

EDUCATIONAL FALSEHOODS

MAXIME LEPOUTRE

University of Reading

Responsible policymaking often relies on esoteric scientific knowledge. Yet, in a democracy, policymaking also requires support from laypeople who lack this knowledge. Is it permissible for experts to communicate falsehoods about esoteric matters to bolster this public support? The standard answer is “no.” A near-consensus view, in democratic theory and philosophy of science, holds that experts should be honest. The present article challenges this position by defending a particular category of falsehoods: educational falsehoods. My argument proceeds in two parts. The first contends that, in non-ideal circumstances, communicating falsehoods can play an indispensable epistemic function, by educating the public about urgent scientific matters. The second part addresses a pressing objection: namely, that educational falsehoods are deceptive, which in turn makes them inimical to autonomy, public trust, and democratic accountability. Although this objection proves unsuccessful, I argue that it helps clarify the conditions under which deploying educational falsehoods is permissible.

1. Introduction

In October 2020, Dr. Anthony Fauci, the director of the US National Institute of Allergy and Infectious Diseases, came under fire for knowingly misrepresenting scientific facts relating to the COVID-19 pandemic. Early on, for instance, Fauci deliberately downplayed the efficacy of mask wearing. He later clarified that he had done so, in part, to avoid a mask shortage among health care workers (Roche 2021).

Is it ever permissible for experts such as Fauci publicly to communicate falsehoods—in particular, falsehoods about scientific matters—in order to promote morally good outcomes? The standard answer is “no.” Vinay Prasad, for example, responds to Fauci’s misrepresentation by asserting that “scientists and public health experts ... must always and only and indefatigably tell you the

Contact: Maxime Lepoutre <m.c.lepoutre@reading.ac.uk>

scientific truth” (Prasad 2020). In her own powerful reflections on Fauci’s case, Zeynep Pamuk echoes this commitment to honesty. “Scientists,” she declares, “should refrain from telling noble lies; all lies, really” (Pamuk 2021b). These responses are by no means unusual. On the contrary, they exemplify a widely held view in the ethics of science communication, according to which honesty is “intrinsic to science: the *sine qua non* for this form of human activity” (Keohane et al. 2014: 353).¹

I share these commentators’ qualms about Fauci’s intervention. Nevertheless, I wish to reject the broader principle they invoke to criticize it—namely, that it is always, or very nearly always, impermissible for experts to communicate scientific falsehoods. I will do so by identifying and defending a particular category of falsehoods. Specifically, I will argue that some falsehoods can perform an important epistemic function, and that, under specific conditions that are characteristic of non-ideal societies, this function makes it permissible to deploy such “educational” falsehoods.

My argument will be organized as follows. In §2, I explain more precisely what it means to communicate falsehoods. Next, I argue for my core positive claim: that, in specific non-ideal circumstances, communicating falsehoods can play an indispensable role in educating the public about urgent scientific matters (§3). I then consider and respond to a pressing objection: namely, that educational falsehoods are deceptive, which in turn makes them injurious to personal autonomy, corrosive of trust in science, and inimical to democratic accountability (§4). I briefly conclude in §5.

Three clarifications are needed before proceeding. First, I will focus on educational falsehoods pertaining to *scientific* matters. This is because science is the domain where the epistemic value of falsehoods, and thus the need for such falsehoods, seems clearest. But my point may have broader applicability. If so, then it may also yield a qualified defense of deploying educational falsehoods regarding, say, political or historical matters.

Second, I am assuming that a key aim of public science communication is educational. In other words, science communication aims, at least in part, to improve the epistemic position of its audience, and thus, the audience’s capacity to make informed decisions, with respect to the scientific matters being discussed (Slovic 1986; Pielke 2007: 1–3; John 2018: 83).² This, as will be explained in §3.4, does not necessarily mean that science communication is wholly “value-free.” Nor does it necessarily entail that education is the *sole* aim of science communication (Pielke 2007: 7). But I nonetheless consider it to be one of its core functions.

1. See also de Melo-Martín & Intemann (2018: 43), Veit et al. (2021: 22–24), and Schroeder (2022: 36–37). Director (2023: 962) acknowledges that dishonesty from public officials could *in principle* be permissible, but ultimately insists that, in practice, it is “almost always” impermissible.

2. For discussion of education’s diverse aims, see Watson (2016).

Finally, it is important to note that my proposal is not without philosophical precedent. Stephen John (2018), in particular, has influentially suggested that scientific dishonesty can sometimes perform an epistemically useful function. I am sympathetic to John's argument and will therefore be building on his suggestion. But it nevertheless remains incomplete with respect to three important issues: first, the nature of the mechanisms connecting falsehoods to educational outcomes; second, the specific social conditions in which it is permissible to deploy epistemically useful falsehoods; and third, whether educational falsehoods violate key values such as personal autonomy and democracy. The account developed in what follows seeks to remedy these limitations—and, by implication, it aims a) to yield a more comprehensive understanding of how deploying misrepresentations can facilitate epistemic goods,³ b) to provide guidance for science communicators regarding the circumstances in which they may appropriately use falsehoods,⁴ and c) to show how educational falsehoods can be reconciled with (among other things) respect for personal autonomy and the demands of democratic accountability.⁵

2. Communicating Falsehoods

What does it mean to communicate a falsehood? A falsehood, as I am using this term, is something that misrepresents something else. On this conception, a falsehood represents something (its target), and that representation is in some way inaccurate.⁶ Thus defined, a falsehood can come in different forms. It can be propositional (e.g., the false proposition that the Earth is flat). But this needn't be the case. It could, alternatively, take a visual form (e.g., a picture of the Earth that represents it as flat).

There is an important sense in which, given the complexity of scientific matters, misrepresentations are inescapable in science. Scientists often disagree about particular scientific claims (e.g., whether masks are effective). And given the incompleteness of scientific evidence, even well-supported predictions might turn out to be inaccurate. So, there are many cases where it is simply not straightforward whether a claim is accurate or not—and where, as a result, even

3. See especially §3.1.

4. §3.3.

5. §4.2, §4.4.

6. This characterization is inspired by Elgin's (2017) influential discussion of "felicitous falsehoods." Elgin characterizes falsehoods as "inaccurate representations," and *felicitous* falsehoods as "inaccurate representations whose inaccuracy does not undermine [their] epistemic function" (2017: 3, 5, 23). For similar uses of the term "falsehood," see also Rancourt (2017: 390) and Potochnik (2017: ix).

scientists who intend to say something accurate end up communicating claims that misrepresent reality.

