Philosophy of Medicine

Original Research: Philosophical Perspectives on Covid-19

Inconvenient Truth and Inductive Risk in Covid-19 Science

Eli I. Lichtenstein[1]

1 School of Philosophy, Psychology, and Language Sciences, University of Edinburgh, Scotland, UK.
Email: eli.lichtenstein@ed.ac.uk

Abstract

To clarify the proper role of values in science, focusing on controversial expert responses to Covid-19, this article examines the status of (in)convenient hypotheses. Polarizing cases like health experts downplaying mask efficacy to save resources for healthcare workers, or scientists dismissing "accidental lab leak" hypotheses in view of potential xenophobia, plausibly involve modifying evidential standards for (in)convenient claims. Societies could accept that scientists handle (in)convenient claims just like nonscientists, and give experts less political power. Or societies could hold scientists to a higher bar, by expecting them not to modify evidential standards to avoid costs only incidentally tied to error.

1. Introduction

The Covid-19 pandemic has raised serious concerns about the politicization of public health and biomedical science. Commentators have, for example, sounded the alarm over how heads of state—from Donald Trump to Emmanuel Macron—"hyped" drugs like hydroxychloroquine (Intemann 2022), and over survey data showing that "Republicans tend to underestimate Covid risks—and Democrats tend to exaggerate them" (Leonhardt 2021). However, this danger extends not just to politicians and the general public, but also to scientists and science communicators. Here especially, political influence may seem to directly undercut the epistemic aims of science.

At the same time, calls to depoliticize Covid-19 science belie a wider wave of modern work that has cast doubt on traditional ideals of science as value-free. Critics have argued that political and other social values affect scientific inquiry in ways that are all too often ignored or downplayed.[1] So, then, if "science has always been politicized" (Gauchat 2012, 168),[2] maybe we would be naive to see responses to Covid-19 as atypical. But as a normative

[1] For example, Longino (1995) argues that traditional standards of theoretical virtue reinforce inegalitarian gender relations.
[2] Compare Redding 2013; Thorp 2020.

ISSN 2692-3963 | Date Received 21-08-2022 | Date Accepted 01-12-2022 | Date First Published 14-12-2022
Philosophy of Medicine | DOI 10.5195/pom.2022.132 | Volume 3 | Issue 1 | pp.1–25

matter, too, critics have argued that social values should inform more stages of science than older ideals allowed.[3]

In view of these insights into the positive role of values in science and worries about bias, it is crucial to clarify where and how, exactly, to divide legitimate and illegitimate uses of non-epistemic values in science. Despite recent progress on this front (see, for example, Wilholt 2013; Intemann 2015; Winsberg, Oreskes, and Lloyd 2020; John 2021; Boulicault and Schroeder 2021; Koskinen and Rolin 2022; Holman and Wilholt 2022; ChoGlueck 2022; Bueter 2022),[4] Covid-19 offers new perspective.

Below, in this spirit, I aim to refine the influential argument from inductive risk (AIR), which concludes that social values can appropriately inform scientists' evidential standards. The core of the AIR is an intuitive idea that Richard Rudner articulated in 1953: "How sure we need to be before we accept a hypothesis will depend on how serious a mistake would be" (1953, 2). In turn, non-epistemic value judgments are often needed to assess the cost of scientific mistakes, such as accepting false hypotheses (*false positives* or *Type I error*) and rejecting true hypotheses (*false negatives* or *Type II error*). For example, as Rudner observed, it is morally worse for a pharmaceutical company to falsely claim that "Our drug has no lethal side-effects" than it is for a belt company to falsely claim that "Our buckles will not break." So, as the moral or political cost of wrongly accepting (or rejecting) a hypothesis increases, the AIR suggests that scientists should require a higher evidential bar for accepting (or rejecting) it. Since being revised and expanded by Heather Douglas (2000), the AIR has remained a fixture in the modern literature on values in science.[5]

Nevertheless, the AIR appears to license *evidential double standards:* requiring different levels of evidence or likelihood for hypotheses, depending on whether they align with our political or other extra-epistemic goals. In some cases, this may yield an undue "preference bias" toward scientists' favored research results (Wilholt 2009).[6] Evidential double standards not only limit the epistemic value of science, but also threaten to intensify polarization and undermine trust in experts. With an eye to these dangers, my aim is to clarify how "cost of mistakes" should be interpreted in the AIR, to help prevent a kind of bias on display in the Covid-19 pandemic.

To this end, I outline a pitfall concerning what I call *(in)convenient hypotheses:* claims whose acceptance by experts is deemed politically or otherwise non-epistemically beneficial (or costly), whether or not these claims are in fact true. Scientific errors then carry costs or benefits that are *only incidentally tied to being mistaken*. As a case study, I examine how, in early 2020, US health officials downplayed the efficacy of masks to help save them for healthcare workers. I argue, however, that benefits like this should not count in the AIR—despite altering the relative cost of Type I and II errors as modeled by Wilholt (2013). Costs or benefits of accepting a given hypothesis, *whether or not it is true,* should not move evidential standards in science up or down. Mistakes can have non-epistemic costs strictly tied to error, though, which the AIR should count. Adjusting evidential standards in view of (in)convenient claims is tantamount to licensing noble lies, manipulation, or wishful

[3] For example, Longino (1995) uses feminist values to motivate alternative theoretical virtues.
[4] Antecedents include Longino (1990); Lacey (1999); Anderson (2004); and Douglas (2009).
[5] See, for example, Steel (2010); Biddle (2016); Elliott and Richards (2017); and the studies cited in the main text corresponding to note 4.
[6] Compare Wagner (2022) on "ends-oriented bias."

thinking. And, whether this adjustment occurs in "pure" research or at the science–policy interface, it often illegitimately trades on the epistemic authority of science.

The fundamental point of this paper is not to criticize specific actions by individual public officials or scientists. Rather, my deeper point is that it is not obvious what, if anything, is wrong with behavior like experts downplaying mask efficacy to save resources for healthcare workers, or scientists rejecting as "conspiratorial" the idea that SARS-CoV-2 accidentally leaked from a lab (see section 6). Existing work on the AIR does not clearly distinguish the issue of (in)convenient hypotheses from other potential sources of bias. And only certain safeguards invoked in previous work on the AIR rule out adjusting evidential standards for (in)convenient claims. Notably, norms like democratic accountability, transparency, and preference for truth over error may not rule out or prevent this sort of adjustment. Other existing work on the AIR, such as research drawing attention to issues of manipulation and wishful speaking (see section 7), is more directly to the point. A range of expert responses to Covid-19 that were met with not universal criticism but rather polarized public response are realistically explained in terms of adjusting evidential standards in view of costs that are incidentally tied to error. In this light, my aim is to clarify why moving the bar for (in)convenient claims is a mistake, and to show how easy it is for experts to fall prey to it.

More precisely, my conclusion is that (in)convenient hypotheses confront society with a choice: We could accept scientists handling (in)convenient claims just as nonscientists do, in which case there would be less reason to give experts political power based on scientific authority. Or, we could hold scientists to a higher epistemic bar, by expecting them not to modify evidential standards to avoid costs that are only incidentally tied to error, insofar as they act with scientific authority. My slogan ("The AIR should only weigh costs/benefits *strictly* tied to error") presupposes the second approach. But the first road may also be viable. My core objection is to a technocracy that combines both: it is inappropriate to give experts great political power based on scientific authority while also allowing them to freely move the bar for (in)convenient claims.

The structure of this paper is as follows: Section 2 fleshes out the case study of US health officials downplaying the efficacy of public masking. Section 3 elaborates operative notions of *(in)convenient hypothesis* and *noble lie*. Section 3 also starts to address the worry that my focus here is on what experts *assert/deny,* whereas the AIR instead focuses on what experts *accept/reject*. Even if this contrast is clear-cut—which I argue is contested—it still raises a question that my account squarely addresses, about how the costs of errors by politicians or the public should inform the evidential standards that experts apply when deciding whether to assert or deny claims in public or political contexts. Section 4 clarifies how someone might realistically take the AIR to license noble lies, or at least adjusting evidential standards in view of (in)convenient claims. Intuitive methods of assessing "costs of error" often indirectly factor in the costs/benefits of accepting or rejecting a given hypothesis regardless of its truth. So, even if advocates of the AIR do not intend to license noble lies, this may unintentionally follow from assessing costs of error without due caution. Section 5 returns to the main case study. Section 6 shows that four safeguards invoked in existing work on the AIR—preferring truth to ignorance and ignorance to error, following scientific conventions, democratic accountability, and transparency—fail to rule out moving the evidential bar for (in)convenient claims. This leaves it unclear whether this sort of adjustment is wrong. Against this skepticism, section 7 shifts focus onto concerns about

manipulation, wishful thinking, and abuse of epistemic authority. In this light, I address the objection that raising the bar for inconvenient claims can be justified in science communication or public policy, even if not in "pure" research. Section 8 concludes with brief comments on unhealthy conventions in the public uptake of science, individual versus group attitudes, and the fundamental aims of the AIR.

2. Case Study: US Public Health Experts Strategically Downplaying Masks

On 17 February 2020, as the pandemic first intensified in the US, Anthony Fauci—the director of the National Institute of Allergy and Infectious Diseases (NIAID), and a lead member of the White House Coronavirus Task Force—downplayed the benefit of universal masking in a meeting with the editorial board of *USA Today:* "If you look at the masks that you buy in a drug store, the leakage around that doesn't really do much to protect you," said Fauci, who is also a widely cited biomedical scientist. "People start saying, 'Should I start wearing a mask?' Now, in the United States, there is absolutely no reason whatsoever to wear a mask" (O'Donnell 2020).

