminative process can yield stronger (but still well-justified) conclusions than robustness analysis and methodological triangulation because the evidential significance of a result is not diluted by a consensus-seeking process among methods.

2. There is a straightforward way to illustrate this intuition in a Bayesian framework. If each method produces a likelihood function for parameter P instead of a simple range, we can imagine a scenario where M_1 and M_2 yield likelihood functions with non-zero values for P only in the interval $[0, 3]$, while M_3 yields a likelihood function with non-zero values only in the interval $[0, 2]$. If the methods are independent such that we can obtain a joint likelihood by multiplying these likelihood functions, then the joint likelihood and the posterior distribution will not have any non-zero values for values of $P > 2$, regardless of the prior.

These considerations suggest an interesting possibility: that an amalgamation process that employs eliminative reasoning benefits from having a large number of evidential modes. While the same can be argued about robustness and methodological triangulation, an eliminative process can employ indirect sources of evidence, unlike these approaches. Additional modes of evidence tend to provide more material for devising inductive arguments to eliminate explanatory possibilities. An eliminative process may thus be particularly well-suited for amalgamating diverse multimodal evidence.

4.3 The Explanatory-Eliminative Approach

I now address how the explanatory and eliminative processes described above can be combined into an approach for amalgamating evidence. The overall approach can be summarized as a two-stage process: Generate explanatory accounts of the full set of evidence, and then attempt to eliminate as many of those as possible. Of course, the usefulness of the approach depends on the details. My aim here is to sketch an amalgamation framework that is general enough to be applied whenever we wish to amalgamate multimodal evidence but that has enough detail to amount to a distinctive and useful approach.

In the first stage, researchers need to do several things: (1) identify the full set of empirical results to be explained; (2) construct explanatory accounts of the evidence that pass basic adequacy criteria; (3) check that *all* explanatory accounts that pass basic adequacy criteria are included. I address each of these in turn.

Researchers first need to identify a set of empirical results that explanatory accounts will need to explain. This is a potentially tricky task because researchers may disagree about what evidence is relevant and some results carry a high risk of bias or error. These two issues, which I explore in depth in Chapter 4, are particularly challenging for quantitative methods such as meta-analysis. However, the explanatory-eliminative approach can largely avoid these issues in the evidence selection process because they are dealt with in the construction of explanatory accounts, which I address later in this section. In short, the explanatory-eliminative approach tells researchers to include all empirical results that are deemed explanatorily relevant by at least some members of the research community. This inclusive strategy helps the explanatory-eliminative approach to avoid charges of bias and incompleteness in the selection of evidence. Of course, this also means that some of the included results may have dubious validity and relevance.

What does it mean for results to be explanatorily relevant? As I understand the term, explanatorily relevant results will be a subset of all evidentially relevant results. In the Covid-19 case, I argued that the fact that common aerosol sampling techniques tend to damage viruses and weaken their infectivity in cell culture was evidentially relevant to assessing hypotheses about Covid-19 transmission. However, this fact cannot be explained by these hypotheses. The Covid-19 transmission hypotheses did not need to explain this information even though it turned out to be evidentially relevant to assessing them. On the other hand, the Covid-19 transmission hypotheses needed to explain many other findings, like the difference in transmission risk outdoors versus indoors. As this example highlights, hypotheses only "need" to explain a subset of all evidentially relevant results. The challenge is to capture this idea in a more precise way. I suggest that a result is explanatorily relevant to a set of hypotheses when it is (a) evidentially relevant to each hypothesis, and (b) potentially explainable by each hypothesis (in conjunction with appropriate auxiliaries). The latter condition distinguishes evidentially relevant results that cannot (and need not) be explained by a hypothesis from those that are explanatorily downstream of the hypothesis.

Of course, the notion of evidential relevance is hypothesis dependent. This implies that identifying the explanatorily relevant evidence requires researchers to first identify a set of hypotheses. This creates a potential problem: how can researchers be sure that the set of hypotheses is exhaustive, as required for successful eliminative induction? The initial set of hypotheses may not be exhaustive, especially if researchers have to identify them prior to working through the evidence. I have several things to say here. First, ensuring that the primary set of hypotheses is exhaustive is sometimes not very difficult. In the climate sensitivity case, there are infinitely many potential values of ECS, but they can be finitely partitioned so that we know the true value has to lie in one of the subsets. Second, background knowledge typically allows researchers to reliably judge when the hypothesis space is exhaustive. I mentioned in Section 3.1 that Covid-19 researchers were confident that one of the following hypotheses had to be true: Covid-19 spreads predominantly through aerosol transmission or predominantly through droplet and contact transmission. The background knowledge in this case comes from the germ theory of disease and basic principles of infectious respiratory illnesses. Of course, it is possible for these judgments to be mistaken; I address this issue later in this chapter.

The next phase of the explanatory-eliminative approach is to construct explanatory accounts of the evidence based on the set of primary hypotheses. Before addressing the structure of explanatory accounts, I identify basic adequacy criteria that all accounts need to meet. Here I largely follow Douglas (2012) in emphasizing the following criteria: coherence, empirical competence, and predictive potential. Coherence here is the idea that the claims that an account uses to explain results must be consistent with one another. Douglas (2012) notes that this sort of coherence is more difficult to achieve than it might appear because explanatory accounts "often contain methodological critiques of some studies (e.g., that confounders were not properly controlled, or that a particular method of dosing is not appropriate)," and these claims have to be applied consistently across studies, including to those with results that seem supportive (153). Empirical competence is the idea that an account must offer a potential explanation of all of the (explanatorily relevant) results. (My understanding of empirical competence differs from Douglas, who views it as consistency with the evidence.) If there is a result that an account fails to provide any potential explanation for, the account is deficient. Additionally, when there are discordant results, empirical competence requires that accounts must offer an explanation of the discordance. An account that does not explain apparent discrepancies may be consistent with the evidence but fails to fully explain why the evidence corpus appears as it does. The requirement of empirical competency thus limits the space of explanatory accounts to those that offer potential explanations for discordant results. Finally, the requirement of predictive potential means that accounts must be able to offer testable predictions, rather than merely conforming to existing data.

Each explanatory account must link a primary hypothesis to the set of explanatorily relevant results. Explanatory accounts typically do so by making phenomena claims that are based on the primary hypothesis and various auxiliary claims. Phenomena claims are usually needed to explanatorily link a primary hypothesis with results. For example, one phenomenon claim associated with the Covid-19 aerosol hypothesis is that ventilation will have a strong effect on transmission risk. This claim can be derived from the hypothesis along with various auxiliaries about the physics of aerosols. It can then be used to help explain results from certain experimental and observational studies on Covid-19 transmission. Thus, the primary hypothesis typically explains results with the help of intermediate phenomena claims. Figure 1 below illustrates the basic structure of a hypothetical explanatory account.

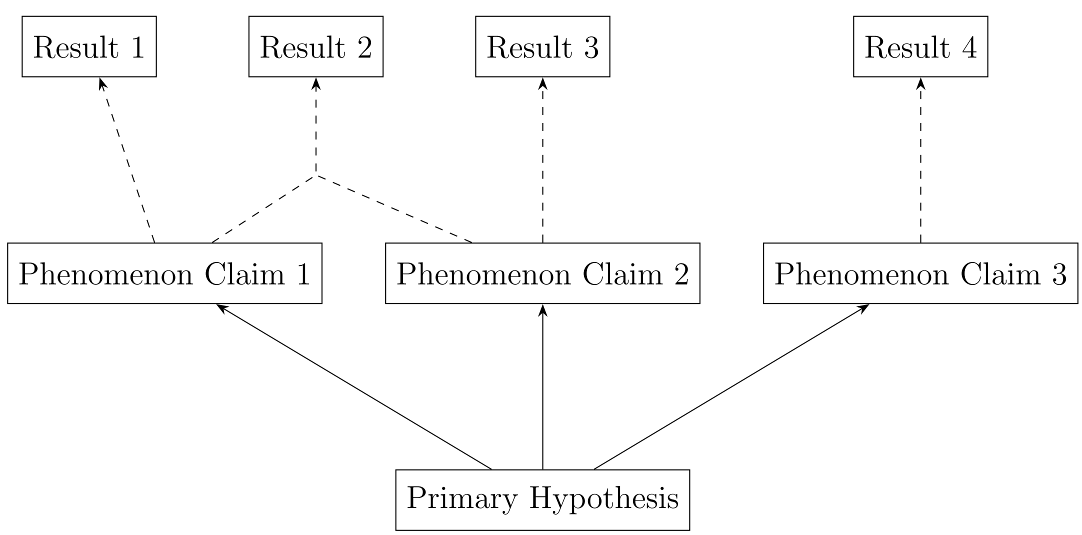

Figure 1: This figure shows the basic structure of a hypothetical explanatory account, omitting the role of auxiliaries. Solid arrows represent explanations of phenomena claims, and dashed arrows represent explanations of results.