My focus, however, will be on a different set of cases. I bracket cases, such as those mentioned above, where scientists mistakenly believe that they are saying something true, and thus accidentally end up saying something false. Instead, I will focus on falsehoods that are communicated *intentionally*. In other words, I will be concerned with situations where scientists know, and where the scientific community may well agree, that what they are communicating is in some meaningful way inaccurate.⁷

Yet even intentional falsehoods might seem inevitable. Scientists are very often in a situation where they must communicate something they know is not strictly accurate. This might be, for instance, because characterizing a phenomenon with what they take to be maximal accuracy is simply impossible given reasonable time constraints, or because, when representing phenomena, any model must be inaccurate on at least some dimensions. So, without further qualification, one might think that almost all scientific statements will count as intentional falsehoods, and, by implication, that it is hard to say what *wouldn't* count as a falsehood.

But even if scientists often have no realistic option other than to communicate something that they know is strictly speaking inaccurate, they sometimes *choose* a more inaccurate representation even though a less inaccurate representation is readily available. These are the kinds of cases I will be concerned with when discussing scientific falsehoods: where scientists know that something constitutes an inaccurate representation, and can readily communicate something that is either accurate or less inaccurate, but opt for the more inaccurate representation instead.

Another important set of questions concerns *how* such misrepresentations are communicated. One can communicate something one takes to be false in direct or indirect ways. For example, one might simply assert the falsehood that the Earth is flat. But one can also communicate the same falsehood more indirectly, through presupposition or conversational implicature. For instance, by saying, in a conspiratorial tone, “They *want* you to believe the Earth is round,” one might imply, without asserting it, that the Earth is flat.

This distinction could conceivably influence the permissibility of educational falsehoods. It is often thought that indirectly communicating a falsehood is less morally problematic than directly doing so. This is reflected in the common intu-

7. I distinguish here between accidental and intentional false claims. Another possibility, which I do not touch on here, is that scientists might intend to say something false (e.g., about face masks), and, out of ignorance, accidentally say something true instead. Such “accidental truths” raise some of the same moral concerns as “intentional falsehoods” (insofar as both can involve deceptive *intent*) but they arguably do not raise *all* of the same concerns (notably, because the inducement of false beliefs can impair autonomy in distinctive ways, to be discussed in §4.2).

ition that lying is morally worse than intentionally misleading (Berstler 2019). But my argument will not depend on this intuition: that is, my defense of educational falsehoods is intended to apply even to assertions of such falsehoods.

Moreover, one can communicate falsehoods in a manner that is transparent or non-transparent. A falsehood is communicated transparently when the speaker intentionally deploys it in a way, or in a context, that makes it plain that it is not true. If, for example, I preface a misrepresentation by saying “this is a gross simplification,” or if it is common knowledge that I am simplifying matters, then the falsehood in question is communicated transparently.⁸ It is communicated non-transparently, in contrast, when the speaker is not open about its falsity—and so, when they intentionally deploy the falsehood in a way, or in a context, that conceals its falsity from the audience.

This distinction is critically important. In defending educational falsehoods, my aim, as stated in §1, is to respond to proponents of absolute or near-absolute scientific honesty. But there is intuitively nothing dishonest about communicating a known misrepresentation transparently. And, relatedly, one of the principal morally worrying features of dishonesty—that it tends to be deceptive—seems inapplicable to transparently communicated falsehoods.⁹ Accordingly, to avoid talking past defenders of absolute or near-absolute honesty, my defence will need to show that some falsehoods can perform a valuable epistemic function, and can be morally permissible, despite being communicated *non-transparently*. This constitutes a difficulty because existing discussions of the positive epistemic functions associated with falsehoods, which I introduce in §3.1, have often focused on what are plausibly *transparent* falsehoods. Part of the challenge, to which I return in §3.2, will therefore be to show that these epistemic functions can carry over to non-transparent misrepresentations.

It also matters, for related reasons, what *audience* falsehoods are being communicated to. As Corey Dethier (2022) has shown in his recent discussion of scientific assertion, the same utterance might be received quite differently by an audience of experts (given their shared background knowledge and assump-

8. For discussion, outside of the scientific context, of settings where expectations of truthfulness are “suspended” (e.g., fiction, theatre), such that false claims are transparently false, see Fallis (2009: 33–37) and Shiffrin (2014: 16–19).

9. On the connection between dishonest and deceptive communication, see, e.g., Director (2023: 951, fn 1, fn 4). Note that, for Director, dishonesty and deceptiveness do not necessarily require a very large departure from the truth. The scientific honesty he defends involves telling the “full truth” and, accordingly, he includes “partial truths” as well as “spin”—which can involve reasonably minor deviations from the truth, or simply “bending” the truth—among the forms of communication that can qualify as deceptive, and thus be impermissible (2023: 1, fn 23). This suggests that even reasonably small departures from the truth can count as deceptive, provided they are intended to induce a false belief. I discuss some such cases, and offer characterizations of them as deceptive, in §4.1.

tions) compared to an audience of laypeople (who may lack this shared background). This point notably applies to the issue of transparency. The same misrepresentation might be, and might reasonably be expected by a speaker to be, transparently false to scientific experts (e.g., because these experts are generally aware that, as will be discussed shortly, scientists often use misrepresentations when modelling a target object) but not to non-experts (who may not be aware of this practice, and may therefore not realize that the misrepresentation in question is a misrepresentation). What this suggests, more generally, is that the epistemic effects and moral desirability of communicating scientific falsehoods can vary meaningfully depending on the audience (Dethier 2022: 12–13, 17). In what follows, my focus will be on falsehoods that are communicated to *non-experts* (e.g., the general public and policymakers).

3. Can Falsehoods Educate?

The present section makes the case that communicating falsehoods can perform a valuable educational function (§3.1). I then stave off three worries concerning this function: that it can only be performed by “transparent” falsehoods (§3.2); that it is unnecessary (§3.3); and, finally, that it requires scientists to make value judgements that it is inappropriate for them to make (§3.4).