A few months later, Fauci was asked, "Why were we told later in the spring to wear [masks], when we initially were told not to?" (Ross 2020). His reply was striking:

> "It was at a time when personal protective equipment [PPE], including the N95 masks and the surgical masks, were in very short supply. And we wanted to make sure that the people, namely, the health care workers … we did not want them to be without the equipment that they needed, so there was not enthusiasm about going out and everybody buying a mask, or getting a mask. We were afraid that that would deter away from the people who really needed it." (Ross 2020)

Nor was Fauci an isolated case. On 8 March 2020, Surgeon General Jerome Adams claimed in a nationally broadcast interview that "masks do not work for the general public in preventing them from getting coronavirus" (CBS News 2020a). Asked in July 2020 if he had said this "because there wasn't enough equipment," Adams replied: "The science at the time suggested that there was not a high degree of asymptomatic spread. We learned more. There also was, as you mentioned, the very real concern about hoarding of PPE … That was a part of it" (CBS News 2020b).

When Adams appeals to "the science at the time," he is invoking evidence for the hypothesis that "Masks do not work to protect the general public" (M). But the cost of hoarding is not evidence for M; it only introduces a pragmatic, moral, or political benefit to publicly endorsing M. So, we can ask: does this extra non-epistemic *benefit* of potential false positives, or accepting M when it is false, justify *lower* evidential standards for accepting M—as the AIR might appear to suggest? Put differently, is it appropriate to *raise* evidential standards for the "opposite" hypothesis that public masking *is* effective, given this *cost* of accepting it in error—if asymptomatic masking by the public in fact would not effectively limit disease spread, but would divert scarce PPE from healthcare workers who would benefit from it, in higher-risk contact with symptomatic patients?

On my view, this kind of cost or benefit should not count. But the problem is not with the AIR, per se. It is just that the relevant notion of "cost of mistakes" has to be qualified. In

the AIR, the "cost of mistakes" should only be their cost *as mistakes,* not also the cost of mistakes simply as instances of accepting an inconvenient hypothesis that happens to be false, or rejecting a convenient hypothesis that happens to be true. In this light, it would be inappropriate for Fauci or Adams, at least as scientists or as officials leveraging the epistemic authority of science, to raise the evidential bar for the hypothesis that "Masks protect the general public" in view of the moral or pragmatic cost of hoarding that diverts scarce PPE away from healthcare workers. Advocating mask use could have this downside whether public masking is actually effective or not, so this downside should not factor into a cost-benefit analysis of mistakes strictly as mistakes. Denying mask efficacy in order to prevent hoarding might not be a full-fledged noble lie, if it actually was significantly uncertain. Still, it would be tantamount to a noble lie, rooted in an abuse of the AIR.

It might seem that the AIR has little to do with noble lies, and that advocates of the AIR never meant to license actions like Fauci's. Here, it is useful to clarify the probative value of this specific case study, before returning to the wider issue of noble lies and their relation to the AIR.

Fauci's initial downplaying of masks arguably played an important role in undermining trust in public health authorities in the US, which in turn plausibly exacerbated the pandemic and stoked political polarization. This polarizing effect is suggested by contrasting coverage in the *National Review* (a conservative magazine) and in *CNN* (catering to a liberal audience): the *National Review* wrote that Fauci "embraced the 'noble lie' as a tool" (Schorr 2021), whereas a *CNN* headline said, "Top health officials have changed their minds about face mask guidance – but for good reason" (Yan 2020). My aim is not to add more fuel to these partisan fires. Given its impact, however, it is worth asking about the justifiability of Fauci's behavior, along with other contested expert responses to the pandemic examined below. This media coverage also illustrates that people who endorse Fauci's action do not see it as a noble lie; instead, they may emphasize rational responses to non-epistemic costs of error, in a way that at least *looks* similar to the AIR.

Moreover, even if most or all scholars who advocate the AIR never meant to endorse behavior like Fauci's, the AIR may be used, rightly or wrongly, to justify his approach. After all, what Fauci did may be intuitively troubling, yet still ultimately appropriate. (Matthew J. Brown's denial of a lexical priority of evidence over values (2020), for example, might be taken to imply that Fauci-style reasoning was acceptable.[7]) Or, even if the AIR, properly applied, does not license Fauci's behavior—which is my own view—still, public *misuse* of AIR-style reasoning may function to "license" irresponsible behavior by expert advisers or officials. Given this potential for misuse, it is important to clarify how the AIR should or should not be applied in contested cases like this.

Two examples will start to show that influential work on values in science often does not clearly rule out Fauci's behavior (see also sections 4 and 6). First, Douglas might object that

[7] I am grateful to an anonymous reviewer for this suggestion. I am also grateful to another reviewer who objected that Brown's account cannot be formalized using decision theory, due to his stressing how, using moral imagination, we creatively develop novel courses of action that might help us avoid or mitigate trade-offs between false positives and false negatives. However, this role for creativity can be accommodated within a decision-theoretic framework: whenever a new course of action is imagined, its costs/benefits can be analyzed and factored into a standard decision-theoretic approach. This process can be iterated as often as creative imagination necessitates. The possibility that someone will come up with a creative solution that avoids unfortunate trade-offs may also be a reason to avoid either endorsing or denying a given claim—that is, it may add utility to strategically holding off.

Fauci used values in a *direct* role; that is, "providing warrant or reasons to accept a claim," rather than in the *indirect* role she endorses, where values only "determine the importance of the inductive gaps left by the evidence" (Douglas 2009, 96). However, it is more charitable to suppose that Fauci used values in an indirect way, by altering his evidential standards for denying mask efficacy in view of the risk of hoarding. It is clear that hoarding is not an *epistemic* reason to accept claims about mask inefficacy.[8] So why impute to Fauci the clearly false view that the risk of hoarding makes mask inefficacy *likelier?* More charitably, but arguably still at the level of direct influence, Fauci may have viewed hoarding as a *pragmatic* reason to accept (or endorse) a claim about mask inefficacy. But given widespread criticism of direct use of values in science, it is even more charitable—and still realistic—to interpret Fauci as simply *adjusting his evidential bar* for endorsing claims about mask efficacy. This raises a broader worry: cases can be analyzed in a biased way by claiming that behavior we antecedently disagree with uses values "directly," while behavior we agree with uses values "indirectly." But this begs the question at issue. At a minimum, the publicly available details of many cases leave it unclear whether they involve a direct or indirect use of values. And interpretive charity favors the latter, even if not decisively.

Second, Fauci arguably violated an epistemic standard proposed by Daniel Steel (2010, 15): "Influences of non-epistemic values on scientific inferences are epistemically bad if and only if they impede or obstruct the attainment of truths." Here, there are several important points to note. To start, Steel's principle describes only the *epistemic* badness of non-epistemic value influence. From a consequentialist point of view, this would then have to be weighed against *non-epistemic* factors to assess how appropriate Fauci's behavior was overall. This is where it is necessary to examine extra-epistemic safeguards on the AIR. My point is that various safeguards emphasized in existing work on the AIR, such as democratic accountability and transparency, may fail to prevent actions such as Fauci's. Norms against manipulation and abuse of scientific authority are more to the point. Of course, many scholars have emphasized issues of public trust and abuse of authority. The added value of my analysis, on this point, is partly a clarification of how abuse of scientific authority comes apart from considerations of transparency, democracy, preference for truth, and adherence to scientific conventions, in ways that existing work on the AIR does not clearly stress.

It is also worth showing how adjusting evidential standards for (in)convenient claims is objectionable from not only consequentialist, but also deontological or virtue-ethical standpoints. From a deontological point of view, Steel's principle can be taken as a side constraint on science, implying that scientists have a strict duty not to impede or obstruct the attainment of truths. Fauci's behavior may then seem to be immediately ruled out by a basic epistemic duty that he has insofar as he acts with scientific authority. Even from this standpoint, however, it is worth clarifying exactly *how* Fauci obstructed the attainment of truth beyond what is typically involved in legitimate uses of the AIR. Again, for instance, it

[8] Hoarding of masks by the public is premised on the idea that masks *are* (at least decently likely to be) effective for the public. If masks were more like homeopathic remedies, hoarding would not matter and the risk of being wrong about that would be like the risk of being wrong about hoarding homeopathic remedies: high cost but very low probability. Part of what is at stake in the Fauci case is the difference between the efficacy of PPE *when used by frontline healthcare workers* and the efficacy of the same PPE *when used by members of the general public*. Conceivably, masking could be far more effective in clinical contexts, for example, as a result of increased viral load in the air, or exposure to expectorated viral matter around symptomatic patients. Charitably viewed, it is likely that Fauci had this sort of asymmetry in mind when he downplayed *asymptomatic masking by the public*.

is perfectly realistic to interpret Fauci as having used values in an *indirect* role, by giving more weight to uncertainty about mask efficacy in view of the shortage of PPE. In this light, the difference between satisfying and violating duties to display epistemic integrity cannot be explained in terms of Douglas's "direct vs. indirect" distinction—if indeed Fauci's behavior actually constituted a breach of epistemic integrity. My account can be understood, then, as a principled explanation of the difference between epistemically legitimate and illegitimate *ways* of using values indirectly. Absent this further explanation, one might think that Fauci *did not actually impede the attainment of truth,* or else that *all* uses of the AIR do so.

It is also debatable whether scientists have a strict duty not to impede or obstruct the attainment of truths.[9] Here, much rests on what exactly it means to "interfere with, block, or obstruct the acquisition of truths" (Steel 2010, 26). For present purposes, the crucial point is that a duty to discount convenient benefits and inconvenient costs when deciding what evidential standard to apply to a given hypothesis may be less restrictive than a Steel-style duty. My view leaves room for behavior like Fauci's to be inappropriate even if there are also cases in which scientists *legitimately* obstruct the attainment of truths in view of non-epistemic values, as in choosing not to work on certain research questions for moral or political reasons. For instance, "gain-of-function" research on bat coronaviruses may attain many truths, yet it may nevertheless be appropriate for scientists and governments to obstruct this process in view of practical risks. Section 7 also argues that we are *epistemically responsible* for mistakes tied strictly to error, but not for mistakes that are tied to error only incidentally. So, on my view, the relevant duty may be one to *act with epistemic responsibility,* rather than a duty *never to block the attainment of truths*.

As mentioned in section 1, my core objection is giving experts like Fauci political power based on scientific authority while also allowing them to freely adjust their standards for (in)convenient claims. One way to resolve this tension would be to allow experts to alter their evidential standards for (in)convenient claims, but in this light to seriously limit their political influence. This would help to address concerns about abuse of scientific authority by way of limiting scientists' political power, rather than by holding scientists to a higher epistemic bar.