Auxiliary information is typically needed to derive phenomena claims from a primary hypothesis, as well as to explain empirical results based on phenomena claims. I discuss auxiliary information in depth in Chapter 3, and will just make a few points about it here. In Figure 1, we can imagine each arrow as being associated with its own set of auxiliaries that help explain the respective result or phenomenon claim. These auxiliaries should be described explicitly in the construction of explanatory accounts. Of course, auxiliary claims can be wrong and can carry substantial uncertainty, which is important for the assessment of explanatory accounts. But during the construction phase, researchers should not focus on the level of support for auxiliaries and should instead develop explanatory accounts even if there is uncertainty or doubt about auxiliaries. This helps ensure that researchers consider explanatory possibilities that are far-fetched or that have low prior probability. Finally, competing explanatory accounts will necessarily differ in some of the auxiliaries they use to explain the results. These differences can be useful when trying to eliminate explanatory accounts.

How can researchers be sure that the explanatory accounts they have developed are exhaustive in the sense required for eliminative induction? It seems there is no guarantee that the truth cannot lie in some "unconceived alternative" account, to use Stanford's (2006) expression. I will make a few conceptual points about this. First, unless an eliminative induction is based on a logically exhaustive hypothesis space, there is no way to guarantee that the truth is inside the hypothesis space. Any eliminative induction not based on a logically exhaustive hypothesis space will need to rely on some claims or assumptions to bound the hypothesis space, which will be defeasible. However, all inductive reasoning is defeasible, so this is not a unique feature of eliminative inference. Stanford's argument about unconceived alternatives gets its force from numerous historical examples that indicate that scientists often fail to imagine alternatives that turn out to be correct. Any program of eliminative inference will need to provide some practical mechanisms for minimizing this prospect. However, there is no way to completely do so while preserving the inductive nature of the inference.

The explanatory-eliminative approach tries to minimize the risk that the correct explanation is not considered through several strategies. First, researchers should consider all primary hypotheses consistent with background knowledge and pursue explanatory accounts based on them. If for a hypothesis no explanatory accounts can be constructed that pass basic adequacy criteria, then the hypothesis does not warrant further consideration. Second, when attempting to explain results based on phenomena claims, and deriving phenomena claims based on a primary hypothesis, researchers need to consider all relevant auxiliaries consistent with background knowledge. If there are multiple sets of auxiliaries consistent with background knowledge that can be used to explain a particular result, this will create multiple explanatory accounts. Of course, an issue for these two strategies is what counts as background knowledge. I cannot resolve that here but will say that background claims should have a rather high level of support to be treated as knowledge. An example would be how researchers in the Covid-19 case did not consider a miasma theory of transmission, which is inconsistent with the germ theory of disease, which has a very high level of support. The construction of explanatory accounts is essentially a search process to find and describe all explanatory possibilities that (a) pass basic adequacy criteria and (b) are consistent with background knowledge.

Once researchers have generated a set of explanatory accounts, the next goal is to rule out as many of them as possible using any relevant evidence. Let me first address in structural terms how explanatory accounts are ruled out. Eliminating explanatory accounts involves breaking their explanatory links by showing that the links are implausible. To eliminate an explanatory account, it suffices to show that *any* of its explanatory links is implausible. Doing so entails that at least one of the results is not successfully explained by the account. As an illustration, Figure 2 depicts a broken explanatory link between the Primary Hypothesis and Phenomenon Claim 3. The explanatory account in Figure 2 can no longer explain Phenomenon Claim 3, nor Result 4, which is downstream of Phenomenon Claim 3. This is sufficient for eliminating the account. Figure 2 also shows a broken explanatory link between

Phenomenon Claim 2 and Result 3, which also suffices to eliminate the account. Now, it might be possible to modify the explanatory account by providing alternative explanations that replace the broken links. If so, how can we say that the account has been eliminated? There are two things to say in response. First, the explanatory-eliminative approach considers the revised explanatory account as a different account. Second, this account would have already been constructed and be subject to the elimination process. There is thus no reason to worry about the ability to revise an account after one or more of its explanatory links have been broken.

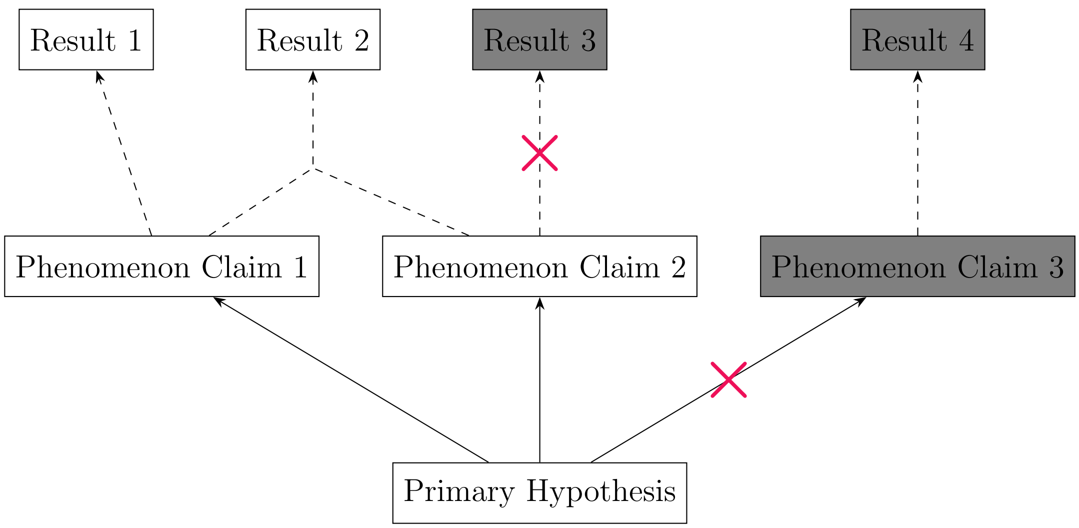

Figure 2: Two ways to eliminate a hypothetical explanatory account. A red 'x' depicts a broken explanatory link, and a shaded node represents a result or phenomena claim that the account can no longer explain.

I briefly address what it means to render an explanatory link implausible. In an ideal world, researchers would show that an explanatory connection does not hold with absolute certainty. In practice, science rarely allows for such certainty, so the eliminative step can involve some type of threshold. In certain contexts it may be possible to explicitly define probabilistic thresholds for when an explanatory link has been broken. In others, the arguments given for why an explanatory link does not hold may not express a precise threshold and instead proceed on a purely qualitative level. For example, the eliminative reasoning used by researchers in the Covid-19 case was largely qualitative in character, and the droplet-and-contact hypothesis was deemed unable to explain the large effect of ventilation on transmission risk largely through qualitative arguments. In the climate sensitivity case, Sherwood et al. (2020) based their analysis on qualitative arguments but also applied Bayesian techniques to quantify their confidence. The explanatory-eliminative approach is flexible in that it allows researchers to use quantitative or qualitative arguments to show that an explanatory link is implausible.

What do these arguments look like? The explanatory-eliminative approach does not specify the form these arguments need to take. Researchers are free to use any relevant evidence to inductively eliminate an explanatory link. However, we can say a few things at a general level. One way to eliminate an explanatory link is to show that one (or more) of its auxiliaries is false or implausible. An explanation that relies on false auxiliary information is a wrong explanation; this is the case even if the auxiliaries could be modified to provide a correct explanation. Another way to break an explanatory link is to show that a phenomenon claim itself is implausible. For example, in the Sherwood et al. (2020) analysis of climate sensitivity, a phenomenon claim that was part of explanatory accounts for ECS values below 2°C was that the net feedback of clouds on global warming is negative. This claim was required to explain various results under low sensitivity scenarios. Sherwood et al. (2020) argued that this claim was implausible. A third way to break an explanatory link is to show that a phenomenon claim (along with a set of auxiliaries) cannot actually explain the result, even though the phenomenon claim and auxiliaries may all be true. A phenomenon claim may initially appear to explain a result, which can lead researchers to incorporate it in an explanatory account that makes it to the elimination phase. However, closer analysis may reveal that the result has nuanced features that cannot be explained by the phenomenon claim.