3.1. Three Mechanisms

There are at least three mechanisms connecting intentional misrepresentations to educational or epistemic gains. The first of these, which is familiar from the practice of scientific modelling, is *simplification*. Filtering out features of the target (the thing that is being represented) can help highlight some of its other features.¹⁰ This is particularly applicable when the target is complex. A representation that strives to represent all features of a complex target might overwhelm us with information, or distract us from properties of the target that are relevant to the matter at hand. Accordingly, a simplified representation can make important properties of the target more salient than they would otherwise be.¹¹

10. For extensive discussion of simplifications, alongside idealizations and approximations more generally, in models, see Rancourt (2017: 390), Strevens (2016: 38–39), Elgin (2017: 23–32), and Potochnik (2017: §2.2), all of whom characterize these as a form of falsehood.

11. Is communicating a simplified representation really *communicating a falsehood*? One might think that it rather involves *not communicating a truth*. In fact, there are two ways this practice can constitute the communication of a falsehood, as defined in §2. First, omitting a feature when characterizing a target can *implicate* the falsehood that the target does not possess this feature. But this practice can also communicate falsehoods in a more direct way, for example when a simplified

Consider how this might apply to climate communication. Imagine a simple visual model of global temperatures over time. At any given time, this model represents the Earth as one color, corresponding to the average global temperature. This representation clearly simplifies its target. It removes, among many other things, information about regional differences in temperature. And, consequently, it misrepresents regional temperatures, with some regions represented as colder or hotter than they really are. But this misrepresentation yields a countervailing epistemic benefit: it helps highlight the size and rate of increase in global average temperatures. Regional differences—and notably the fact that some regions have grown colder while others have grown hotter—may otherwise make these properties less visible.

The second process whereby falsehoods can facilitate epistemic access is *exaggeration*. We can make a feature of the target more visible, not just by filtering out some of its other features, but also by exaggerating the feature in question—that is, by distorting it in a way that makes it more striking or visible.

Exaggeration is commonly deployed in the context of scientific consensus statements. Consensus statements aim to convey that the scientific community agrees on a particular proposition. Such statements are widely thought to play an important role in convincing the public that particular scientific findings have strong evidential support (Bayes et al. 2020: 2). This is, in part, because disagreement among scientists tends to reduce their credibility in the eyes of the public.¹² Hence, consensus statements have long been a central part of science communication, particularly in the context of climate communication (e.g., “97 percent of scientists agree that anthropogenic climate change is real”) (Bayes et al. 2020: 2).

But getting the public to recognize the existence of a broad scientific consensus on climate change is challenging. According to Karen Kovaka (2021: 2368), this is partly because climate skeptic media tend to amplify what dissent there is, and partly because—independently of this amplification—laypeople often lend excessive credibility to scientists who dissent from the mainstream.¹³ The public therefore typically vastly underestimates the level of scientific consensus over climate change (Duffy 2022).

representation is asserted. The case of the simplified global temperature model I offer below can be interpreted along these lines. We can imagine a policy advisor asserting that the model represents temperatures across the planet over the last thousand years. Yet as discussed below, because of its simplifications, the model partly misrepresents these temperatures.

12. Chinn & Hart (2022: 122); Gustafson & Rice (2020: 621). See fn 48 for further discussion of the empirical evidence. Beatty and Moore (2010) persuasively argue that disagreement is not necessarily indicative of bad science. But this is consistent with people believing that it is.

13. Note that Kovaka herself does not use this observation to argue in favor of educational falsehoods. She instead recommends correcting misconceptions about the nature of science (2021: 2368–2369). I return to this alternative proposal in §3.3.

To counteract this phenomenon, climate communicators often exaggerate the extent of this consensus through the practice of “masking disagreement”. On this approach, members of scientific panels who disagree with a particular claim agree to refrain from expressing their disagreement, and to let this claim stand as the group’s position, so that the group can speak with one voice.¹⁴ Take, for example, a scientific panel that disagrees on the likely extent of climate-induced sea level rises. A member whose estimate (Y) is far lower or higher than their colleagues’ estimates (which cluster closely around X) might decide to let the much more widespread estimate stand as the group’s position. This allows the group to state a single “consensus” position.¹⁵ Strictly speaking, this needn’t involve asserting the falsehood that everyone agrees on X . For example, the panel might simply say “It is the panel’s position that climate change will cause sea levels to rise by approximately X ” *without* explicitly saying that every panel member agrees with the panel’s position. Nonetheless, the practice *is* plausibly designed to implicate a misrepresentation. The panel could readily declare their actual level of disagreement. By masking their disagreement instead, they intentionally make it seem as though there is a higher level of consensus than is really the case.

This exaggeration, in turn, can help secure important epistemic benefits.¹⁶ Given the common perception that the existence of disagreement is a sign of bad science, presenting an artificially united front may help the public appreciate, fittingly, that climate scientists are credible sources on matters surrounding climate change (Beatty 2006: 54). And, more specifically, it may encourage the true belief that the object of the apparent consensus (e.g., that climate change will cause sea levels to rise by approximately X) is strongly supported by the evidence.¹⁷

The final epistemic mechanism I wish to outline here concerns *familiarity*. This mechanism builds on the previous two. Simplifying or exaggerating falsehoods matter, not simply because they allow us to filter out distracting features,

14. For analysis of this practice, see Beatty (2006: 52–53), John (2018: 79–80), and de Melo-Martín & Intemann (2018: 82–83).

15. This is loosely inspired by a real case involving the IPCC, discussed in Schroeder (2022: 49–51).

16. Some worry that masking disagreement nevertheless comes with countervailing epistemic costs: first, it hides the true extent of disagreement; and second, suppressing disagreement could lead scientists to dismiss fruitful avenues of research. See Beatty & Moore (2010: 202–204) and de Melo-Martín & Intemann (2018: 82–83). The second worry is too quick. Scientists can mask disagreement when communicating their results, while continuing to consider and debate dissenting results when conducting research. The first worry is better placed. But my contention is that, in at least some cases, this epistemic cost is overridden by the epistemic benefits outlined above. See §3.2 for discussion of this condition.

17. The fact that some scientists dissented from p , but nonetheless agreed to let it stand, does not mean that p lacks strong evidentiary support: as John (2018: 79–80) notes, one of the reasons why scientists agree to let claims they disagree with stand as the group’s position *just is* that they recognize that the claim nonetheless satisfies a high evidentiary standard.

or to enlarge features we wish to highlight, but also because, in doing so, they can help present information in a more familiar—and therefore, more cognitively accessible—way.