3. Noble Lies, (In)Convenient Hypotheses, and Acceptance

Before elaborating how the AIR might be taken to license *noble lies,* or at least moving the evidential bar for *(in)convenient* claims, it is crucial to clarify how these terms are used here.

Sometimes experts may identify a moral, pragmatic, or political benefit to accepting or endorsing a given hypothesis, whether or not it is actually true. As we have seen, for example, accepting the claim that "Asymptomatic masking by the public does not significantly reduce the spread of Covid-19" plausibly had the benefit of saving scarce PPE for healthcare workers dealing with symptomatic patients. As section 4 describes in detail, this reduces the cost of false positives relative to false negatives, since accepting a *convenient* hypothesis like this brings the benefit in question, even if it is certainly false. The benefit of accepting a convenient hypothesis can be reframed as a relative cost of

[9] To be clear, I am not arguing that Steel's principle is misguided, only that it could be viewed as overly restrictive and it is worthwhile to see how behavior like Fauci's might also be ruled out by weaker epistemic duties.

rejecting it.[10] So, mistakenly rejecting a convenient hypothesis, if it is true, carries costs—but "costs of error" that are *tied to error only incidentally*.

In extreme cases, experts may deem it better to accept or endorse a given hypothesis even when highly confident that it is false, in view of non-epistemic benefits of accepting it or costs of rejecting it. If it is benevolently motivated, endorsing a hypothesis in this situation is a *noble lie*. Altruistic intent may distinguish noble lies from biases toward research results favorable to corporate sponsors in industry-funded science (on which, see, for example, Holman and Elliott 2018).

Accepting or endorsing a claim may also impose costs whether or not it is true, making it *inconvenient*. Denying a hypothesis even when highly confident that it is true, due to perceived societal benefits of doing so, is another *noble lie*. I use the term "(in)convenient" to refer to either convenient or inconvenient claims, or to both at once. Claims are (in)convenient specifically insofar as mistakenly rejecting (accepting) them has costs that are tied to error only incidentally.

As section 4 elaborates, the AIR might reasonably seem to license noble lies, or at least lowering the evidential bar for convenient claims and raising the evidential bar for inconvenient ones. But, on my view, this misapplies the AIR. Costs or benefits of saying that a hypothesis is true *whether or not it is* should not influence scientists' evidential standards. By contrast, non-epistemic costs of mistakes *strictly as mistakes* should influence scientists' evidential standards.

Finally, a few remarks on the idea of "acceptance" are in order. My analysis may seem to equivocate between experts' *acceptance* or *rejection* of hypotheses and what they *publicly assert* or *deny*. For example, Fauci might have publicly downplayed public masking without privately rejecting its efficacy. So, insofar as the AIR focuses on what scientists accept or reject, it may seem unrelated to the Fauci case. But this is misleading, for four key reasons (see also section 8).

First, "acceptance" is a contested notion in philosophy of science. Some theorists take it to be a narrowly epistemic attitude that involves commitment to the truth of a given claim (for example, Goldman 1999; Adler 2002). But others see "acceptance" as a more pragmatically inflected attitude. For instance, Catherine Z. Elgin understands "acceptance" of *p* in terms of "being willing to take *p* as a premise, as a basis for action or … as an epistemic norm or a rule of inference, when one's ends are cognitive" (2017, 19). Similarly, for Bas C. van Fraassen, "accepting" a theory involves not only a belief that it is empirically adequate, but also a more practical "commitment" to "confront any future phenomena by means of the conceptual resources of this theory" (1980, 12). For Van Fraassen, moreover, theoretical virtues beyond empirical adequacy are "*pragmatic* virtues" that concern "the use and usefulness of the theory" rather than "the relation between the theory and the world" (1980, 88). The crucial point is that pragmatically inflected notions of scientific "acceptance" are less clearly opposed to "assertion"—to accept a given hypothesis, in a pragmatically inflected sense, may require certain patterns of assertion or practical uptake in contact with other experts, but *not* personal belief in its truth or likelihood. In this light, assertion or denial by experts speaking publicly with scientific authority (like Fauci) may in fact be aptly understood as a component of acceptance or rejection, in one relevant sense.

[10] Section 2 gave one example of this sort of reframing; section 4 gives another.

Second, and relatedly, the AIR can be reframed in terms of attitudes beyond acceptance. For example, Stephen John distinguishes "acceptance" from "assertion" before "follow[ing] Douglas in framing the problem of inductive risk in terms of assertion, rather than acceptance" (2015, 81; compare Franco 2017). Thus, the appearance that work on inductive risk focuses on acceptance, rather than assertion, may be misleading. Likewise, Hugh Lacey contrasts "holding," "adopting," and "endorsing" before advancing an AIR-style argument about *endorsement,* or treating the evidence supporting a given claim, *p,* as "*sufficiently strong* that the legitimacy of actions informed by it should not be challenged on the ground that *p* has insufficient empirical support" (2015, 93). Lacey, too, cites Douglas's work on inductive risk: "What counts as 'sufficiently strong' varies with how ethically salient are the consequences of acting on *p*, should it actually be false; the more ethically problematic they are, the stronger the supporting evidence should be for *p* to be endorsed (cf. Douglas, 2009)" (Lacey 2015, 93). So, even if Fauci did not *accept* a claim about the inefficacy of asymptomatic masking by the general public, he may still have *endorsed* this claim; and the AIR, or essentially the same logic, may apply—perhaps even better—to attitudes like endorsement. This is why I used phrases like "accepting or endorsing" earlier in this section.

Third, while the "benefits" of Fauci's decision only directly follow from what he says, not from what he believes to be true or likely, note that the same can be said for most, if not all, costs of error in the context of the AIR. Revisiting one of Rudner's cases: the non-epistemic costs of mistakenly accepting the hypothesis that "This drug is nonlethal" (if it is lethal) only follow from *acting* on this belief—for example, by distributing the drug or publicly endorsing its safety. Plausibly, then, the non-epistemic costs of scientific error *always* only directly follow from what scientists do or say—not from mistaken acceptance or rejection viewed in isolation from the actions that follow.

Fourth, even if a distinction between assertion or endorsement and (veritistic) acceptance is sustainable, the AIR has bearing on both. Traditionally, the AIR concerns adjusting evidential standards for scientific acceptance or rejection in view of costs of mistakes by scientists. But the core logic of the AIR is naturally extended to a related issue: whether experts may legitimately adjust their evidential standards for *publicly asserting or denying* hypotheses in view of costs of error by the public, politicians, or other nonexperts who may mistakenly *accept or reject* these hypotheses as a result. This issue arises even if "acceptance" is viewed as a narrowly epistemic or veritistic attitude. The key point is that what experts publicly endorse or deny (downplay, and so on), when speaking with perceived expert authority, may result in mistaken acceptance or rejection of hypotheses by nonexperts. And my view squarely addresses this issue: experts *can* legitimately adjust their evidential standards for endorsing hypotheses, in view of the non-epistemic costs of errors *by nonexperts,* but these "costs of error" should not include costs tied *incidentally* to error.

4. Noble Lies and (In)Convenient Hypotheses in the AIR?

To see how Fauci and others might realistically take the AIR to license noble lies, or at least adjusting evidential standards for (in)convenient claims, it helps to ask: how exactly should one assess the cost of Type I/II error? Here we may first follow Carl G. Hempel (1965, 92)—who coined "inductive risk" in a 1960 essay that Douglas (2017, ix) acknowledges as an early influence—in identifying four possible "outcomes" of a scientist's decision to accept or reject a hypothesis, H: (1) accepting H when it is in fact true; (2) rejecting H when it is false; (3)

accepting H when it is false; and (4) rejecting H when it is true. In modern terms, these outcomes are (1) a *true positive;* (2) a *true negative;* (3) a *false positive;* and (4) a *false negative*. "Inductive risk" is the risk of either (3) or (4).

Hempel frames the core of the AIR in this context: "The problem of formulating adequate rules of acceptance and rejection has no clear meaning" without "assigning definite values or disvalues to those different possible 'outcomes' of acceptance or rejection" (Hempel 1965, 92). And here non-epistemic values enter, as in cases of "monetary gains or losses" when accepting or rejecting a hypothesis may lead to the success/failure of some "intended practical application" (1965, 92–93). For present purposes, it helps to use a more abstract notion of *utility* that includes not only epistemic and monetary costs/benefits, but also other pragmatic, political, or moral costs/benefits. For ease of reference, let us designate the utilities of outcomes (1)–(4), for a generic hypothesis, H, as follows: p_T (true positive), n_T (true negative), p_F (false positive), and n_F (false negative). Following Torsten Wilholt (2009, 2013), these utilities may be depicted as in Figure 1.

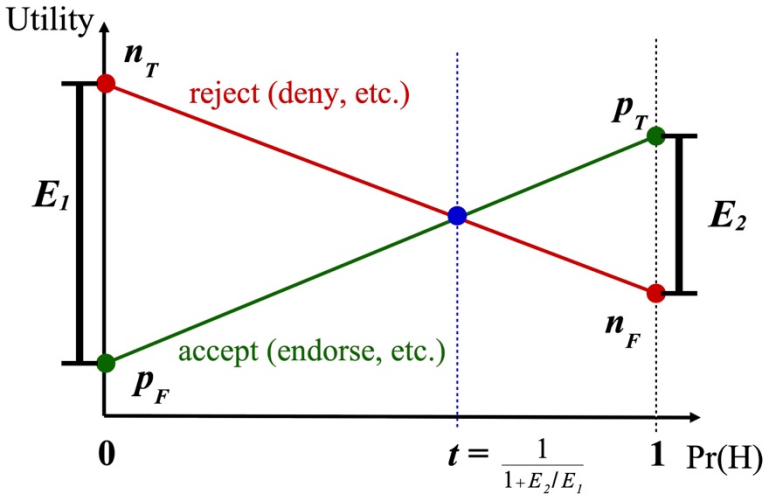

Figure 1. A model of how utilities affect inquiry in the simplified case of two options: "accept" or "reject" the given hypothesis, H. Pr(H) is the subjective probability assigned to H by a given expert or group of experts in light of available evidence. As indicated in parentheses, similar models may be framed in terms of alternative pairs of attitudes or actions, such as publicly "endorse" vs. "deny" (note that the utility of acceptance may differ from that of endorsement or assertion—see sections 3 and 8). Adding additional options like "no result" yields additional confidence thresholds between rational changes in attitude or behavior. The *cost of a false positive* (E_1) is the difference in utility between a true negative (n_T) and false positive (p_F). The *cost of a false negative* (E_2) is the difference in utility between a true positive (p_T) and false negative (n_F). When Pr(H) < $t = 1/(1+E_2/E_1)$, *rejecting* H maximizes expected utility. When Pr(H) > t, *accepting* H maximizes expected utility. Compare Wilholt (2013, figures 1–4).