If many explanatory accounts pass basic adequacy criteria, the elimination stage potentially poses a long and challenging task to researchers. Attempting to refute every explanatory account individually could require a great deal of time and effort. However, there are several things researchers can do to make the elimination process more efficient. First, it is often possible to eliminate multiple explanatory accounts at the same time. If multiple explanatory accounts have an explanatory link in common, and that link can be refuted, each of those accounts will be eliminated. For example, climate science researchers ruled out ECS values below 2°C by eliminating one of the physical storylines shared by all explanatory accounts with ECS values below 2°C—namely, that the net cloud feedback is negative. Because some explanatory links will often be found across multiple explanatory accounts, researchers can speed up the elimination process by prioritizing these links. When there is a small number of accounts remaining, researchers can focus their attention on explanatory links that differ between accounts. For example, in the Covid-19 case, I noted that the aerosol account and the droplet-and-contact account gave competing explanations of apparent long-range transmission reported by epidemiological investigations. Researchers were able to rule out the droplet-and-contact account in part by assessing these explanations. Thus, the elimination process need not proceed as an exhaustive assessment of each explanatory account.

At the end of the eliminative process, there should be one or more explanatory accounts that remain. The explanatory-eliminative approach says that these are the only accounts that are plausible in light of the evidence. Since there can be multiple explanatory accounts based on the same primary hypothesis, it is possible that multiple explanatory accounts remain but researchers conclude that there is only one plausible primary hypothesis. The explanatory-eliminative approach does not say anything about how to assess the relative strength of the remaining accounts. If two explanatory accounts remain after the eliminative process, there could be good reasons for placing more confidence in one than the other, but this is beyond the amalgamation process sketched here. However, one useful feature of the explanatory-eliminative approach is that it can be paired with Bayesian or other techniques for quantifying evidential support, like in the Sherwood et al. (2020) analysis of climate sensitivity.

The explanatory-eliminative approach attempts to overcome the major challenges of amalgamating multimodal evidence. The approach satisfies the four desiderata mentioned at the end of Section 2 in the following ways. First, the approach can work with qualitative evidence because there is nothing about the amalgamation process that requires researchers to input quantitative results, as in meta-analysis, for example. Second, the approach has a principled way of handling discordant results, which is to seek explanations of discordance and incorporate them into the explanatory accounts being assessed. We saw an example of this in how researchers reconciled studies that failed to detect viable SARS-CoV-2 from aerosols with other evidence that supported the aerosol hypothesis. Third, the approach is able to incorporate information about evidential context, because each explanatory link is associated with a set of auxiliaries needed for explaining a result. Finally, the approach does not assume that various lines of evidence are independent, and explanatory accounts highlight potential dependencies between lines of evidence.

I now show how the explanatory-eliminative approach can be applied to a real example. Based on my discussion in Section 3.1, I have constructed diagrams for the two competing explanatory accounts in the Covid-19 transmission case. These accounts represent the evidence base as it was in 2020 and early 2021. Figure 3 depicts the overall structure of the aerosol account, incorporating key auxiliaries. Figure 4 does the same for the droplet-and-contact account. I have included key results that each account needed to explain. For simplicity, I have grouped them into four categories, which represent sets of individual study results: results indicating specific outbreak patterns, estimates of Covid-19's basic reproduction number R_0, results indicating efficacy for surgical and cotton masks in limiting transmission, and negative findings in several studies that attempted to detect viable SARS-CoV-2 from aerosols. For each of these, the aerosol account and the droplet-and-contact account provide differing explanations, as they rely on differing phenomena claims. I would like to highlight two important differences. First, whereas the aerosol account claims that ventilation will have a strong effect on transmission (PC2 in Figure 3), the droplet-and-contact account claims that ventilation will have a minimal effect (PC7 in Figure 4). Second, in order to explain apparent long-range transmission (as reported by some outbreak investigations), the droplet-and-contact account needed to postulate unobserved contact or proximity (A3 in Figure 4).

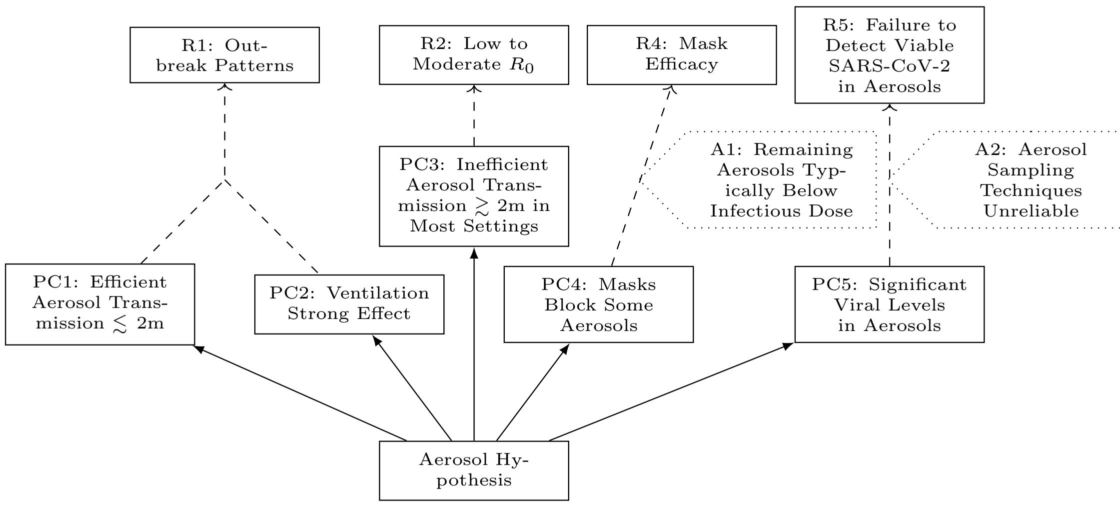

Figure 3: The overall structure of the explanatory account for the aerosol hypothesis for Covid-19 transmission

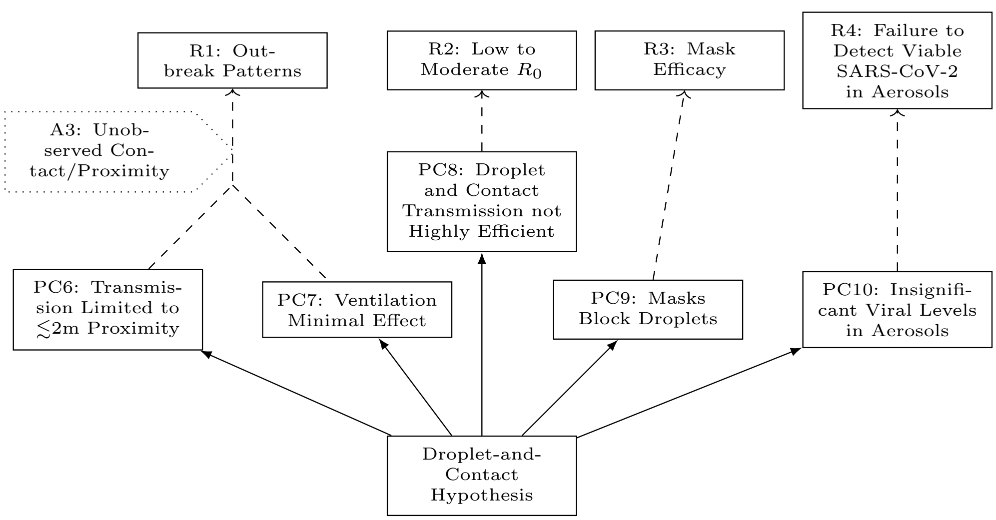

Figure 4: The overall structure of the explanatory account for the droplet-and-contact hypothesis for Covid-19 transmission

Next, Figure 5 illustrates how the droplet-and-contact account was eliminated. Two important phenomena claims were shown to be implausible: that transmission can only occur within approximately two meters of proximity (PC6), and that ventilation has a minimal effect on transmission (PC7). As I discussed in Section 3.1, this occurred because of a build up of studies indicating long-range transmission and studies indicating a strong effect of ventilation on transmission risk. As a consequence, the droplet-and-contact account could no longer explain key patterns about Covid-19 transmission. In particular, it could not explain a large difference in transmission risk between indoor and outdoor settings, a strong effect of ventilation in indoor settings, and long-range transmission reported in rigorously conducted outbreak investigations. Later, another phenomena claim was also shown to be implausible: that there are insignificant levels of viable SARS-CoV-2 in aerosols (PC10). This means that the droplet-and-contact account could also no longer explain the failure of early studies to detect viable SARS-CoV-2 in aerosols (R4). However, I have not included this in Figure 5 since it occurred after the droplet-and-contact account had already been effectively eliminated.