There are two notable ways this can happen. To begin, falsehoods can help package information in a familiar *format*. A key example of this relates to narrative. Suppose we wish to explain how climate research operates. We can present this information in different formats. We might, for instance, enumerate facts relating to the process of climate research. But we might instead choose to communicate this process in narrative (or story) form. Opting for the latter option invariably involves some measure of falsification. A story has a standard structure: prototypical stories involve one or more protagonists, who encounter a meaningful problem, and who then confront that problem, leading to some form of resolution (Fraser 2021). Making information fit this structure almost always involves simplifications and exaggerations. For example, identifying clear protagonists might require focusing disproportionately on the contributions of one or two scientists or scientific groups. Moreover, to dramatize the problem, communicators might disproportionately emphasize the obstruction of climate research by a few states or corporations, while omitting details of the challenging peer review process (Flottum & Gerjstad 2017).

Such simplifications and distortions help improve the audience's understanding of the climate research process partly for reasons already mentioned: they help filter out material that may have overwhelmed or distracted us; and they enlarge important pieces of information we may otherwise have overlooked. But the crucial point here is that this is not the only reason they improve epistemic access. The further reason is that the story format resulting from these simplifications and exaggerations is cognitively familiar. As social psychologists have argued, we therefore often find it easier to process and remember information when it is presented in narrative form (Mandler & Johnson 1977; Fraser 2021).¹⁸

So far, the point has been that falsehoods can help present informational content about the target in a familiar format. But falsehoods can also contribute to making the target familiar—and so, more epistemically accessible—by likening it to some other, more familiar, object. The paradigmatic example of this is metaphor. Metaphor involves characterizing one thing (the target) in terms of another (the source). Hence, metaphors help us understand unfamiliar things by drawing on what we know about more familiar things. As Therese Asplund

18. To see that familiarity makes an *added* epistemic contribution, consider that we could provide the same simplified and exaggerated information in a less familiar format (e.g., as an unordered list of claims, or in reverse chronological order). Although the information would be equally simplified and exaggerated, the lack of familiarity would arguably make it less epistemically accessible.

(2011) has shown, for instance, climate change is often publicly explained using metaphors that liken it to a greenhouse or to war.

How does this relate to falsehoods? Metaphorical representations of a target characteristically require some measure of simplification and distortion. For one thing, metaphors are always partial, in that the source helps understand *some* parts of the target, while hiding or concealing other parts. For instance, the greenhouse gas metaphor for climate change highlights some features of climate change (e.g., the fact that heat from the Sun is trapped, and that this engenders temperature increases). But it simultaneously conceals other aspects of the process (e.g., oceans' role in thermal transfers) (Asplund 2011: 3–4; Chen 2012: 109–110). As for the parts of the target that *are* meant to be represented by the source, metaphors often exaggerate the similarity between the two, and thus yield a partially distorted representation of the target. Although the greenhouse metaphor highlights the causal connection between trapped sunlight and global temperature increases, it also invites us to see this connection as much more rapid than it actually is (Chen 2012).

Let us take stock. I have introduced three mechanisms through which the communication of falsehoods can facilitate epistemic access to a target. Falsehoods can simplify a target, by removing some of its features to make other features more visible. They can exaggerate, and thereby highlight, features of the target. And they can help present the target in a more familiar way, either by packaging information about it in a familiar format, or by likening it to something familiar.

Now, even though these processes are conceptually distinct, in practice they are often entangled with one another. Indeed, we have already seen that presenting a target in a more familiar light generally requires simplifying some of its features, and exaggerating others. To further exemplify these processes, and to illustrate their possible entanglement, I therefore wish to conclude this section by offering a final example in which all three combine to produce an epistemically felicitous outcome.

In the 1980s, atmospheric scientists became increasingly concerned about the growing depletion of the ozone layer, which protects the Earth from dangerous ultraviolet radiation. These concerns intensified in 1985 with the discovery of an especially thin area above the Antarctic, which came to be known as the ozone “hole.” The discovery prompted a rapid international response, culminating in the accelerated phaseout of depletion-accelerating chemicals (Ungar 1998; Grevsmuhl 2018).

In his fascinating historical exploration of this episode, Sebastian Grevsmuhl observes that the depleted area was not initially referred to or represented as a “hole.” Because, strictly speaking, what had been observed was a thinning or

decrease of the ozone above the Antarctic (but not a complete lack), references to a “hole” were considered somewhat misleading and “metaphorical.” Accordingly, “one referee of [Richard Stolarski’s groundbreaking scientific article on ozone depletion] objected to the use of the ‘hole’ metaphor in the title of the paper,” so that the more accurate term “decrease” was eventually preferred by its authors (2014: 46).¹⁹ This was mirrored in visual representations. The key findings were initially represented in a simple graph showing ozone units above the Antarctic decreasing over time. This graph, Grevsmuhl notes, represented the findings “very efficiently,” but it “clearly could not make the case for a ‘hole’ in the ozone layer” (2018: 78).²⁰

Yet, though it was considered less accurate, the idea of an ozone “hole” soon gained currency in scientific discourse and, Grevsmuhl argues, ultimately played a key role in successfully educating the public about the real threat posed by ozone depletion. Importantly, for our purposes, a number of intentional misrepresentations featured in the construction of this idea. To see this, consider the famous NASA satellite picture of the ozone hole (Figure 1).

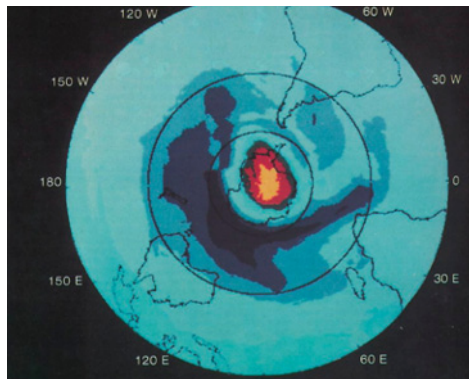


Figure 1: NASA Satellite Image of the Ozone Hole (Grevsmuhl 2018).

This visual representation was importantly *simplified*. As Grevsmuhl explains, the image “homogeni[zes] a substantial number of satellite measurements by correlating a certain data interval with a specific color” (2018: 78). Put differently,

19. Grevsmuhl (2014: 46) further emphasizes that “one definitely has to speak of a metaphor since stratospheric ozone is never removed completely.” The misleading quality of the term “hole” is also acknowledged by the Environmental Protection Agency (EPA) (2023), which, today, clarifies that the ozone “hole” is “not really a hole through the ozone layer, but rather a large area of the stratosphere with extremely low amounts of ozone.”