Our initial question may now be restated more clearly: how are "costs of error" in the AIR related to these four utilities? Two crucial points will begin to reveal the relevance of noble lies.

First, the cost of a false positive for a given hypothesis, H, is naturally understood as *how much worse it is to mistakenly accept H than to correctly reject it, given that it is actually false;* that is, the cost of a Type I error is $E_1 = n_T - p_F$ (the difference in utility

between a true negative and false positive). Likewise, the cost of a false negative is *how much worse it is to mistakenly reject H than to correctly accept it, given that it is true;* that is, the cost of a Type II error is $E_2 = p_T - n_F$ (the difference in utility between a true positive and a false negative) (Figure 1). This is how costs of Type I/II errors are treated in influential work by Wilholt (for example, 2013, 243).[11] The utilities of true positives/negatives may thus be *indirectly* relevant to "costs of error" in the AIR, even if the AIR explicitly seems to focus only on the (dis)utilities of the two "error" outcomes (p_F and n_F).

But perhaps other approaches to the AIR, like Douglas's, do not require any role for the utility of true positives/negatives, even if Wilholt's does? In fact, however, this is far from clear. Douglas (2009) does focus on costs of error, devoting little attention to benefits that accepting a claim may bring. But this explicit focus on costs of error is compatible with the implicit or even unintended relevance of (1) the utility of true positives/negatives, as above; and (2) both the costs *and the benefits* of a given error. To the latter point, note that the benefit of a convenient false positive can always be reframed as the *relative cost* of a corresponding false negative (or vice versa). For example, if rejecting the hypothesis that "SARS-CoV-2 accidentally leaked from the Wuhan Institute of Virology" has the benefit of limiting xenophobia (section 6), this benefit can be reframed as a relative cost of Type I error, or accepting this hypothesis if it is false. So, a use of the AIR that seems to ignore the benefits of convenient error may in fact smuggle them in.

Second, in the AIR, the cost of false positives must always be weighed against the cost of false negatives, and vice versa: neither Type I nor Type II errors should be considered in isolation. Failure to compare the cost of Type I and Type II errors lets inappropriate forms of politicization enter into science. For example, when the political left or right endorses a measure to prevent or treat Covid-19, the "other side" may focus only on potential false positives, and so claim that this measure fails to meet a *higher* evidential bar that is warranted ("You claim this drug/vaccine is safe and effective—but what if you're wrong? Think of the suffering!"). On the other hand, when either political side looks at its own favored policies, it may instead stress potential false negatives, leading each side to *lower* the evidential bar that it applies to itself ("What if the drug/vaccine we support is safe and effective, but you wrongly reject it? Think of the suffering!").

Wilholt's model elegantly captures these two features of "cost of error" in the AIR. First, the cost of Type I errors seems to be well described by E_1, the utility difference between correctly rejecting and mistakenly accepting a false claim (and likewise for Type II errors as E_2). Second, the costs of false positives and negatives must be weighed against each other—which Wilholt's model captures insofar as it yields a precise expression for the confidence threshold (t) between rational acceptance and rejection, which depends only on the *ratio* E_2 / E_1: accepting H maximizes utility if $\Pr(H) > t = 1 / (1 + [E_2 / E_1])$, and rejecting H maximizes utility if $\Pr(H) < t$ (Figure 1).[12]

These expressions distill the AIR. As false positives get costlier versus false negatives, it maximizes expected utility to *raise the evidential bar* for accepting the given hypothesis (that is, as E_2 / E_1 decreases, the rational acceptance-rejection threshold t approaches Pr(H)

[11] To be precise, Wilholt (2013, 241) identifies "the 'costs' of false positive errors" with the difference in utility between communicating "no result" and "communicating *S*" for a given false hypothesis *S*. This slight deviation from my formulation just reflects that Wilholt considers three options—basically, *accept, reject,* and *no result*—instead of two. For simplicity's sake, I am presenting the corresponding analysis for a two-option setup. The dynamics at issue in my argument are unaffected by this difference: more options simply add more thresholds.
[12] See note 11.

= 1). On the other hand, as false negatives get relatively costlier, *lowering the evidential bar* for H maximizes expected utility (that is, as E_2 / E_1 increases, the acceptance-rejection threshold t approaches Pr(H) = 0) (Figure 2).[13] At the limit of full certainty that H is true/false, a basic epistemic preference for correctness over error is decisive, but non-epistemic values also matter given any inductive risk. The AIR can be viewed as downstream of decision theory, insofar as Wilholt-style models show how the evidential bar-moving that the AIR recommends is a special case of maximizing expected utility—even if, in practice, this happens informally and without being seen as utility-maximizing. This is evidently why Hempel (1965, 93) takes "decision rules" used in contexts like "industrial quality control" to be valid "rules for acceptance" of scientific hypotheses with practical bearing.

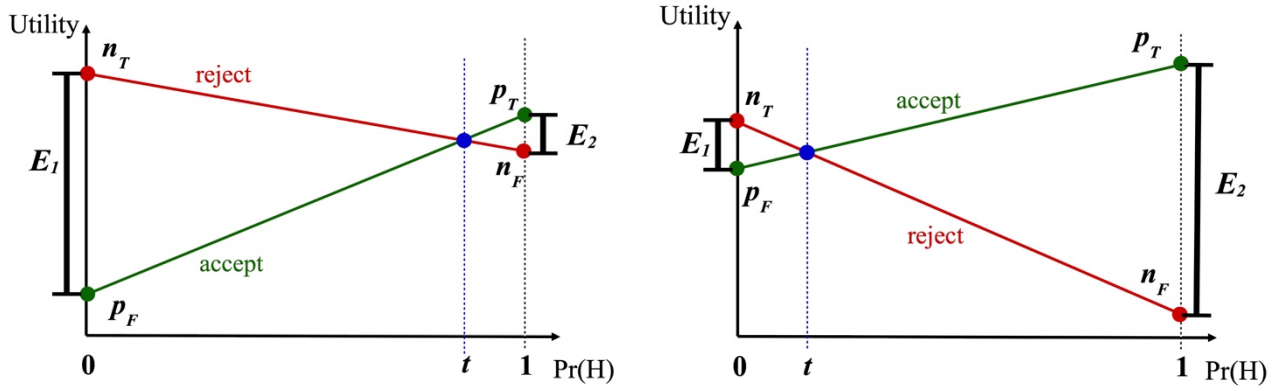

Figure 2. (a) (left) As false positives become relatively costlier (as E_1 / E_2 increases), applying a *higher* evidential standard to H maximizes expected utility—as reflected in the threshold, t, approaching Pr(H) = 1. **(b)** (right) As false negatives become relatively costlier (as E_2 / E_1 increases), applying a *lower* evidential standard to H maximizes expected utility.

So far, so good. But this is where noble lies and inconvenient truth enter the picture. In visual terms, the problem is that inconvenient claims raise evidential standards *for the wrong reason:* the threshold t moves rightward (toward Pr(H) = 1) because the line representing utilities of accepting H at every confidence level moves uniformly downward (Figure 3), but evidential standards in science should only move in view of costs *anchored* either to the truth or falsity of a given H (Figure 4).[14] In conceptual terms, the problem is

[13] Compare Wilholt (2013, 239).

[14] A reviewer objected that, for any two pairs of utilities p_T and p_F, they can be trivially construed as "bearing" the same contributing cost/benefit u for *any conceivable value of u.* In the same way that it would be trivial to claim that a 5'6" and 5'8" person "bear" in common every value of height up to 5'6", it would be trivial to claim that the true positive (p_T) and false positive (p_F) scenarios "bear" every imaginable (positive or negative) utility in common with each other. That is true—however, this sort of arbitrarily posited utility that p_T and p_F "share" *will not in general correspond to actual benefits or costs.* The cost of keeping masks from healthcare workers is an actual cost (if masks are effective). An arbitrarily large artificial "cost" that one could *stipulate* is borne by both the "true positive" and "false positive" outcomes in any given case *is not a real cost.* It is an *imaginary* inconvenient cost, if not just a *mathematical artifact,* lacking even an imaginary referent. My focus is on actual (in)convenient benefits (costs). It might be further objected, however, that total utility has no nonarbitrary decomposition that would allow this sort of distinction between "real" and "unreal" costs/benefits to be drawn. For example, a reviewer suggested that typed utilities may be used to decompose total utility in nonarbitrary ways, but only at the cost of raising worries about incommensurability, which would block any use of Wilholt's acceptance/rejection decision formula. But it is not clear that typed utilities are needed to distinguish between real costs/benefits and mathematical artifacts. Moreover, incommensurability is arguably a problem for *all*

that, without my proposed addendum, Wilholt-style approaches to the AIR fail to distinguish between (a) costs of mistakenly accepting a false hypothesis, H, which *would not* have been incurred if H were true (Figure 4); and (b) costs of mistakenly accepting H, which *would* still have been incurred if H were true because H is *inconvenient* (Figure 3).[15] Both (a) and (b) are "costs of error" in some sense. Even so, on my view, only costs of type (a) should be factored into the AIR. Insofar as Wilholt's model clarifies how to do the cost-of-error calculations that Douglas's approach requires, and insofar as Douglas does not offer an alternative, my critique also extends to nonformal approaches like Douglas's.