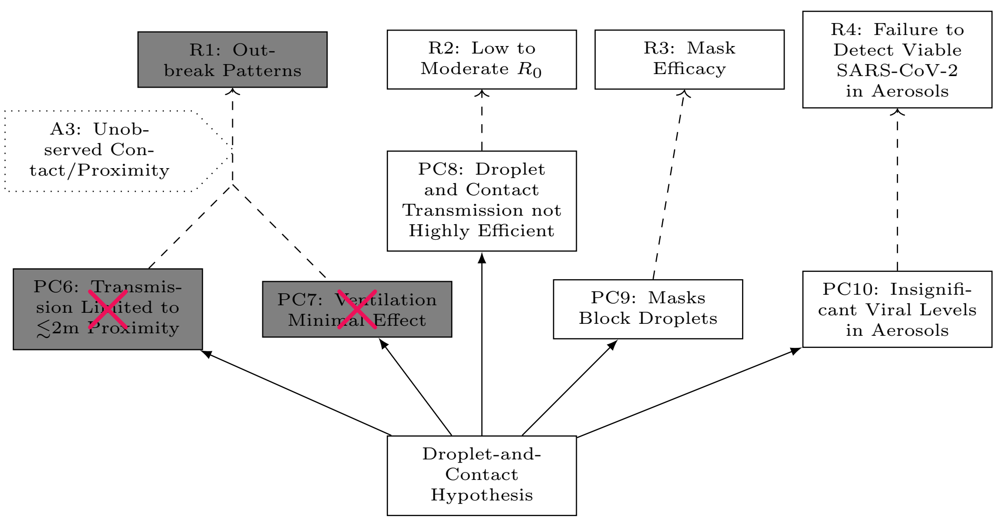

Figure 5: Elimination of the explanatory account for the droplet-and-contact hypothesis for Covid-19 transmission

4.4 Discussion

Before addressing some potential concerns with the explanatory-eliminative approach, I will discuss how it relates to some recent proposals in philosophy of science. The first proposal is Douglas's (2012) explanatory approach for amalgamating complex evidence in policy-making contexts, which I have already touched on previously. In particular, I mentioned that I adopted Douglas's concept of *explanatory accounts* but expanded and deviated from her proposal in a few important ways. First, whereas Douglas's notion of explanatory accounts is fairly minimal—sets of potential explanations—I fleshed out the structure of explanatory accounts in terms of hierarchical relationships between physical hypotheses, phenomena claims, and empirical results. This structure allows us to more clearly apply explanatory evidence amalgamation methods, whether they be *explanatory*$_1$ or *explanatory*$_2$ approaches. Second, I emphasized that Douglas's account was an *explanatory*$_1$ approach because it is a version of IBE, whereas my explanatory-eliminative approach is an *explanatory*$_2$ approach. The role of explanations in an *explanatory*$_2$ approach—constructing and considering explanations for why the full body of evidence appears as it does in order to arrive at a conclusion—has been seriously underconsidered in philosophy of science literature. My account further diverges from Douglas's approach through its reliance on eliminative reasoning to evaluate explanatory accounts.

The idea of deploying explanatory considerations in evidence amalgamation also emerges in different ways in Schupbach's (2018) account of robustness analysis and in Reiss's (2015) pragmatist theory of evidence. Schupbach introduced a novel account of what makes certain pieces of evidence valuable in robustness analysis. Schupbach's central claim is that pieces of evidence are valuable in supporting a hypothesis H within a robustness analysis if they are explanatorily discriminating between H and competing hypotheses. For Schupbach (2018), a result explanatorily discriminates between H and an alternative hypothesis H' if H would "explain well" why we detected the result and H' would "explain well" a failure to detect the result (288). Schupbach's account leaves ambiguous what it means to explain something well. It is thus not clear whether we should interpret this along the lines of IBE (where explaining something well would involve exhibiting explanatory virtues) or in some other sense. Schupbach's other work on IBE suggests the former interpretation (e.g., Schupbach 2017). In any case, Schupbach's account gives explanatory reasoning a central role and also emphasizes the evidential value on ruling out competing hypotheses. In these two aspects, my explanatory-eliminative approach is similar in spirit, though it differs in three major ways. First, rather than assessing how individual pieces of evidence discriminate between hypotheses, it focuses on constructing and evaluating explanatory accounts of the full evidence corpus. Second, it explicitly incorporates auxiliary claims and contextual information about data production processes in constructing explanations, aspects that are not present in Schupbach's account. Finally, my approach overcomes two major limitations of robustness analysis faces when attempting to amalgamate multimodal evidence: the inability to handle discordant evidence and the failure to incorporate indirect evidence. The explanatory-eliminative approach can handle both through its systematic construction and evaluation of explanatory accounts.

The explanatory-eliminative approach embodies a few lessons from Reiss's (2015) pragmatist theory of evidence. Reiss frames his account as an alternative to what he calls the "experimental paradigm" in biomedical and social sciences, which views RCTs as intrinsically reliable and downgrades other methodologies based on how much they deviate from RCTs (341). In contrast, the pragmatist theory aims to explain why diverse sources of evidence can be valuable in assessing causal hypotheses. Reiss provides a theory of evidential support and warrant rather than a method for amalgamating evidence. Nevertheless, this theory has several valuable lessons for multimodal evidence amalgamation. First, it highlights the importance of evidential context by emphasizing how "background knowledge about how the world works" affects what data patterns we are entitled to expect under the truth of a hypothesis (349). Second, Reiss's theory emphasizes eliminative reasoning by giving a prominent confirmatory role to the elimination of alternative hypotheses. Third, Reiss emphasizes the value of indirect evidence in assessing hypotheses. The last two features are intertwined in Reiss's account, as he defines indirect evidence as evidence that helps eliminate alternative hypotheses.

The explanatory-eliminative embodies these three lessons but differs in the details. While Reiss and I both emphasize evidential context, my notion of evidential context includes information about data production processes in addition to background assumptions about the world. As I have tried to show using the example of Covid-19 studies that sampled aerosols, information about data production processes can be crucial to understanding the evidential significance of results. Reiss does not address this kind of information in his account, but in principle it could be incorporated. With respect to eliminative reasoning, our accounts differ on the objects of elimination: hypotheses versus explanatory accounts. Additionally, because there can be multiple explanatory accounts that use the same hypothesis, to eliminate a hypothesis, researchers would need to eliminate all of those explanatory accounts. Finally, Reiss's notion of indirect evidence is significantly different than mine. Indirect evidence for Reiss is evidence that can be used to eliminate alternative hypotheses. In contrast, my notion of indirectness concerns the "distance" between the evidence and a hypothesis—evidence is indirect if it requires a long chain of reasoning using auxiliary claims to be related to a hypothesis. What Reiss considers indirect evidence would often be part of the direct evidence that all explanatory accounts need to explain under the explanatory-eliminative account, while evidence about measurement techniques and data production processes that I consider indirect plays no explicit role in his framework.

In addition to these subtle differences, there are several more prominent ones. While the explanatory-eliminative approach provides a systematic way to handle discordant results, Reiss's framework does not address how to handle discordant results. It is not clear, for example, what we should do if one result provides direct support to a hypothesis and another appears incompatible with that hypothesis. This is a serious limitation to Reiss's account because discordant evidence is common in the biomedical and social sciences, where

Reiss's account is supposed to be most applicable. A more fundamental difference is that whereas Reiss's account provides a theory of evidential support and warrant, the explanatory-eliminative approach offers a systematic method for amalgamating multimodal evidence. The method provides clear guidance for each stage: first in identifying the empirical results to be explained, then in systematically developing explanatory accounts that pass basic adequacy criteria, and finally in eliminating accounts. Reiss's account, while valuable as a theory of support and warrant, does not outline an amalgamation process. Multiple amalgamation approaches are consistent with his theoretical framework, such as the explanatory-eliminative approach and Douglas's approach, and researchers relying on his account would lack guidance on how to systematically develop and evaluate alternative explanations of the evidence.