20. See also Grevsmuhl (2014: 45) for a detailed analysis of the way false-color imaging was used, not only to homogenize a number of measurements, but also to create “the illusion of continuous measurement.”

the picture color-codes measurements of the ozone layer in a way that lumps together large measurement intervals. Thus, it filters out substantial amounts of information about the different thickness of different parts of the ozone.

This simplification in turn leads to an *exaggeration*. It results in—and indeed, was designed to result in—an artificially clear contrast between the comparatively depleted area situated over Antarctica, and the rest of the ozone layer. By exaggerating this discontinuity between the two areas, the image “convey[s] the erroneous impression of a discrete hole in the atmosphere over the South Pole” (Mazur & Lee 1993: 711).²¹

Finally, artificially creating the appearance of an actual hole—and metaphorically referring to the depleted area as such—matters because it helps present the problem in terms of something *familiar*. To most people, the idea of thinning layers of ozone in the stratosphere is unfamiliar, and somewhat abstract. By contrast, the image of a hole in a shield—and even of a hole in a shield protecting us from lethal rays—is a familiar one. As Sheldon Ungar (1998: 523) explains, this image “meshes nicely with abiding and resonant cultural motifs. These Hollywood affinities range from the shields on *Starship Enterprise* to *Star Wars*.” So, the construction of the ozone “hole” helped convey, in familiar terms, the severity and urgency of the problem posed by ozone depletion.

The ozone case thus exemplifies how the three falsehood-involving epistemic mechanisms I have outlined can come together fruitfully. Misrepresentations of a target can simplify it, exaggerate it, present it in a familiar light—and thus, improve epistemic access to it. In the rest of §3, I wish to introduce several worries with, and corresponding qualifications of, this epistemic defense of falsehoods.

3.2. *Must the Falsehoods be Recognized as Falsehoods?*

I have argued that deploying falsehoods (understood as intentional misrepresentations) about a target can facilitate epistemic access to this target. But my argument in §3.1 might seem to imply that the epistemic value of falsehoods depends on their being transparently false to, and thus recognized as false by, the audience.

For one thing, my discussion of simplification is directly inspired by the practice of scientific modelling. But scientists generally know, and know that other scientists know, that scientific models are simplified. Philosophical discus-

21. See also Ungar (1998: 519), who observes that “the ozone ‘hole’ was an exaggeration ... satellite pictures were doctored and colored to make them more graphic.”

sions of scientific modelling are often explicit about this. Catherine Elgin, whose influential analysis of epistemically felicitous falsehoods I am indebted to, is clear that scientists are generally aware that the misrepresentations embedded in their models do not “purport to be true.” As a result, scientists typically do not *believe* these falsehoods (2017: 3, 23). One might think that this shared background knowledge, in the scientific community, is a key reason why the simplifications and other misrepresentations that feature in scientific models are capable of generating epistemic benefits. What is more, some of my examples in §3.1 were underspecified, such that we could readily imagine the public knowing that the misrepresentations in question are misrepresentations. For instance, the public *know*, given their background knowledge, that the greenhouse metaphor for climate change is a metaphor. Hence, they know that it does not purport to represent climate change with full accuracy.

More generally, then, one might worry that the positive epistemic functions associated with falsehoods *exclusively* apply to intentional misrepresentations that are communicated transparently, such that they are recognized as false by the audience. As explained in §2, this conclusion is unacceptable for my purposes. I aim to challenge an absolute or near-absolute commitment to scientific honesty. But falsehoods that are communicated transparently are plausibly neither dishonest nor deceptive. Accordingly, I need to show that the epistemic benefits discussed above can also be generated by misrepresentations that are communicated non-transparently, and whose falsity is therefore not recognized by the audience.

To argue for this claim, one might be tempted to appeal to ordinary pedagogical practice. Secondary school science teachers often teach theories that are strictly speaking false. As Sindhuja Bhakthavatsalam observes:

Today, we consider Newtonian mechanics to be false in light of Einstein’s theories of gravitation and quantum mechanics; we consider Rutherford’s and Bohr’s planetary models of the atom to be false in light of quantum mechanical models; ... But these continue to be taught in K-12 science. In fact, it is not until late in high school or in many cases only in college that accepted-as-true — and often more sophisticated — theories are taught[.] (2019: 6)

This, Bhakthavatsalam goes on to argue, is not an accident. Teaching false theories, even when more sophisticated theories are available, can yield important educational benefits, not least by improving students’ understanding of the world (2019: 11–20). For example, though Newtonian mechanics may be false, accepting Newton’s laws of motion can help students better grasp how two bod-

ies attract one another, and make accurate predictions about their movement, in immensely many cases.²² Now, if teaching false theories is educationally valuable, one might conclude, by analogy, that non-transparently communicating scientific falsehoods can be as well.

On closer inspection, however, this first argument runs into difficulties. Notice, first of all, that many schoolteachers teach false theories *as false*.²³ For example, they might preface the teaching of Newtonian mechanics by noting that it has been superseded by Einstein's theory of general relativity, which rejects some of Newtonian mechanics' theoretical claims (e.g., about the nature of space). Second, even when teachers do not preface false theories in this way, they often intend to reveal that these theories are false in due course—for example, when explaining, subsequently, why a new theory must be introduced.²⁴ Finally, and partly as a result of these first two practices, students might come to *expect* that the theories taught in school are not always strictly true. Consequently, even when a teacher does not explain that a theory is false, before *or* after it is taught, the educational context might make it reasonable for students to doubt that this theory is strictly true.

The point is that the falsehoods deployed by teachers are often either intended to be transparent to students (when the teacher explicitly discloses their falsity) or at least foreseeably transparent to students (when, for contextual reasons, students can reasonably expect that theories they are taught are not strictly true).²⁵

22. Here, I am assuming, following Slater (2008), Rancourt (2017: 390–391), Elgin (2017: 60–61), and Bhatkhavatsalam (2019: 6), that Newtonian mechanics is false. For extensive discussion of this point, see Slater (2008: 533–539), who also problematizes the claim that Newtonian mechanics is even *approximately* true. Bhatkhavatsalam's approach (2019: 8–10) is slightly more concessive: she suggests that Newtonian mechanics may be empirically adequate (such that most of its claims about observable phenomena within its domain are true) but nonetheless insists that many of its theoretical claims are strictly speaking false. Note, finally, that even if one remained unconvinced that Newtonian mechanics is false in any meaningful sense, Slater (2008: 533, 537, 539), Elgin (2017: 29), and Bhatkhavatsalam (2019: 9–10) all suggest that there are *other* examples of false taught theories, which they believe are not even approximately true (e.g., the Hardy-Weinberg theory of population genetics; Bohr's theory of the hydrogen atom).