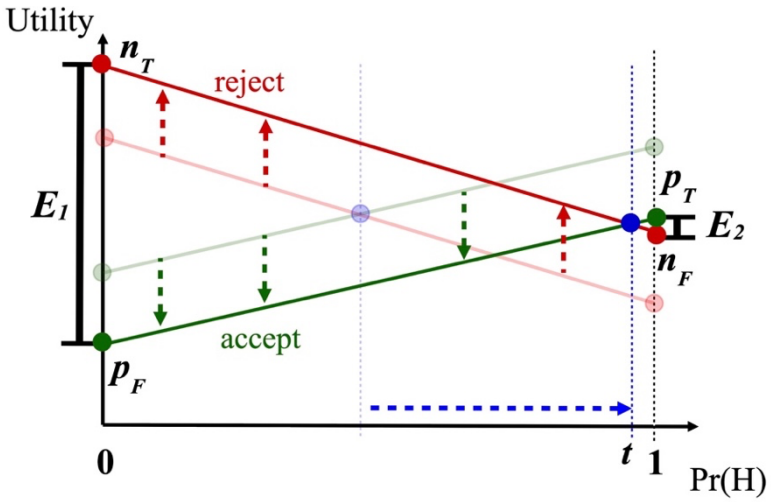

Figure 3. How *inconvenience* may affect evidential standards via misuse of the AIR: costs of accepting H (or benefits of rejecting H), irrespective of its truth or falsity, lower the "accept" line (or raise the "reject" line). This makes it rational to raise the evidential bar for H (reflected here in t approaching Pr(H) = 1), but for the wrong sort of reason. Inconvenience adds real costs to potential errors, but costs tied to error only *incidentally*.

Douglas's account might seem to rest on deontic concepts of *recklessness* and *negligence* rather than on the consequentialist framework that the above model could be thought to require. My response is severalfold. First, the basic question is whether or not scientists should consider costs tied *incidentally* to error; Douglas must answer this question whether or not her account can be recast in decision-theoretic terms. Second, even if there is a strict duty to avoid recklessness or negligence, Douglas evidently thinks that scientists discharge

versions of the AIR. For example, Douglas is committed to considering the non-epistemic costs of potential error. But what if these costs are incommensurable? In this light, any problem that incommensurability might present is not a distinctive problem for the approach adopted here. A reviewer argued that incommensurability is only a problem for consequentialism, as other traditions rely on *judgment* to adjudicate incommensurability in given cases. Whether judgment is up to that task lies beyond the scope of this paper. Regardless, my basic goal is to establish that people acting with scientific authority should not adjust their evidential standards for (in)convenient claims. Douglas takes no explicit stance on this issue and adjusting evidential standards for (in)convenient claims involves only an "indirect" use of values (which Douglas allows), so it is important to clarify how it should be treated by *both* consequentialists and deontologists. Also my analysis may help to clarify exactly *how* Fauci-style behavior is *reckless* or *negligent*—which is not obvious.

[15] Analogously, one may distinguish (a) benefits of mistakenly accepting a false hypothesis, H, which *would not* have been incurred if H were true from (b) benefits of mistakenly accepting H, which *would* still have been incurred if H were true, because H is *convenient*.

this duty by considering consequences of error: we "have a general moral responsibility to consider the consequences of error, based on our concern over reckless or negligent behavior" (Douglas 2009, 69; compare with Havstad 2022, 306). It may be impossible to assess whether an action is negligent or reckless without considering the relative likelihood of outcomes it could yield. In this light, comparing the relative cost of false positives and negatives may be needed to decide whether using a given evidential standard is reckless or negligent. This analysis may be informal and imprecise, but that is enough for present purposes.[16]

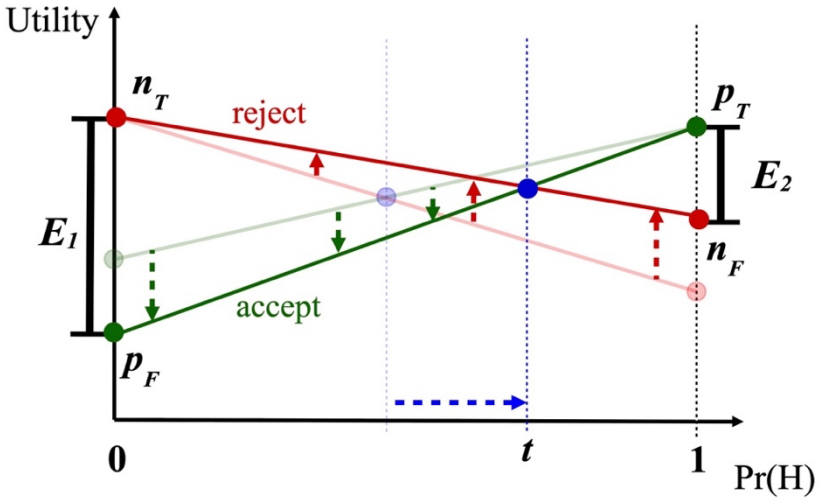

Figure 4. Appropriately raising evidential standards in view of the non-epistemic costs or benefits of error: when costs or benefits are tied to error *strictly as such,* the AIR applies. For example, if H is "Hydroxychloroquine (HCQ) is a safe/effective treatment for Covid-19," many costs of endorsing HCQ when it is in fact unsafe or ineffective are valid reasons to raise the evidential bar for H. Unsafe drugs cause harms that safe drugs do not, and endorsing ineffective drugs diverts focus from effective treatments in a way that endorsing effective drugs does not. So, these are costs of false positives strictly as errors, not simply as instances of accepting a claim that happens to be false. Relevant costs/benefits are *anchored* to the truth or falsity of the hypothesis at issue.

From a decision-theoretic standpoint, my proposal to ignore "convenient benefits" and "inconvenient costs" might appear to be irrational—after all, if these are real costs and benefits, it would be unreasonable to attempt to maximize expected utility without factoring them in. But this is crucially misleading. To see this, note that my proposal can be represented in two ways, which will yield essentially the same practical upshot: (i) scientists can *ignore* inconvenient costs when thinking about how to set evidential standards; or (ii) scientists can *factor in* inconvenient costs, but also factor in counteracting *higher-order* benefits tied to factors like avoiding abuse of scientific authority. The only way to avoid abusing authority, manipulation, or fostering wishful thinking among members of the public (sections 7–8) is to avoid adjusting evidential standards for (in)convenient claims. So, there is a massive higher-order benefit to setting evidential standards at the level that expected-utility maximization *would* dictate if one simply discounted inconvenient costs

[16] Also worth noting is the "consequentializing" program—see, for example, Portmore (2007).

altogether.[17] With this higher-order benefit included, expected-utility maximization yields essentially the same practical recommendation: when setting evidential standards, people acting with scientific authority should act *as if* inconvenient costs and convenient benefits did not exist.

Earlier work on inductive risk has not isolated this issue, let alone taken a stance on it. For example, Wilholt defines *preference bias* as violating a "conventional standard" of a research community "in order to increase the likelihood of arriving at a preferred result" (2009, 99). But it is unclear that any conventional standards in public health sciences rule out what top US health officials did in early 2020 when they downplayed the efficacy of masks to help save them for healthcare workers (section 6). Indeed, Wilholt concedes that his account does not cover cases where "the conventional standards of a research community are themselves distorted by interests and preferences in an epistemologically problematic way" (2009, 99). My analysis helps fill this gap: if a research community has a convention of adjusting evidential standards in order to realize benefits or avoid costs tied to errors, not strictly *as errors,* but rather just as cases of accepting or rejecting a claim *regardless of its truth,* this creates epistemological problems that Wilholt's account of preference bias does not address. I return to this and other links to earlier work in section 6. First, however, revisiting the mask case will help to clarify my position.

5. Fauci Revisited

Downplaying mask efficacy to prevent hoarding of PPE is, on my view, tantamount to a noble lie, rooted in abuse of the AIR. If Fauci's aims had been transparent, the public rightly would have taken his critiques of mask efficacy to be seriously unreliable. The solution is not greater transparency about adjusting standards for (in)convenient claims, but rather greater reluctance among experts to indulge in it.

Several further points of clarification are also in order. First, my position still allows Fauci and other public health experts to use different tactics to save masks for healthcare workers—just not by altering evidential standards. For example, Fauci could have simply said, "Please do not buy masks: there is a shortage and we need to save them for healthcare workers, who deserve and benefit more from masks since they face an especially high risk." This might have had the desired effect, without requiring experts to selectively raise the evidential bar for claims about mask efficacy. Of course, it might result in more people hoarding masks out of self-interest. But this is neither obvious nor inevitable: asking people to be responsible may have had a similar benefit, while more effectively avoiding a loss of credibility.

Second, even if it was inappropriate to say that there was "no reason" to wear masks, one might ask: would it at least have been fair to stress that there is *uncertainty* about how protective masks are? (Compare Douglas's claim that scientists "choose whether or not to emphasize the importance of the uncertainties," and that "this is where the weighing of the consequences of error [legitimately] comes in" (2009, 81).) This is still problematic. Selectively emphasizing uncertainty for political benefit is just a higher-order instance of

[17] For consequentialists, even if this higher-order benefit is "massive," in principle it may be outweighed by other factors, so that altering standards for (in)convenient claims is preferable. Nothing I say is meant to rule this out. Rather, I stress higher-order costs that consequentialists might overlook or downplay and I try to clarify how a deontological duty to avoid altering standards for (in)convenient claims may be contextualized and justified.

lowering evidential standards to make it easier to accept a convenient claim. In this light, it would be concerning for Fauci to selectively stress uncertainty about mask efficacy to divert masks to healthcare workers. It may be fitting for *other* officials to do this—but, as I elaborate in section 7, it is less fitting for those who "represent science," as Fauci said he does, in a televised interview (CBS News 2021).