What is perhaps most distinctive about the explanatory-eliminative approach is not found in any of the three accounts discussed above. It is that researchers assess hypotheses *in conjunction with* the auxiliaries and phenomena claims needed to explain the full set of empirical results. Hypotheses are not evaluated separately from the auxiliaries and phenomena claims they require to explain the evidence base. In a sense, the explanatory-eliminative approach internalizes the central claim of confirmational holism: that individual statements cannot be (dis)confirmed in isolation from a web of background claims. This holistic treatment of hypotheses is not found in existing explanatory proposals nor in accounts of eliminative reasoning. Forber (2011) argues that holism undermines eliminative inference because we "have no justifiable reason for eliminating the hypothesis instead of the auxiliaries we must use to generate predictions" (189). While philosophers of science have offered various solutions to this problem (e.g., Kitcher 1993; Sober 1999), the explanatory-eliminative account provides a novel strategy by treating the entire explanatory account as the fundamental unit of scientific assessment. Whether this strategy can truly overcome skeptical worries based on holism must await future work, but it represents a promising and distinctive approach.

Taking a step back, we can see that the explanatory-eliminative approach is part of a broader turn toward explanatory frameworks for evidence amalgamation in philosophy of science. Schupbach, Douglas, and Reiss all emphasize explanatory factors in evidence assessment, though in different ways. Schupbach (2018) focuses on the significance of evidence that can discriminate between competing explanations, Douglas (2012) develops an approach that assesses competing explanatory accounts of evidence, and Reiss (2015) provides a theory of evidential support and warrant that centers the elimination of alternative hypotheses. The explanatory-eliminative approach builds on these insights while attempting to provide a general and systematic approach for amalgamating real-world evidence.

Recently, Fuller, Chin-Yee, and Upshur (2024) have proposed an "argument framework" that represents another flexible approach for real-world evidence amalgamation. Their framework aims to overcome limitations of the traditional measurement-focused approach to evidence in medicine, especially the inability of meta-analysis to incorporate diverse types of evidence. Like the explanatory-eliminative approach, their framework allows diverse sources of evidence to be brought together through chains of reasoning, though they represent these connections through argument diagrams rather than explanatory accounts. The argument framework and the explanatory-eliminative approach share an emphasis on making explicit the reasoning that connects diverse pieces of evidence to hypotheses. Both can also incorporate highly indirect evidence. While there are some deep structural differences between the two approaches, both attempt to overcome major weaknesses of existing amalgamation methods.

I close this section by addressing how the explanatory-eliminative approach relates to theories of evidential support. The approach is not an account of evidential support or warrant, but should be understood as a methodological strategy for amalgamating evidence, similar to methodological triangulation and meta-analysis. Methodological strategies can be assessed in part by how well the conclusions they reach accord with theories of evidential support. In principle, it should be possible to compare various methodological strategies for amalgamating evidence in a Bayesian framework, or in Reiss's (2015) pragmatist theory of evidence, for instance. This is a complex task that future work could explore. However, there are a few reasons for thinking that the explanatory-eliminative approach would fare well, at least when assessed in a Bayesian framework. For one, eliminative reasoning has often been thought to be well-accommodated by Bayesianism (Earman 1992; Hawthorne 1993; Vineberg 1996). Moreover, as I described earlier, robustness analysis and methodological triangulation can dilute the evidential significance of results in seeking consensus among methods, but the explanatory-eliminative approach avoids this issue because it does not rely on a consensus-seeking process. Finally, if there are multiple hypotheses that cannot be ruled out by the evidence, IBE and methodological triangulation often tell us to accept just one hypothesis—the one that would best explain the evidence, or the one supported by the greatest number of methods—whereas the explanatory-eliminative approach does not force us to accept one hypothesis.

4.5 Worries and Objections

I now address several concerns with the explanatory-eliminative approach. There are two prominent objections to any inferential strategy centered on elimination: the problem of unconceived alternatives, as I mentioned earlier, and the worry that eliminative strategies ignore or underutilize probative evidence. Because both apply to eliminative strategies generally, a complete response to them would be beyond this chapter's scope. However, I can address the specific features of the explanatory-eliminative approach that at least partially alleviate these worries.

The explanatory-eliminative approach is designed to help researchers reveal previously unconceived possibilities. Features in the construction phase for explanatory accounts help ensure that researchers consider all explanatory accounts consistent with background knowledge. The construction of detailed explanatory accounts that include auxiliaries and phenomena claims, as opposed to merely listing potential hypotheses, helps researchers identify explanatory possibilities they might not otherwise consider. When researchers systematically work through how primary hypotheses connect to phenomena claims and empirical results through various auxiliaries, they often discover explanatory possibilities that were not initially apparent. For example, early in the Covid-19 pandemic, some health authorities deemed the aerosol transmission hypothesis very unlikely because of the basic reproduction number of Covid-19, which is substantially lower than other respiratory diseases known to transmit through aerosols, such as measles (Conly et al. 2020). These authorities appeared to assume that aerosol transmission would cause a high basic reproduction number (Jimenez et al. 2022). By systematically considering explanatory possibilities, researchers considering the aerosol hypothesis were able to offer an explanation for the apparent discrepancy. The realization that aerosol transmission does not require a high basic reproduction number has prompted a broader reconsideration by health authorities of transmission mechanisms for other respiratory illnesses such as influenza (World Health Organization 2024; Maxmen 2024). This example illustrates how systematically articulating explanatory accounts helps prevent researchers from overlooking explanatory possibilities.

Additionally, the explanatory-eliminative approach can help reveal when our explanatory resources are inadequate. When existing explanatory accounts fail to fully explain the evidence or require impossible auxiliaries to accommodate certain results, this indicates that researchers have to look for unconceived alternatives. In the explanatory-eliminative approach, if researchers are unable to construct any explanatory accounts that meet basic adequacy criteria and are consistent with background knowledge, then they need to stop the process and reconsider items that were taken to be background knowledge. Because the requirements on explanation and coherence are rather demanding, in situations where none of the hypotheses under consideration are correct, the explanatory-eliminative approach will be more likely to raise the need for reconsideration than most other amalgamation methods. The explanatory-eliminative approach will thus be more effective at mitigating the problem of unconceived alternatives than many other forms of inductive inference.

The worry that the explanatory-eliminative approach ignores or underutilizes probative evidence threatens to limit its usefulness when more than one explanatory account remains after the elimination stage. We can use the climate sensitivity case as an example. The Sherwood et al. (2020) study effectively ruled out scenarios in which ECS is below 2°C and above 4.5°C. However, the difference between 2°C and 4.5°C has very different implications for climate change and it would be very useful to know the likelihood of different possibilities within this range. The explanatory-eliminative approach does not attempt to provide this. This may seem like a serious problem because we often have access to probative evidence that is useful in assessing remaining hypotheses. If the explanatory-eliminative approach has nothing to offer here, it seems that we would be forced to withhold judgment about remaining hypotheses rather than making informed judgments about their likelihood or strength.

First, probative evidence is utilized in the explanatory-eliminative approach, because it forms part of the evidence corpus that each explanatory account must explain. This does not directly address the objection but highlights the fact that probative evidence is not ignored by the approach. This evidence can actually be useful in constraining explanatory possibilities because some hypotheses may not be able to provide a coherent explanation of all pieces of probative evidence, and thus will not make it past the elimination phase. Another preliminary point is that the explanatory-eliminative approach does not require withholding judgment about remaining hypotheses. It instead does not say what researchers should believe about remaining hypotheses beyond the conclusion that they are plausible in light of existing evidence.

This is a substantial limitation that may strike some as a serious problem. However, I argue that this limitation actually represents a strength of the approach. First, the explanatory-eliminative approach can be usefully paired with methods that harness probative evidence to assess the degree of empirical support for remaining explanatory accounts. I have already mentioned that the Sherwood et al. (2020) study used a Bayesian analysis to calculate a posterior probability density function for climate sensitivity values in conjunction with its "storyline" approach. There is nothing preventing researchers from applying Bayesian or other techniques to assess empirical support for remaining explanatory accounts.