23. Slater (2008: 541) explicitly recommends this approach to teaching false theories.

24. In a similar spirit, Bhatkhavatsalam denies that a theory's falsity must be foregrounded by teachers (2019: 7) but suggests that teachers should eventually explain that such theories are false (12, 14).

25. A further possible disanalogy concerns the respective aims of school teaching and public science communication. One might think that public science communication is distinctive in that it aims, at least in part, to ensure that citizens can engage in informed political deliberation on issues to which the scientific matters at hand are relevant (which, in turn, may require fostering distinctive epistemic capacities). However, I do not believe that this is specific to public science communication. According to many philosophers of education, one of the purposes of education—including school science education—*just is* to prepare future citizens for competent democratic participation. And this democratic competence is often taken to include the capacity to deliberate with others on matters of public concern. See, on this point, Brighouse (2009: 40), Elgin

Accordingly, the pervasiveness of falsehoods in ordinary teaching contexts does not obviously show that non-transparent falsehoods (and, relatedly, dishonesty) can be epistemically valuable.

Even so, there is another, more direct, reason for thinking that non-transparent falsehoods can be epistemically valuable. Indeed, we can appreciate why this is the case by attending more closely to the mechanisms outlined in §3.1. Upon closer examination, none of these mechanisms depends on whether the falsehoods are recognized as false. Rather, they depend on whether, as a result of removing, or exaggerating, or otherwise misrepresenting (parts of) the target, some of the target's features appear more visible or more familiar than they otherwise would.

To illustrate, suppose for the sake of argument that I do not know that the picture of the ozone hole described in §3.1 homogenizes many of the ozone layers, and that, consequently, it exaggerates the boundary between parts of the ozone situated above the Antarctic, and the rest of the ozone. Even if this is the case, these simplifications, and the resulting exaggeration, will still make it easier for me to see that an area of the stratospheric ozone is especially depleted, and to identify that area. Relatedly, even if I do not know that references to an ozone "hole" are metaphorical, this metaphor can still improve my epistemic access to an important fact—namely, that there is a critical weakness in the protective screen safeguarding us from radiation.²⁶

An important qualification is nonetheless needed here. Even if non-transparent falsehoods can yield epistemic benefits, they also come at an epistemic cost. The person who believes that there really is a hole in the ozone (as opposed to simply a substantially thinner area) represents the world less accurately than someone who knows that the "hole" is a metaphor, which highlights, but also exaggerates, a real phenomenon. Similarly, the person who believes that the consensus on sea-level rises is greater than it really is believes something false, which constitutes an epistemic cost, even if that false belief also facilitates true beliefs relating to climate change.

The implication is that, when deciding whether to communicate a (non-transparent) educational falsehood, we must weigh its epistemic benefits against its epistemic costs. In some cases, the benefits outweigh the costs. For instance, making it seem as though there is a literal hole in the ozone may be overall epistemically justified if this exaggeration successfully highlights the existence of a substantially depleted area. In other cases, however, misrepresentations conceal

(2020: 72–73), Nussbaum (2009: 55–56), and—in the context of *science* education—Ratcliffe & Grace (2003: 38) and Slater (2008: 530). So, the present aim alone cannot explain why falsehoods would be more problematic in public science communication than in schools.

26. Rancourt (2017) likewise suggests that falsehoods can facilitate epistemic gains even if they are *believed*.

more than they reveal (Moser & Dilling 2004: 37). My defense of communicating falsehoods does not extend to these cases. Thus it is qualified, first of all, in that it applies exclusively to falsehoods whose epistemic benefits are reasonably expected to outweigh their epistemic costs.

This first qualification assumes that it is *possible* for science communicators to weigh these epistemic benefits and costs against each other. This may seem controversial. Yet I believe this assumption is warranted. To see why, consider that similar assessments are commonplace in both scientific research and pedagogical practice. It is widely acknowledged that scientific research pursues, not just any truths, but “*significant truths*”—for, as Philip Kitcher influentially observes, “there are vast numbers of true statements [e.g., about the number of blades of grass in London] it would be utterly pointless to ascertain” (2001: 65, emphasis added; see also Elgin 2017: 10–11, 82–83). Bhakthavatsalam applies this insight to science teaching: “it is not the case,” she insists, “that we teach any and all truths—we want to weed out the trivial, the obvious, the tautological, and so on” (2019: 6; see also Slater 2008: 528). So, both scientific researchers and teachers recognize, in their respective practices, that some true propositions are more significant than others, such that it is more important to believe them than others.

How do practitioners assess the significance of different true propositions? Practical considerations undoubtedly play a role. We research and teach about greenhouse gases or the ozone hole, in part, because there is a lot at stake (Kitcher 2001: 65; Slater 2008: 528). But judgments of significance are also guided, at least in part, by considerations of comparative *epistemic* significance. There are different ways of making such epistemic assessments. One common approach appeals to explanatory power. Elgin (2017: 58, 82–84), for example, observes that some propositions are “central” to our understanding of particular subject matters—by which she means, notably, that they help explain a wide range of phenomena within this subject matter—whilst others are comparatively peripheral to our understanding—such that comparatively few features of this subject matter depend, for their explanation, on this proposition.²⁷ For instance, one might think that the true proposition *that the ozone layer is substantially depleted* is needed to explain various biological or environmental phenomena (e.g., increased risks of skin cancer, decreases in phytoplankton production) (EPA 2023), but that the true proposition *that ozone depletion consists in thinning layers rather than a literal hole* is not.

I am not suggesting that such assessments of comparative epistemic significance are easy to make. Nor do I mean to suggest that judgments of overall epistemic significance can be wholly divorced from practical considerations (for

27. Strevens (2016: 38) focuses, relatedly, on the idea of explanatory *difference-makers*: Some true beliefs make a difference to explaining that something happened, whereas others are “explanatorily irrelevant.” For detailed discussion of epistemic significance, which acknowledges that epistemic significance may depend on context-specific factors (e.g., what questions we are interested in answering), see Kitcher (2001: 63–82).

reasons to be discussed in §3.4). But the important point is that they are routinely performed by scientific researchers and teachers. And so, from the fact that my defense of educational falsehoods requires science communicators to make such judgments, we should not immediately conclude that it is overly demanding.