Third, my view still allows that value-laden adjustments to evidential standards are often appropriate—because scientific errors often have costs that are more directly tied to being wrong. Mistakes often have costs strictly as mistakes, and these should count in the AIR. For example, if the hypothesis at issue is that "Masks can significantly protect the public against Covid-19," then a false negative—rejecting this hypothesis if it is in fact true—would have certain costs directly tied to error. In particular, members of the public who are told that masks are ineffective may be infected at higher rates. This cost is strictly tied to error since it is only incurred if masks actually offer protection; if not, then the public would not be missing a genuine benefit. Experts can therefore legitimately take this cost of a potential false negative as a reason to lower the evidential bar for accepting the claim that masks protect the general public.

Similarly, if scientists or public health experts are weighing the hypothesis that a given drug is a safe and effective treatment for some disease—for example, hydroxychloroquine, which was "hyped" as a Covid-19 treatment (Intemann 2022)—the cost of endorsing this drug when it is in fact unsafe or ineffective is a valid reason to raise evidential standards (Figure 4). Unsafe drugs cause harms that safe drugs do not, and endorsing ineffective drugs diverts people from effective treatments in a way that endorsing effective drugs does not. So, these are costs of false positives strictly as errors, not just as instances of accepting a hypothesis that then happens to be false. By the same token, however, pandemics may make false *negatives* far costlier, since preventing access to safe and effective treatments may cause suffering and death. In this light, there is also a countervailing reason to *lower* evidential standards—as in the recent distribution of SARS-CoV-2 vaccines via the lower-than-normal evidential standards of emergency authorization protocols. Again, this cost of error is appropriately tied to error *as such* and so, on my view, is admissible.

Finally, even if one agrees that Fauci adopted extremely high standards for claims about mask efficacy, it may be unclear that this involved any inconsistency or change from a "normal" approach on his part. After all, it is not clear that there is a "normal" bar for claims about mask efficacy, let alone what it might be. By contrast, in many cases where the AIR is applied, there is a far clearer sense of what "normal" practice would be—as in conventionally requiring a specific threshold of statistical significance within a given research community, like a maximum p-value of 0.05 in many areas of social science. However, the absence of explicit or precise norms does not mean that it is impossible to make more or less plausible assessments of whether a given act of acceptance (or endorsement) involves deviation from more informal evidential norms or standards. These assessments may just require more nuanced judgments, informed by social scientific or common knowledge about relevant factors such as cognitive biases or polarization. This assessment may also often involve counterfactual reasoning. For example, it is relevant to ask, "Had PPE not been scarce, would Fauci have applied lower evidential standards for endorsing claims about the efficacy of public masking?" To the extent that the answer is plausibly "Yes" (as suggested by Fauci's public comments), this counterfactual analysis reveals an implied alternative evidential bar. This alternative is not necessarily a "normal" state. But it is a lower evidential bar for an

analogous claim that is less inconvenient than the claim actually at issue. Alternatively, we can consider counterfactuals that *invert* the cost-benefit calculus at stake. For example, if Fauci had found claims of mask efficacy to be *convenient,* rather than *inconvenient,* would he have adopted a different evidential standard when being interviewed by *USA Today?* Insofar as the answer is plausibly "Yes," there is reason to believe that Fauci performed the kind of bar-moving at issue, even if there is no "neutral" or no overt "normal" evidential bar to invoke.

Context may also play a crucial role here. For example, peer-reviewed research may have higher evidential standards, in general, than various forms of public speech by scientific experts (section 7 returns to this point). In a broadly related spirit, Corey Dethier has recently argued that context-sensitivity explains why "some scientific conclusions seem to be appropriately asserted even though they are not known, believed, or justified on the available evidence" (2022, 1). The key point is simply that context may partly fix relevant alternative evidential standards relative to which assessment of bar-moving can be made. For example, being interviewed on national television as the director of the NIAID and a lead member of the White House Coronavirus Task Force is an importantly different context from peer-reviewed research. But *relative to this context* one can still pose counterfactual questions such as those raised above. And these context-specific questions can often be enough to establish whether (in)convenient claims are altering evidential standards.

6. Beyond Democracy, Transparency, Conventionalism, and Preference for Truth

To see more clearly how this account differs from earlier approaches, it is helpful to briefly establish that four safeguards invoked in influential analyses of the AIR do not clearly rule out altering evidential standards in view of the cost or benefit of (in)convenient hypotheses.

The first of these safeguards is requiring that, for any given hypothesis, scientists *prefer a true result over ignorance and ignorance over an error* (Wilholt 2013, 237). (Options beyond "accept," "reject," and "inconclusive" can be modeled, but are excluded here for simplicity.) This norm entails that there will be some level of confidence in any given hypothesis above which acceptance maximizes utility, and below which reporting "no result" maximizes utility. A preference for ignorance over error ensures that there will be another confidence threshold above which reporting "no result" maximizes utility and below which rejecting maximizes utility.

Crucially, the costs and benefits of a given (in)convenient hypothesis, H, can move these thresholds arbitrarily close to Pr(H) = 1 (if H is inconvenient), or to Pr(H) = 0 (if H is convenient), without violating the constraint that truth has more utility than ignorance and ignorance has more utility than error. For instance, if saving masks for healthcare workers was valuable enough, then even if the hypothesis "Masks protect the public against infection" (M) is certainly true, rejecting M may have only slightly less utility than accepting M or finding "no result." This could raise the confidence threshold below which accepting M fails to maximize expected utility arbitrarily close to Pr(M) = 1, so that accepting M would only be favored if it was 100 percent certain—which no empirical claim ever is.

The constraint that scientists prefer truth over ignorance and ignorance over error is thus compatible with noble lies, like denying a claim even when highly confident that it is true. This constraint does rule out endorsing a claim when certain it is false. But it only rules

out edge cases like this, even though noble lies in the presence of minor uncertainty are almost as criticizable.

A second widely invoked safeguard is requiring scientists to *conform to the conventional standards of a given research community* (Wilholt 2009, 2013; see also De Ridder 2022). As noted in section 4, Wilholt (2009, 99) concedes that this fails to address the issue of research communities with problematic standards. To supplement conventionalism, then, we may propose standards that research communities *should* uphold—like restricting the "cost of mistakes" in the AIR to costs that are only incurred in the case of error, not costs only incidentally tied to error.

One might object that public health sciences actually do have conventional standards that rule out downplaying the utility of masks in the way that Fauci did. Or even if public health does not, perhaps other relevant fields do—such as areas of science that are less closely tied to politics and policymaking. However, another recent incident casts doubt on this: in February 2020, 27 public health and biomedical scientists published a letter in the *Lancet,* which described claims to the effect that SARS-CoV-2 "does not have a natural origin" as "conspiracy theories" that "do nothing but create fear, rumours, and prejudice" (Calisher et al. 2020, e42). But endorsing a hypothesis like "SARS-CoV-2 accidentally leaked from the Wuhan Institute of Virology" may foster xenophobic prejudice or excessive fear about lab safety, even if it is true. So, on my view, scientists would misapply the AIR if they took this cost as a reason to raise the evidential bar for hypotheses about nonnatural origins. Yet the *Lancet* letter signatories arguably did exactly this, or at least lowered their bar for dismissing the idea of nonnatural origins as conspiratorial. (Even if claims about intentional release of the virus were unbelievable, the hypothesis of a lab accident was significantly more credible, at least initially. For instance, emails between Fauci and other scientists show that, in February 2020, various influential experts privately judged a lab accident to be a likely source of the virus (Knapton 2022).) Paired with the mask case, this incident suggests that at least some leading communities in public health and biomedical science lack conventions that rule out modifying their evidential standards for (in)convenient hypotheses.

An interesting contrast is a 2013 debate about a study which found that being overweight, but not obese, was associated with lower all-cause mortality than being "normal" weight (Flegal et al. 2013). After criticism from public health experts warning against sending "mixed messages," a *Nature* editorial came to the researchers' defense by stressing how simplified, "black-and-white" claims "tend to be easiest … to falsify" ("Shades of Grey" 2013; cf. Broadbent 2013). Further research is required to clarify the dynamics undergirding contrasts in editorial posture between different scientific venues and across different cases, like obesity research versus SARS-CoV-2 origins. One worry is that experts may express concern about simplified claims in a selective way, based on how politically (in)convenient "black-and-white" analysis is taken to be in a given case.

Two last widely invoked safeguards are *democratic accountability* and *transparency* (see, for example, Douglas 2000, 2009). But these also fail to rule out moving the evidential bar for (in)convenient claims. If a non-epistemic benefit is widely valued, after all, moving evidential standards to help realize it could be democratically endorsed. For example, a democratic majority might support a higher evidential standard for hypotheses about nonnatural origins of SARS-CoV-2, in view of potential costs of error that are not directly tied to error—like risks of geopolitical destabilization or prejudice. Nor would transparency

necessarily rule this out; it may just require figures such as the *Lancet* letter signatories and Fauci to flag more clearly when they adjust their evidential standards.

7. Manipulation, Wishful Thinking, and Abuse of Epistemic Authority

Why is it inappropriate for scientists to change their evidential standards when handling (in)convenient hypotheses in cases where this does not violate ideals like democratic accountability, transparency, preferring truth over error, or conforming to scientific conventions?

One obvious reason, not restricted to science, is that altering evidential standards to avoid costs or gain benefits that do not depend on whether a hypotheses is true is similar to *lying*. So, it may be criticizable for the same reasons that lying is. Or, even if it is not lying, modifying evidential standards for (in)convenient claims may be criticizable as *manipulation* or *wishful thinking*. For instance, when he downplayed mask efficacy to prevent hoarding, Fauci arguably manipulated the public. And either under- or overconfidence in the hypothesis that SARS-CoV-2 accidentally leaked from a lab may be politically motivated wishful thinking. So, my approach—discounting convenient benefits and inconvenient costs—helps prevent the AIR from giving license to systematic lying, manipulation, or wishful thinking in science.

If Fauci lied or manipulated the public, he would have been directing the public away from a simple intervention that he thought could realistically protect them. This likely violates basic bioethical principles like beneficence, nonmaleficence, or respect for patient autonomy. If Fauci was engaged in wishful thinking—if he genuinely believed that masks would not help the public because he *wanted* to be able to direct scarce PPE to healthcare workers without cost to the public—then he may have violated the same principles unintentionally, but still culpably.