However, the explanatory-eliminative approach also offers something valuable here. We can use climate sensitivity again as an example. Because Bayesian techniques are so flexible, there have been a large number of Bayesian analyses of ECS, often with major differences in their model structure and priors that had a large impact on results (Knutti and Hegerl 2008). Because of this flexibility, these analyses did little to usefully constrain ECS, which is partly why the IPCC did not rely on them in its Fifth Assessment Report to assess ECS (Collins et al. 2013). However, the explanatory-eliminative approach can help structure and constrain probabilistic analyses. By requiring researchers to develop comprehensive explanatory accounts with explicit phenomena claims and auxiliary assumptions, it creates a more rigorous framework for probabilistic analyses. As I have already mentioned, the Sherwood et al. (2020) study proved influential for the IPCC's Sixth Assessment Report in its assessment of ECS partly due to its novel methodology in the climate science context. The explanatory-eliminative approach thus helps constrain probabilistic techniques that assess the probative value of evidence, allowing for more reliable and informative probabilistic assessments.

Second, the explanatory-eliminative approach avoids overconfident conclusions, which is a major risk for many amalgamation methods. Methods that attempt to directly assess the strength of probative evidence for competing hypotheses can lead researchers to place more confidence in certain hypotheses than the evidence warrants. This tendency arises through several mechanisms. First, uncertainty about model structure can be difficult to account for properly. Techniques such as Bayesian Model Averaging attempt to account for this kind of uncertainty but have well-known shortcomings in doing so (Oelrich et al. 2020; Yao 2021). Second, the desire to extract positive support from evidence can lead researchers to overlook shared assumptions between lines of evidence. For example, many Bayesian estimates of ECS prior to Sherwood's study incorrectly assumed that lines of evidence were independent, producing overconfident assessments (Knutti et al. 2010). Third, and most generally, a desire to extract positive support from evidence can lead researchers to downplay or overlook alternative explanations. In the case of Covid-19, health authorities viewed the lower basic reproduction number compared to SARS-CoV-2 compared to known aerosol-spreading diseases as strong evidence for droplet spread, without adequate consideration of how this apparent evidential link could fail (Jimenez et al. 2022). By maintaining focus on the explicit explanatory relationships between hypotheses and results, the explanatory-eliminative approach helps prevent researchers from overlooking plausible alternative explanations of results.

I now consider another objection. One might think that in some contexts, it will be virtually impossible to eliminate explanatory accounts because any apparently disconfirming results could be "explained away" by postulating bias. The postulation of bias would serve as an auxiliary assumption allowing an explanatory account to explain a result that appears to conflict with its primary hypothesis and/or its phenomena claims. This might be particularly troublesome in the biomedical sciences, where it is often difficult to know whether study results are tainted by biases such as publication bias and confounding bias. If this objection is correct, the utility of the explanatory-eliminative approach will be very limited whenever the presence of bias is difficult to assess.

My response is two-fold: first, I will show how explanatory accounts that postulate bias to explain away apparently conflicting results can be eliminated; second, I will argue that when there is serious uncertainty about bias, other amalgamation approaches will not fare any better than the explanatory-eliminative approach in constraining conclusions. The first point to recognize is that auxiliary assumptions themselves are subject to elimination. Additionally, eliminating an auxiliary assumption suffices to break its associated explanatory link. As I noted in Section 4.3, one way to eliminate an explanatory link is to show that one (or more) of its auxiliaries is false or implausible. We now need to consider how auxiliaries that postulate bias could be shown to be false or implausible. There are several main ways: Researchers could try to show that (1) a result is too large or significant for the bias to possibly account for it; (2) the postulated bias is incompatible with specific features of the result; (3) the postulated bias is very unlikely in light of study's methodological features; (4)

the postulated bias is incompatible with other features of the evidence corpus. Additionally, when there are many studies that reach the same overall conclusion, explaining away study results would require the postulation of a systematic bias mechanism. In response, researchers could try to establish variants of (1)-(4) applicable to the collection of study results, as well as showing that (5) the studies use diverse methodologies that would be vulnerable to different biases.

Researchers may be able to quantitatively demonstrate that a result is too large or significant for the postulated bias to plausibly account for it. For example, sensitivity analyses can establish bounds on how much an unmeasured confounder would need to influence treatment assignment and outcome to nullify an observed effect (D'Agostino McGowan 2022). When these bounds exceed what is plausible given in light of domain-specific background knowledge, the bias explanation becomes implausible. Second, researchers can show that the postulated bias is incompatible with specific features of a result. Reiss (2015) points out that genetic accounts of the association between smoking and lung cancer, such as Fisher's (1958) constitutional hypothesis—which proposed that genetic predisposition confounds epidemiological studies by influencing both smoking behavior and cancer susceptibility independently—could not account for the large dose-response effect between smoking intensity and cancer risk observed in epidemiological studies. Third, researchers can argue that a postulated bias is very unlikely in light of study's methodological features, such as blinding, randomization, and pre-registration. For instance, if an explanatory account claims that observer expectancy effects biased the results of a clinical trial, this claim becomes implausible if the trial used triple-blinding procedures where participants, clinicians, and data analysts were all unaware of treatment assignments. Fourth, researchers can attempt to find patterns in the evidence corpus that are incompatible with the bias hypothesis. For example, if an explanatory account claims that publication bias explains positive findings about a drug's efficacy, this explanation becomes implausible if large, well-funded industry studies with pre-registered protocols show similar effect sizes to smaller academic studies. Finally, if an explanatory account posits a systematic bias across a group of studies, researchers can apply similar strategies as above or also argue that the studies employ methodologically diverse approaches that would be vulnerable to different biases, making a systematic effect implausible.

My second point is that when there is serious uncertainty about bias (such that the explanatory-eliminative approach will be unable to rule out some explanatory accounts), alternative amalgamation approaches will not reach more strongly constrained conclusions than the explanatory-eliminative approach. Consider the methodological triangulation strategy of Heesen, Bright, and Zucker (2019) I mentioned in Section 2.2. This strategy will arrive at the conclusion that a specific hypothesis (from a set of potential hypotheses) is correct even when there is substantial uncertainty about bias for each source of evidence. Some versions of IBE will also do this. These strategies seem to provide more informative and better constrained conclusions than the explanatory-eliminative approach. In endorsing a specific hypothesis under uncertainty, these approaches effectively neglect uncertainty about bias and run a substantial risk of endorsing a false conclusion. This is why sophisticated versions of IBE require that the best explanation is "sufficiently good" (Lipton 2004, 154), although there has been little agreement about how to establish such a threshold. Amalgamation methods that appear to yield more definitive conclusions under substantial uncertainty about bias typically do so by making stronger assumptions about the evidence than are warranted. The explanatory-eliminative approach, by contrast, makes these assumptions explicit as auxiliary claims within explanatory accounts and subjects them to critical scrutiny. When this scrutiny cannot eliminate all competing accounts due to uncertainty about bias, the approach acknowledges these epistemic limitations rather than artificially resolving them.

Finally, I address a different sort of objection: that the explanatory-eliminative approach is impractical for many scientific contexts because it is very resource intensive. Because approach requires researchers to develop detailed explanatory accounts for all plausible hypotheses, it can require a large amount of time and labor. For example, the Sherwood et al. (2020) assessment of climate sensitivity cited over 500 climate studies and took four years to complete (Forster et al. 2020). For complex scientific problems with many variables and potential mechanisms, the explanatory-eliminative could become unwieldy or infeasible. It is true that in many contexts the explanatory-eliminative approach will require more time and effort to implement than some other evidence amalgamation methods, such as expert elicitation and methodological triangulation. As I argue in Chapter 4, it can be reasonable for researchers to use less evidence and rely on less demanding methods when practical constraints require it. The explanatory-eliminative approach is best suited for cases where the stakes are high, resources are available, and careful assessment of complex evidence is needed.

Even in such challenging cases, though, several features of the approach help make it more tractable than it might initially appear. First, background knowledge substantially constrains the space of possibilities that need to be explored. Researchers do not need to consider every logically possible explanation, but only those consistent with well-established scientific understanding. Second, the approach's emphasis on explanatory structure can actually help make amalgamation more manageable by revealing patterns shared across multiple explanatory accounts. Explanatory structures will often appear in multiple explanatory accounts, which speeds up both the construction and elimination phases for explanatory accounts. Finally, the approach is collaborative and cumulative in ways that other methods like meta-analysis are not. Researchers can build on and modify existing explanatory accounts developed by others rather than starting from scratch. In contrast, researchers carrying out a meta-analysis on a topic effectively start from scratch. Each new meta-analysis typically requires re-extracting all data, recalculating effect sizes, and reapplying inclusion criteria across the entire evidence base.[3] Taken together, these three considerations suggest that

3. Here I am referring to meta-analyses conducted independently on the same topic. A few organizations such as the Cochrane collaboration periodically update their meta-analyses by incorporating newly published results, but these would not be examples of new meta-analyses.

while the explanatory-eliminative approach requires serious effort, it may be considerably more manageable in practice than the objection suggests.