3.3. *Are Falsehoods Necessary?*

But even with this first qualification, the fact that non-transparent educational falsehoods come at an epistemic cost might raise another question: namely, are these falsehoods really necessary? Put differently, would it be possible to help the audience achieve the same epistemic benefits *without* incurring these epistemic costs?

Take the case of consensus statements. Instead of masking scientific disagreement about p (“climate change will cause sea-levels to rise by approximately X ”), one might provide accurate information about the extent of disagreement, while explaining that disagreement is not a sign of bad science, and that, despite residual disagreement, p is well supported by the evidence (see, for a similar suggestion, Pamuk 2021a: 84–85; Moore & MacKenzie 2020: 1–8; and Kovaka 2021: 2368–2369). Such an approach might seem to achieve the best of both worlds, epistemically speaking. It retains the epistemic benefits associated with masking disagreement (by helping the audience see that climate scientists and their conclusions are credible) while avoiding its epistemic cost (the concealment of disagreement).

This question of necessity matters. If there *is* an alternative way of achieving the epistemic benefits associated with (non-transparent) educational falsehoods, which avoids their epistemic costs, then, other things being equal, we should forego communicating such falsehoods.²⁸

Yet communicating non-transparent falsehoods is in fact sometimes epistemically necessary. Whether or not we can achieve certain epistemic benefits without deploying such falsehoods depends significantly on context. More specifically, it depends on at least four contextual factors. First, how complex is the matter at hand? The less complex the matter is, the more epistemically accessible it is—and so, the less necessary it is to misrepresent it so as to enhance its simplicity, exaggerate its features, or render it more familiar. Second, how much background knowledge does the public have?²⁹ Even if a matter is complex, the public might already know a significant amount about it, thus making it easier

28. Director (2023: 962) takes this “necessity” requirement to constitute an important consideration against dishonesty.

29. For the broader point that the appropriateness of particular scientific assertions depends importantly on the audience’s common background knowledge (or lack thereof), see Dethier (2022: 12–13).

to convey information concerning it without distortion. Third, how pressing is the political issue on which the scientific matter bears? Insofar as we have time to thoroughly explain complex matters, and dispel public misconceptions, falsification seems less necessary. Finally, how cooperative is the broader communicative environment? Holding complexity, background knowledge, and time equal, it also matters what other communicators are doing. Trying to accurately explain scientific matters is more likely to succeed when the media, and other influential communicators, are cooperative, in that they are willing and able to help the public improve their scientific understanding.

What does this mean for the necessity of communicating falsehoods? The upshot is that educational falsehoods are more likely to be necessary, and thus more likely to be permissible, insofar as we are communicating about something complex, pressing, about which the public is uninformed or misinformed, and in a context where other influential communicators actively foster misunderstanding. Accordingly, my defense of (non-transparent) educational falsehoods will be restricted to these conditions. I agree that, the more we move away from these conditions, the harder it is to make the case that educational falsehoods are epistemically needed.³⁰

Though important, this qualification does not render my defense trivial by any means. This is because the non-ideal conditions just described are common in real-world science communication.³¹ This is notably the case with communication about climate change and climate research. The issue is extremely pressing; it raises technical or esoteric issues; these issues are often poorly understood, or indeed gravely misunderstood, by significant portions of the public; and rampant disinformation helps sustain these misunderstandings.

To make this more concrete, consider again the proposal that, instead of masking climate-related disagreement, we should communicate honestly and transparently about it. Explaining why disagreement is not necessarily a sign of bad science is complex. Epistemologists themselves continue to debate why and when disagreement should reduce our confidence in a proposition (Christensen 2009). Moreover, this complexity is compounded by public misconceptions about the significance of scientific disagreement. For instance, Kovaka argues that science is often taught in a way that lionizes scientists, such as Galileo, who dissent from the mainstream. This, she suggests, encourages skepticism toward

30. My defense of educational falsehoods is therefore compatible with thinking that there is a presumption in favor of honesty. My point is that this presumption can be defeated in a meaningful set of non-ideal circumstances.

31. For relevant examples, see Anderson (2011: 153–157), John (2018: 81), Kovaka (2021: 2368). In making this point, I am disagreeing with Director (2023: 956, 962), according to whom the conditions under which scientific dishonesty would be justified are “exceedingly rare,” such that dishonesty is “almost always” wrong.

mainstream positions whenever they face dissent (Kovaka 2021: 2368). Given the urgency of climate change, there is little time to correct these misconceptions. But even if there were, climate deniers actively obstruct attempts at doing so. They strategically exploit misconceptions about scientific disagreement by elevating dissenting voices, thus further unsettling trust in climate scientists and their conclusions. Put together, all of this makes it extremely risky, epistemically speaking, to be fully honest and transparent about climate-related disagreement (on this point, see also John 2018: 81). In such circumstances, it may therefore be necessary to mask disagreement, and thus exaggerate the degree of consensus, to help the public form the true belief that climate scientists are credible, and that the positions held by panels of climate scientists are well supported by the evidence.

3.4. *Beyond Education?*

My purpose so far has been to show that non-transparent falsehoods are sometimes needed to perform a valuable educational function. At this point, however, one might worry that the falsehoods I am defending require scientists to go beyond a strictly educational role. Central to my argument, in §3.3, was the idea that the scientific issue at hand is “pressing” or “urgent.” Relatedly, many of my examples refer to the way falsehoods can help highlight “dangers” or “critical” problems. But the judgment that something is pressing, or urgent, or dangerous, or critical, is, at least in part, a value judgment—in particular, it depends on ideas about what is morally or politically desirable. And if scientists are making such practical judgments when deciding whether or not to deploy falsehoods, then one might think that they are no longer acting primarily as educators, who seek to inform their audiences. Rather, they are acting as “issue advocates,” who seek to mobilize their audiences to promote ends that they consider important. This, in turn, might seem like an inappropriate role for scientists to perform.³²