To be clear: lying, manipulation, and wishful thinking are all distinct. John, for example, suggests that belief or acceptance is required for wishful thinking, but not for the sort of "wishful speaking" that he links to manipulation and argues better captures a variety of real-life cases (2019, 67). In this light, one might insist that Fauci more plausibly engaged in wishful *speaking* than in wishful *thinking*. However, insofar as manipulation and wishful thinking are both worth avoiding, it often may not matter much that we can decisively say, for example, whether Fauci engaged in wishful thinking versus lying or manipulation. The public record will often be too sparse for this sort of interpretive question to be settled realistically. But if Fauci's action is inappropriate regardless—because either way it stems from adjusting his evidential standards for acts of public assertion *or* private acceptance in view of the costs or benefits of (in)convenient claims—this indeterminacy does not undermine my view or its core practical upshot. (See also section 8.)

Greater transparency might seem to address these worries about lying or manipulation. In fact, however, it may just be self-defeating. If Fauci were to say, "It's uncertain whether typical masks offer significant protection," and immediately add, "I am stressing uncertainty because I want to keep you from buying masks, due to shortages," then his first statement would lose its force. Ironically, in this case, transparency could even lead to a greater loss of credibility. Or, if "transparency" meant regulatory oversight—for example, if a government committee vetted Fauci's statements—then concerns about manipulating the public could extend to this regulatory body.

A second reason not to alter evidential standards for (in)convenient hypotheses is that we are only epistemically responsible for costs of mistakes strictly as such. If potential errors about accepting a given hypothesis have costs that do not depend on whether this hypothesis is actually true, but rather just on the effect of endorsing it, then we may be responsible for them morally or politically—but not epistemically. If scientists alter evidential standards to avoid costs for which they are not epistemically responsible, this comes at the expense of avoiding other costs for which they are responsible: the cost of mistakes strictly as such. So, insofar as scientists should value epistemic responsibility more than politicians do, scientists should be more reluctant than politicians to alter evidential standards for (in)convenient hypotheses.

A final reason not to alter evidential standards for (in)convenient hypotheses is that doing so allows scientists to trade on the epistemic authority of science without paying the price that grounds this authority. It is unclear why we should look to arbitrate political disagreements by "following the science" or deferring to scientists, if they do not see conveniently modifying their evidential standards as any more costly than politicians and the general public do. This presents a choice: we could accept scientists acting just like nonscientists in this regard, in which case there would be less reason to give experts like Fauci political power based on scientific authority; or we could hold scientists to a higher epistemic bar, by expecting them not to modify evidential standards to avoid costs that are only incidentally tied to error, insofar as they act with scientific authority.

But even if altering evidential standards for (in)convenient hypotheses is inappropriate in "pure" science, perhaps it can still be appropriate in science communication or science-based policymaking? After all, Fauci and the *Lancet* letter signatories clearly acted in roles like these—whereas, for example, peer-reviewed research on mask efficacy may be less affected by politics.

Modifying evidential standards for (in)convenient claims is less inappropriate in contexts like science communication, but it is still often unacceptable. A higher-order version of the previous worry arises: if moving the evidential bar for (in)convenient hypotheses is widely accepted in science-based policymaking and science communication, then people in these fields can too easily trade on the epistemic authority of more "basic" areas of science without following the protocols that actually ground this authority. For instance, the *Lancet* letter dismissed the idea of lab accidents in a leading medical journal, and Fauci downplayed mask efficacy as the director of the NIAID. It would be naive to ignore the authority of the *Lancet* or the NIAID as bastions of basic scientific research, when considering the societal impact of these statements. Wider norms against moving the evidential bar for (in)convenient claims, extending into the science–policy interface, would be less vulnerable to systematic manipulation or miscommunication.

8. Unhealthy Conventions, Group Acceptance, and the New Demarcation Problem

Even if the distinction between *acceptance* and *assertion* can be sustained (section 3), it raises a crucial question that my view squarely addresses: is it appropriate for scientists to tailor their acts of public assertion or denial in view of the non-epistemic costs of mistaken acceptance or rejection by nonexperts who defer to their authority? On my view, the answer to this question is *Yes*—as long as "cost of error" is qualified to exclude costs that are only

incidentally tied to error. This rules out what Fauci did when he publicly downplayed masks, insofar as he raised his evidential bar for the hypothesis that asymptomatic masking by the public is effective, in view of the social cost of hoarding PPE. But my proposal still allows for behavior like applying a lower-than-normal evidential bar for Covid-19 vaccines, in view of the emergency situation.

My subsequent analysis has clarified the justificatory basis for this position. The problem with modifying evidential standards for (in)convenient claims does *not* stem from a failure to respect traditional safeguards like transparency, democratic accountability, following scientific conventions, or preferring truth over ignorance and ignorance over error. The core problem with adjusting evidential standards for (in)convenient claims is instead rooted in issues like lying, manipulation, wishful thinking, and abuse of scientific authority. In contexts where politicians or other nonexperts in the general public *do not know* that scientists are adjusting their evidential standards in view of what they take to be (in)convenient, this behavior amounts to a violation of public trust. And, to give Wilholt's conventionalist approach its due, this behavior also plausibly violates a convention governing the interaction between scientists and the public, even if not an convention internal to the practice of a scientific research community (compare Wilholt 2013, 250).

Even if politicians or members of the wider public *know* that scientists modify evidential standards for (in)convenient hypotheses, however, this may still be inappropriate—because this convention may itself be epistemologically problematic. In particular, a convention like this may result in systematic forms of *wishful thinking among nonexperts* fueled by *expert manipulation*. Political leaders and the general public may indulge in politically motivated wishful thinking, for instance, by being more receptive to experts whose views conform to their own partisan political wishes. And experts who are willing to lie or manipulate others in view of the costs of the public uptake of inconvenient claims may systematically assert exactly what partisans wish to hear from them. This is a convention, of sorts—but one that is epistemologically and politically unhealthy.

Two final points about "acceptance" in the AIR warrant mention. First, beyond the ambiguity between more or less veritistic notions of acceptance that was reviewed in section 3, one might also distinguish between *individual* and *collective* acceptance.[18] For example, perhaps the *Lancet* letter signatories collectively rejected the "accidental lab leak" hypothesis, even if many of them privately harbored greater uncertainty. In turn, the AIR could be applied to this sort of collective acceptance even when individual scientists only assert claims.

Second, to assess which notions of acceptance are (ir)relevant, it helps to step back and consider the AIR in broader context. Here there is a crucial difference between how Rudner and other early proponents viewed the AIR and how I am approaching it. For Rudner, the ultimate point of the AIR was to show that "the scientist *qua* scientist makes value judgments" (Rudner 1953). My aim is to clarify where and how, exactly, to divide legitimate and illegitimate uses of non-epistemic values in science. This has been termed the "new demarcation problem" (Holman and Wilholt 2022). And, crucially, in this context, it is less clear that the relevant attitude of *acceptance* must be narrowly epistemic—let alone that it must require veritistic belief in the truth of an accepted proposition, à la Alvin I. Goldman.

[18] For an account of which propositions to *assert* in collaborative scientific documents, see also Bright, Dang, and Heesen (2018).

For Rudner's critique, it may be necessary to focus on narrowly epistemic forms of acceptance or rejection that even an ardent supporter of the "value-free" ideal might acknowledge are "internal" to science. But delimiting what is "inside" or "outside" science is far less important, from the perspective of the new demarcation problem. My analysis has focused on controversial expert responses to Covid-19 that lie along more or less vague borders between science, public policy, and politics. But this is a strength, not a weakness. Some of the greatest non-epistemic costs of scientific error stem from errors that scientists make when acting with political power or interacting with nonexperts. In these contexts, it matters less whether a given action is "internal to science" and more that it has perceived scientific authority.

The AIR is fundamentally sound, but must be used with care. Value-laden assessments of the "cost of mistakes" should not include costs or benefits that are tied to error only incidentally—just the cost of errors strictly as such. This helps to prevent excessive politicization, while still allowing political and other values to play a basic role in setting evidential standards in science. Crucially, however, the relevant errors may include, not only the mistaken acceptance or rejection of hypotheses by scientists, but also the mistaken acceptance or rejection of these hypotheses by politicians or members of the general public who are influenced by what experts assert and deny.

Acknowledgments

For helpful feedback, I am grateful to Filipa Melo Lopes, an audience at the University of Edinburgh, two anonymous reviewers, and the editorial team at *Philosophy of Medicine*.

Disclosure Statement

No competing interest was reported by the author.

References

Adler, Jonathan E. 2002. *Belief's Own Ethics*. Cambridge, MA: MIT Press.

Anderson, Elizabeth. 2004. "Use of Value Judgments in Science: A General Argument, with Lessons from a Case Study of Feminist Research on Divorce." *Hypatia* 19, no. 1: 1–24. https://doi.org/10.1111/j.1527-2001.2004.tb01266.x.

Biddle, Justin B. 2016. "Inductive Risk, Epistemic Risk, and Overdiagnosis of Disease." *Perspectives on Science* 24, no. 2: 192–205. https://doi.org/10.1162/POSC_a_00200.

Boulicault, Marion and S. Andrew Schroeder. 2021. "Public Trust in Science: Exploring the Idiosyncrasy-Free Ideal." In *Social Trust,* edited by Kevin Vallier and Michael Weber, 102–21. New York: Routledge.

Bright, Liam Kofi, Haixin Dang, and Remco Heesen. 2018. "A Role for Judgment Aggregation in Coauthoring Scientific Papers." *Erkenntnis* 83, no. 2: 231–52. https://doi.org/10.1007/s10670-017-9887-1.

Broadbent, Alex. 2013. "Is Being Overweight Good for You?" *philosepi: Philosophy of Epidemiology, Medicine, and Public Health,* 3 June. https://philosepi.org/2013/06/03/overweight/.

Brown, Matthew J. 2020. *Science and Moral Imagination: A New Ideal for Values in Science*. Pittsburgh, PA: University of Pittsburgh Press.