5 Conclusion

Existing evidence amalgamation methods are poorly suited for assessing complex, multimodal bodies of evidence. This poses serious challenges to scientists and policymakers. The explanatory-eliminative approach I have developed in this chapter overcomes major weaknesses of existing methods and offers a systematic strategy to amalgamate multimodal evidence. It draws inspiration from the reasoning used by Covid-19 researchers to identify the primary mode of Covid-19 transmission and by climate scientists to reduce uncertainty about equilibrium climate sensitivity. These cases illustrate how scientists can effectively amalgamate multimodal evidence using a combination of explanatory and eliminative reasoning.

In the explanatory-eliminative approach, these two components work in tandem. Researchers systematically construct explanatory accounts of the evidence corpus that specify how hypotheses relate to empirical results through chains of reasoning that incorporate auxiliaries and phenomena claims. This explicit charting of explanatory relationships allows for the incorporation of indirect evidence and contextual information about data production processes. In the eliminative phase, researchers attempt to eliminate explanatory accounts by identifying implausible explanatory links. Unlike most accounts of eliminative reasoning in science, the objects of elimination are explanatory accounts rather than hypotheses themselves. This approach represents an explanatory$_2$ rather than an explanatory$_1$ strategy—it constructs and considers explanations of why the full body of evidence appears as it does in order to arrive at a conclusion. In making this distinction I have tried to highlight an underappreciated type of explanatory reasoning in science, one that does not rely on explanatory virtues as criteria for inference as in IBE.

The explanatory-eliminative approach has several advantages over existing amalgamation methods. Unlike meta-analysis, it can incorporate qualitative results and evidence about different phenomena. Unlike robustness analysis, it can handle discordant evidence and incorporate highly indirect evidence. Unlike some Bayesian approaches, it does not require precise probability assignments or strong independence assumptions between lines of evidence. The explanatory-eliminative approach also avoids diluting the evidential significance of results by consensus-seeking processes such as those found in methodological triangulation.

Future work on the explanatory-eliminative approach could pursue several promising directions. First, it may be possible to develop formal criteria to assess the plausibility of explanatory links, which would enhance the rigor of the elimination phase. Second, exploring in more depth how this approach can be combined with Bayesian techniques could help researchers quantify confidence in non-eliminated explanatory accounts, following the example of Sherwood et al. (2020) in their analysis of climate sensitivity. Third, considering how the approach might be adapted for contexts with severe time or resource constraints could make it more practical for urgent decision-making scenarios. Addressing these questions could help further refine the explanatory-eliminative approach and improve its usefulness in both scientific and policy-making contexts.

References

Annan, J. D., and J. C. Hargreaves. 2011. “On the Generation and Interpretation of Probabilistic Estimates of Climate Sensitivity.” *Climatic Change* 104 (3): 423–436.

Azimi, Parham, Zahra Keshavarz, Jose Guillermo Cedeno Laurent, Brent Stephens, and Joseph G. Allen. 2021. “Mechanistic Transmission Modeling of COVID-19 on the Diamond Princess Cruise Ship Demonstrates the Importance of Aerosol Transmission.” *Proceedings of the National Academy of Sciences* 118 (8): e2015482118.

Berlin, Jesse A., and Robert M. Golub. 2014. “Meta-Analysis As Evidence: Building a Better Pyramid.” *JAMA* 312 (6): 603–606.

Bhagat, Rajesh K., M. S. Davies Wykes, Stuart B. Dalziel, and P. F. Linden. 2020. “Effects of Ventilation On the Indoor Spread of COVID-19.” *Journal of Fluid Mechanics* 903: F1.

Bird, Alexander. 2005. “Abductive Knowledge and Holmesian Inference.” In *Oxford Studies in Epistemology,* edited by Tamar Szabo Gendler and John Hawthorne, 1–31. Oxford University Press.

———. 2010. “Eliminative Abduction: Examples from Medicine.” *Studies in History and Philosophy of Science Part A* 41 (4): 345–352.

CDC. “Scientific Brief: SARS-CoV-2 Transmission.” Centers for Disease Control / Prevention. May 7, 2021. https://www.cdc.gov/coronavirus/2019-ncov/science/science-briefs/sars-cov-2-transmission.html.

Collins, Matthew, Reto Knutti, Julie Arblaster, Jean-Louis Dufresne, Thierry Fichefet, Pierre Friedlingstein, Xuejie Gao, et al. 2013. "Long-Term Climate Change: Projections, Commitments and Irreversibility." In *Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel On Climate Change,* 1029–1136. Cambridge University Press.

Conly, John, W. H. Seto, Didier Pittet, Alison Holmes, May Chu, Paul R. Hunter, John Conly, et al. 2020. "Use of Medical Face Masks Versus Particulate Respirators As a Component of Personal Protective Equipment for Health Care Workers in the Context of the COVID-19 Pandemic." *Antimicrobial Resistance & Infection Control* 9 (1): 126.

D'Agostino McGowan, Lucy. 2022. "Sensitivity Analyses for Unmeasured Confounders." *Current Epidemiology Reports* 9 (4): 361–375.

Deeks, Jonathan J., Julian P. T. Higgins, and Douglas G. Altman. 2008. "Chapter 9: Analyzing Data and Undertaking Meta-Analyses." In *Cochrane Handbook for Systematic Reviews of Interventions,* edited by Julian P. T. Higgins and Sally Green, 243–296. John Wiley & Sons.

Douglas, Heather. 2012. "Weighing Complex Evidence in a Democratic Society." *Kennedy Institute of Ethics Journal* 22 (2): 139–162.

Drossinos, Yannis, Thomas P. Weber, and Nikolaos I. Stilianakis. 2021. "Droplets and Aerosols: An Artificial Dichotomy in Respiratory Virus Transmission." *Health Science Reports* 4 (2): e275.

Earman, John. 1992. *Bayes or Bust?: A Critical Examination of Bayesian Confirmation Theory.* MIT Press.

Fisher, Ronald A. 1958. "Cancer and Smoking." *Nature* 182 (4635): 596–596.

Forber, Patrick. 2011. "Reconceiving Eliminative Inference." *Philosophy of Science* 78 (2): 185–208.

Forster, P., T. Storelvmo, K. Armour, W. Collins, J.-L. Dufresne, D. Frame, D. J. Lunt, et al. 2021. "The Earth's Energy Budget, Climate Feedbacks and Climate Sensitivity." In *Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel On Climate Change,* 923–1054. The Intergovernmental Panel on Climate Change.

Forster, Piers, Zeke Hausfather, Gabi Hegerl, Steven Sherwood, and Kyle Armour. "Guest Post: Why Low-End 'Climate Sensitivity' Can Now Be Ruled Out." Carbon Brief. 2020. https://www.carbonbrief.org/guest-post-why-low-end-climate-sensitivity-can-now-be-ruled-out/.

Frame, D. J., B. B. B. Booth, J. A. Kettleborough, D. A. Stainforth, J. M. Gregory, M. Collins, and M. R. Allen. 2005. "Constraining Climate Forecasts: The Role of Prior Assumptions." *Geophysical Research Letters* 32 (9).

Fuller, Jonathan, Benjamin Chin-Yee, and Ross E. G. Upshur. 2024. "The Argument Framework Is a Flexible Approach to Evidence in Healthcare." *Nature Medicine* 30 (8): 2113–2116.

Greenhalgh, Trisha, Jose L. Jimenez, Kimberly A. Prather, Zeynep Tufekci, David Fisman, and Robert Schooley. 2021. "Ten Scientific Reasons in Support of Airborne Transmission of SARS-CoV-2." *The Lancet* 397 (10285): 1603–1605.

Hawthorne, James. 1993. "Bayesian Induction Is Eliminative Induction." *Philosophical Topics* 21 (1): 99–138.

Heesen, Remco, Liam Kofi Bright, and Andrew Zucker. 2019. "Vindicating Methodological Triangulation." *Synthese* 196 (8): 3067–3081.

Hey, Spencer Phillips. 2015. "Robust and Discordant Evidence: Methodological Lessons from Clinical Research." *Philosophy of Science* 82 (1): 55–75.

Imrey, Peter B. 2020. "Limitations of Meta-Analyses of Studies With High Heterogeneity." *JAMA Network Open* 3 (1): e1919325.

Jiang, Guanyu, Can Wang, Lu Song, Xing Wang, Yangyang Zhou, Chunnan Fei, and He Liu. 2020. "Aerosol Transmission, an Indispensable Route of COVID-19 Spread: Case Study of a Department-Store Cluster." *Frontiers of Environmental Science & Engineering* 15 (3): 46.

Jimenez, Jose L., Linsey C. Marr, Katherine Randall, Edward Thomas Ewing, Zeynep Tufekci, Trish Greenhalgh, Raymond Tellier, et al. 2022. "What Were the Historical Reasons for the Resistance to Recognizing Airborne Transmission During the COVID-19 Pandemic?" *Indoor Air* 32 (8): e13070.

Kitcher, Philip. 1993. *The Advancement of Science: Science Without Legend, Objectivity Without Illusions.* Oxford University Press.

Knutti, Reto. 2008. "Should We Believe Model Predictions of Future Climate Change?" *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* 366 (1885): 4647–4664.

Knutti, Reto, Reinhard Furrer, Claudia Tebaldi, Jan Cermak, and Gerald A. Meehl. 2010. "Challenges in Combining Projections from Multiple Climate Models." *Journal of Climate* 23 (10): 2739–2758.

Knutti, Reto, and Gabriele C. Hegerl. 2008. "The Equilibrium Sensitivity of the Earth's Temperature to Radiation Changes." *Nature Geoscience* 1 (11): 735–743.

Kuorikoski, Jaakko, and Caterina Marchionni. 2016. "Evidential Diversity and the Triangulation of Phenomena." *Philosophy of Science* 83 (2): 227–247.

Lipton, Peter. 2004. *Inference to the Best Explanation.* Routledge.

Maxmen, Amy. "WHO Overturns Dogma on Airborne Disease Spread. The CDC Might Not Act on It." *KFF Health News.* May 1, 2024. https://kffhealthnews.org/news/article/airborne-disease-transmission-who-cdc-new-evidence/.

Morawska, Lidia, and Donald K. Milton. 2020. "It Is Time to Address Airborne Transmission of Coronavirus Disease 2019 (COVID-19)." *Clinical Infectious Diseases* 71 (9): 2311–2313.

Murad, Mohammad Hassan, Noor Asi, Mouaz Alsawas, and Fares Alahdab. 2016. "New Evidence Pyramid." *BMJ Evidence-Based Medicine* 21 (4): 125–127.

Norton, John D. 1995. "Eliminative Induction as a Method of Discovery: How Einstein Discovered General Relativity." In *The Creation of Ideas in Physics: Studies for a Methodology of Theory Construction,* edited by Jarrett Leplin, 29–69. Dordrecht: Springer Netherlands.

Oelrich, Oscar, Shutong Ding, Måns Magnusson, Aki Vehtari, and Mattias Villani. 2020. *When Are Bayesian Model Probabilities Overconfident?* arXiv: 2003.04026 `[math.ST]`.

Pan, M., J. A. Lednicky, and C.-Y. Wu. 2019. "Collection, Particle Sizing and Detection of Airborne Viruses." *Journal of Applied Microbiology* 127 (6): 1596–1611.

Parker, Wendy. 2011. "When Climate Models Agree: The Significance of Robust Model Predictions." *Philosophy of Science* 78 (4): 579–600.

Randall, K., E. T. Ewing, L. C. Marr, J. L. Jimenez, and L. Bourouiba. 2021. "How Did We Get Here: What Are Droplets and Aerosols and How Far Do They Go? A Historical Perspective on the Transmission of Respiratory Infectious Diseases." *Interface Focus* 11 (6): 20210049.

Reiss, Julian. 2015. "A Pragmatist Theory of Evidence." *Philosophy of Science* 82 (3): 341–362.

Saatsi, Juha. 2017. "Explanation and Explanationism in Science and Metaphysics." In *Metaphysics and the Philosophy of Science: New Essays,* edited by Matthew Slater and Zanja Yudell, 163–192. Oxford University Press.

Schupbach, Jonah N. 2017. "Inference to the Best Explanation, Cleaned Up and Made Respectable." In *Best Explanations: New Essays on Inference to the Best Explanation,* edited by Kevin McCain and Ted Poston, 39–61. Oxford University Press.

———. 2018. "Robustness Analysis as Explanatory Reasoning." *The British Journal for the Philosophy of Science* 69 (1): 275–300.

Shen, Ye, Changwei Li, Hongjun Dong, Zhen Wang, Leonardo Martinez, Zhou Sun, Andreas Handel, et al. 2020. "Community Outbreak Investigation of SARS-CoV-2 Transmission Among Bus Riders in Eastern China." *JAMA Internal Medicine* 180 (12): 1665–1671.

Sherwood, S., M. J. Webb, J. D. Annan, K. C. Armour, P. M. Forster, J. C. Hargreaves, G. Hegerl, et al. 2020. "An Assessment of Earth's Climate Sensitivity Using Multiple Lines of Evidence." *Reviews of Geophysics* n/a (n/a): e2019RG000678.

Sherwood, Steven C., and Chris E. Forest. 2024. "Opinion: Can Uncertainty in Climate Sensitivity Be Narrowed Further?" *Atmospheric Chemistry and Physics* 24 (4): 2679–2686.

Sober, Elliott. 1999. "Testability." *Proceedings and Addresses of the American Philosophical Association* 73 (2): 47–76. JSTOR: 3131087.

Stanford, P. Kyle. 2006. *Exceeding Our Grasp: Science, History, and the Problem of Unconceived Alternatives.* Oxford, New York: Oxford University Press.

Stegenga, Jacob. 2009. "Robustness, Discordance, and Relevance." *Philosophy of Science* 76 (5): 650–661.

Stegenga, Jacob. 2012. "Rerum Concordia Discors: Robustness and Discordant Multimodal Evidence." In *Characterizing the Robustness of Science: After the Practice Turn in Philosophy of Science,* edited by Léna Soler, Emiliano Trizio, Thomas Nickles, and William Wimsatt, 207–226. Dordrecht: Springer Netherlands.

Stevens, Bjorn, Steven C. Sherwood, Sandrine Bony, and Mark J. Webb. 2016. "Prospects for Narrowing Bounds on Earth's Equilibrium Climate Sensitivity." *Earth's Future* 4 (11): 512–522.

Tebaldi, Claudia, and Reto Knutti. 2007. "The Use of the Multi-Model Ensemble in Probabilistic Climate Projections." *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* 365 (1857): 2053–2075.

Tellier, Raymond. 2009. "Aerosol Transmission of Influenza A Virus: A Review of New Studies." *Journal of the Royal Society Interface* 6 (suppl_6): S783–S790.

———. 2022. "COVID-19: The Case for Aerosol Transmission." *Interface Focus* 12 (2): 20210072.

Vineberg, Susan. 1996. "Eliminative Induction and Bayesian Confirmation Theory." *Canadian Journal of Philosophy* 26 (2): 257–266.

Wang, Chia C., Kimberly A. Prather, Josué Sznitman, Jose L. Jimenez, Seema S. Lakdawala, Zeynep Tufekci, and Linsey C. Marr. 2021. "Airborne Transmission of Respiratory Viruses." *Science* 373 (6558): eabd9149.

WHO. "Modes of Transmission of Virus Causing COVID-19: Implications for IPC Precaution Recommendations." World Health Organization. March 29, 2020. https://www.who.int/news-room/commentaries/detail/modes-of-transmission-of-virus-causing-covid-19-implications-for-ipc-precaution-recommendations.

WHO. "Transmission of SARS-CoV-2: Implications for Infection Prevention Precautions." World Health Organization. July 9, 2020. https://www.who.int/news-room/commentaries/detail/transmission-of-sars-cov-2-implications-for-infection-prevention-precautions.

———. "Coronavirus Disease (COVID-19): How Is It Transmitted?" World Health Organization. December 23, 2021. https://www.who.int/news-room/questions-and-answers/item/coronavirus-disease-covid-19-how-is-it-transmitted.

Woodward, James. 2000. "Data, Phenomena, and Reliability." *Philosophy of Science* 67: S163–S179.

World Health Organization. 2024. *Global Technical Consultation Report on Proposed Terminology for Pathogens That Transmit Through the Air.* Geneva: World Health Organization.

Yao, Yuling. 2021. *Bayesian Aggregation,* arXiv:1912.11218. arXiv: 1912.11218 [stat].