Bueter, Anke. 2022. "Bias as an Epistemic Notion." *Studies in History and Philosophy of Science* 91: 307–15. https://doi.org/10.1016/j.shpsa.2021.12.002.

Calisher, Charles, Dennis Carroll, Rita Colwell, Ronald B. Corley, Peter Daszak, Christian Drosten, Luis Enjuanes, et al. 2020. "Statement in Support of the Scientists, Public Health Professionals, and Medical Professionals of China Combatting COVID-19." *The Lancet* 395, no. 10226: e42–e43. https://doi.org/10.1016/S0140-6736(20)30418-9.

CBS News. 2020a. "Full Transcript of 'Face the Nation' on March 8, 2020." https://www.cbsnews.com/news/full-transcript-of-face-the-nation-on-march-8-2020/.

———. 2020b. "Transcript: Surgeon General Jerome Adams on 'Face the Nation,' July 12, 2020." https://www.cbsnews.com/news/transcript-surgeon-general-jerome-adams-on-face-the-nation-july-12-2020/.

———. 2021. "Transcript: Dr. Anthony Fauci on 'Face the Nation,' November 28, 2021." https://www.cbsnews.com/news/transcript-dr-anthony-fauci-on-face-the-nation-november-28-2021/.

ChoGlueck, Christopher. 2022. "Still No Pill for Men? Double Standards & Demarcating Values in Biomedical Research." *Studies in History and Philosophy of Science* 91: 66–76. https://doi.org/10.1016/j.shpsa.2021.11.010.

De Ridder, Jeroen. 2022. "How to Trust a Scientist." *Studies in the History and Philosophy of Science* 93: 11–20. https://doi.org/10.1016/j.shpsa.2022.02.003.

Dethier, Corey. 2022. "Science, Assertion, and the Common Ground." *Synthese* 200, article no. 30. https://doi.org/10.1007/s11229-022-03580-7.

Douglas, Heather. 2000. "Inductive Risk and Values in Science." *Philosophy of Science* 67, no. 4: 559–79. https://doi.org/10.1086/392855.

———. 2009. *Science, Policy, and the Value-Free Ideal*. Pittsburgh, PA: University of Pittsburgh Press.

———. 2017. "Foreword." In *Exploring Inductive Risk: Case Studies of Values in Science,* edited by Kevin C. Elliott and Ted Richards, ix–xi. Oxford: Oxford University Press.

Elgin, Catherine Z. 2017. *True Enough*. Cambridge, MA: MIT Press.

Elliott, Kevin C. and Ted Richards, eds. 2017. *Exploring Inductive Risk: Case Studies of Values in Science*. Oxford: Oxford University Press.

Flegal, Katherine M., Brian K. Kit, Heather Orpana, and Barry I. Graubard. 2013. "Association of All-Cause Mortality with Overweight and Obesity Using Standard Body Mass Index Categories: A Systematic Review and Meta-Analysis." *JAMA* 309, no. 1: 71–82. https://doi.org/10.1001/jama.2012.113905.

Franco, Paul L. 2017. "Assertion, Nonepistemic Values, and Scientific Practice." *Philosophy of Science* 84, no. 1: 160–80. https://doi.org/10.1086/688939.

Gauchat, Gordon. 2012. "Politicization of Science in the Public Sphere: A Study of Public Trust in the United States, 1974 to 2010." *American Sociological Review* 77, no. 2: 167–87. https://doi.org/10.1177/0003122412438225.

Goldman, Alvin I. 1999. *Knowledge in a Social World*. Oxford: Clarendon.

Havstad, Joyce C. 2022. "Sensational Science, Archaic Hominin Genetics, and Amplified Inductive Risk." *Canadian Journal of Philosophy* 52, no. 3: 295–320. https://doi.org/10.1017/can.2021.15.

Hempel, Carl G. 1965. "Science and Human Values." In *Aspects of Scientific Explanation,* 81–96. New York: Free Press.

Holman, Bennett and Kevin C. Elliott. 2018. "The Promise and Perils of Industry-Funded Science." *Philosophy Compass* 13: e12544. https://doi.org/10.1111/phc3.12544.

Holman, Bennett and Torsten Wilholt. 2022. "The New Demarcation Problem." *Studies in History and Philosophy of Science* 91: 211–20. https://doi.org/10.1016/j.shpsa.2021.11.011.

Intemann, Kristen. 2015. "Distinguishing between Legitimate and Illegitimate Values in Climate Modeling." *European Journal for Philosophy of Science* 5: 217–32. https://doi.org/10.1007/s13194-014-0105-6.

———. 2022. "Understanding the Problem of 'Hype': Exaggeration, Values, and Trust in Science." *Canadian Journal of Philosophy* 52, no. 3: 279–94. https://doi.org/10.1017/can.2020.45.

John, Stephen. 2015. "Inductive Risk and the Contexts of Communication." *Synthese* 192: 79–96. https://doi.org/10.1007/s11229-014-0554-7.

———. 2019. "Science, Truth, and Dictatorship: Wishful Thinking or Wishful Speaking?" *Studies in History and Philosophy of Science Part A* 78: 64–72. https://doi.org/10.1016/j.shpsa.2018.12.003.

———. 2021. "Science, Politics and Regulation: The Trust-Based Approach to the Demarcation Problem." *Studies in History and Philosophy of Science Part A* 90: 1–9. https://doi.org/10.1016/j.shpsa.2021.08.006.

Knapton, Sarah. 2022. "Scientists Believed Covid Leaked from Wuhan Lab—but Feared Debate Could Hurt 'International Harmony'." *The Telegraph,* 11 January. https://www.telegraph.co.uk/news/2022/01/11/scientists-believed-covid-leaked-wuhan-lab-feared-debate-could/.

Koskinen, Inkeri and Kristina Rolin. 2022. "Distinguishing between Legitimate and Illegitimate Roles for Values in Transdisciplinary Research." *Studies in History and Philosophy of Science* 91: 191–98. https://doi.org/10.1016/j.shpsa.2021.12.001.

Lacey, Hugh. 1999. *Is Science Value-Free? Values and Scientific Understanding*. London: Routledge.

———. 2015. "'Holding' and 'Endorsing' Claims in the Course of Scientific Activities." *Studies in History and Philosophy of Science Part A* 53: 89–95. https://doi.org/10.1016/j.shpsa.2015.05.009.

Leonhardt, David. 2021. "Covid's Partisan Errors." *The New York Times,* 18 March. https://www.nytimes.com/2021/03/18/briefing/atlanta-shootings-kamala-harris-tax-deadline-2021.html.

Longino, Helen E. 1990. *Science as Social Knowledge: Values and Objectivity in Scientific Inquiry*. Princeton, NJ: Princeton University Press.

———. 1995. “Gender, Politics, and the Theoretical Virtues.” *Synthese* 104: 383–97. https://doi.org/10.1007/BF01064506.

O’Donnell, Jayne. 2020. “Top Disease Official: Risk of Coronavirus in USA is ‘Minuscule’; Skip Masks and Wash Hands.” *USA Today,* 17 February. https://www.usatoday.com/story/news/health/2020/02/17/nih-disease-official-anthony-fauci-risk-of-coronavirus-in-u-s-is-minuscule-skip-mask-and-wash-hands/4787209002/.

Portmore, Douglas W. 2007. “Consequentializing Moral Theories.” *Pacific Philosophical Quarterly* 88: 39–73. https://doi.org/10.1111/j.1468-0114.2007.00280.x.

Redding, Richard E. 2013. “Politicized Science.” *Society* 50: 439–46. https://doi.org/10.1007/s12115-013-9686-5.

Ross, Katherine. 2020. “Why Weren’t We Wearing Masks from the Beginning? Dr. Fauci Explains.” *The Street,* 12 June. https://www.thestreet.com/video/dr-fauci-masks-changing-directive-coronavirus.

Rudner, Richard. 1953. “The Scientist *Qua* Scientist Makes Value Judgments.” *Philosophy of Science* 20, no. 1: 1–6. https://doi.org/10.1086/287231.

Schorr, Isaac. 2021. “Fauci Admits Post-vaccination Masking Was about ‘Signals’ Weeks after Insisting Otherwise.” *The National Review,* 18 May. https://www.nationalreview.com/news/fauci-admits-post-vaccination-masking-was-about-signals-weeks-after-insisting-otherwise/.

“Shades of Grey.” 2013. *Nature* 497: 410. https://doi.org/10.1038/497410a.

Steel, Daniel. 2010. “Epistemic Values and the Argument from Inductive Risk.” *Philosophy of Science* 77, no. 1: 14–34. https://doi.org/10.1086/650206.

Thorp, H. Holden. 2020. “Science Has Always Been Political.” *Science* 369, no. 6501: 227. https://doi.org/10.1126/science.abd7628.

Van Fraassen, Bas C. 1980. *The Scientific Image*. Oxford: Oxford University Press.

Wagner, Wendy E. 2022. “No One Solution to the ‘New Demarcation Problem’? A View from the Trenches.” *Studies in History and Philosophy of Science* 92: 177–85. https://doi.org/10.1016/j.shpsa.2022.02.001.

Wilholt, Torsten. 2009. “Bias and Values in Scientific Research.” *Studies in History and Philosophy of Science Part A* 40: 92–101. https://doi.org/10.1016/j.shpsa.2008.12.005.

———. 2013. “Epistemic Trust in Science.” *British Journal for the Philosophy of Science* 62, no. 2: 233–53. https://doi.org/10.1093/bjps/axs007.

Winsberg, Eric, Naomi Oreskes, and Elisabeth Lloyd. 2020. “Severe Weather Event Attribution: Why Values Won’t Go Away.” *Studies in History and Philosophy of Science Part A* 84: 142–49. https://doi.org/10.1016/j.shpsa.2020.09.003.

Yan, Holly. 2020. “Top Health Officials Have Changed Their Minds about Face Mask Guidance—but for Good Reason.” *CNN,* 20 July. https://edition.cnn.com/2020/07/19/health/face-masks-us-guidance/index.html.