The judgment that a scientific issue is pressing or urgent does indeed require appealing to nonepistemic (e.g., moral or political) values. From this, however, it does not necessarily follow that scientists who make such judgements are acting as issue advocates. To appreciate this, it is important to note that science is inescapably laden with nonepistemic values. That is, philosophers of science have widely argued that such values usually feature, and should feature, at numerous stages of scientific practice—for example, when deciding which questions to pursue, which methodology to use, which variables to represent in models, how

32. On the distinction between advocacy and non-advocacy roles performed by scientists, see Pielke (2007: 15–21). For the view that scientists should not act as advocates, see Schroeder (2022: 54–56).

to interpret events when recording them as data, whether to accept a hypothesis based on existing data, and so on.³³

The significance of nonepistemic values carries over to science *communication*—including science communication that aims to educate or inform the audience. A key function of educating or informing people is to empower them to engage in rational, autonomous decision-making (Brighouse 2009: 35–36; Elgin 2020: 64–65). Now, as Stephen John (2018: 83; 2019: 69–70) and S. Andrew Schroeder (2022: 41–48) have both observed, properly informing people—improving their epistemic position in a way that facilitates their capacity for rational, autonomous decision-making—requires attending to their values.³⁴ This is because what pieces of information are needed for autonomous decision-making depends significantly on one’s values. To cite one of Schroeder’s examples, taken from the medical context, being informed *that a particular medicine contains gelatin* matters for a strict vegetarian’s capacity to make an autonomous decision to take, or not take, the medicine; but this same piece of information might not matter, or matter nearly as much, for a non-vegetarian (2022: 43). The broader point is that, insofar as education aims to facilitate autonomy—as we will further discuss in §4.2—it must consider nonepistemic values.

This, however, does not mean that education simply collapses into issue advocacy. For it matters *whose* values are being used. Scientists who act as issue advocates often seek to mobilize others to promote goals that they consider to be important.³⁵ In contrast, on the picture outlined above, the values relevant to educating one’s audience, and thus empowering them to engage in informed decision-making, are in the first instance the *audience’s* values (Schroeder 2022:

33. For discussion of the different ways in which scientific practice appeals to nonepistemic values, see, e.g., Douglas (2009: ch. 5), Elliott (2022: 7–15), and Winsberg & Harvard (2024: §4). There are different arguments as to why this should be. Some suggest that epistemic values alone cannot always uniquely determine which methodology we should choose, how we ought to interpret data, whether we should accept a given hypothesis on the basis of existing data, etc. See, e.g., Steel (2010: 25–32). But even if they could, many argue that scientists should still appeal to nonepistemic values, for example because they have a moral responsibility to consider the costs of potential errors. For this latter argument, see Douglas (2009: ch. 4). For an excellent overview, see Elliott (2022: 15–34).

34. As they both acknowledge, it may be very difficult to ascertain and “align” with the audience’s values, particularly in contexts marked by evaluative disagreement. But this problem is not specific to educational falsehoods. Rather, it stems from the value-ladenness of science, combined with the (common) view that scientists should try to align the values they use with those of their audience (e.g., the public). For discussion of this view, see also Elliott (2022: 46) and Winsberg & Harvard (2024: 60–61).

35. Birch (2021: 12), for example, defends a form of advocacy (“normatively heavy advice”), and acknowledges that “there is no reason to think” that the values it involves will match the public’s values.

54–56).³⁶ We can apply this to my running examples. When judging that climate change or ozone depletion constitutes a *pressing* issue, scientists needn't be appealing to their own values, or seeking to promote their own goals. Their judgment might instead be that these issues are pressing *in light of the audience's own values and concerns*. The thought might be, for instance, that given their audience's own concerns for their health and security, or for the environment, or for their descendants' well-being, they have a strong interest in being made aware of the reality of climate change or ozone depletion. In this way, we can account for the value judgements in question even if the scientist deploying falsehoods is primarily aiming to educate or inform, rather than advocate.³⁷

4. The Moral Costs of Educational Falsehoods

I have argued that communicating falsehoods can provide meaningful epistemic benefits to the audience, even if these falsehoods are non-transparent; and that, in an important set of cases, it would be very difficult to perform this educational function without resorting to such falsehoods. One might nonetheless insist that, despite their epistemic benefits, communicating non-transparent educational falsehoods remains morally wrong. Why might that be? The most immediate moral worry is that doing so is deceptive. The rest of this paper aims to examine this concern. Are educational falsehoods deceptive (§4.1)? And if so, does this mean that, despite their epistemic benefits, it is impermissible to deploy them (§§4.2–4.4)?

36. This feature (being governed by the audience's values) also distinguishes the educational falsehoods defended in this paper from what Manson refers to as "spin." While both can involve selective representations of the world, in spin "the process of selection is governed by *self-interested*, promotional reasons" (2012: 205, emphasis added). This arguably makes spin different from many kinds of issue advocacy too, since advocacy can be guided by the advocate's *values* without being in the advocate's *self-interest* (e.g., if the advocate values social justice and social justice does not straightforwardly align with their own self-interest).

37. The point here is that my argument is not committed to the view—which many, like Schroeder (2022), reject—that it is permissible for scientists to act as issue advocates. And I will focus, in what follows, on scientists who pursue a primarily educational aim. But nor is my defense of educational falsehoods strictly *incompatible* with thinking that scientists can sometimes engage in issue advocacy. If, like Birch (2021), one thinks such advocacy is sometimes permissible, the idea of educational falsehoods can help distinguish between two ways of mobilizing people to promote the speaker's aims: Some forms of advocacy are mediated by an educational function (such that they mobilize *by putting* people in a better epistemic position to assess the situation), whereas others are not (such that they mobilize by impairing people's epistemic position). For reasons to be outlined in §4, the former are significantly easier to justify than the latter. For discussion of the moral difference between educational and non-educational mobilizing in a non-scientific context, see Lepoutre (2024: 134–35).

4.1. Are Educational Falsehoods Deceptive?

Deception involves intentionally causing beliefs in one's audience that the speaker knows to be false (Carson 2010: 50). As discussed, many epistemically beneficial falsehoods are *not* deceptive. In many cases, scientists intentionally deploy misrepresentations in a context or manner that makes it plain to the audience that what they are communicating is a misrepresentation. But my focus, once more, is on educational falsehoods that are communicated *non-transparently*. Here, the speaker intentionally does not reveal, and may actively conceal, the falsity of what they are communicating. The case of masking disagreement is a paradigmatic example: the scientific panel exaggerates the level of consensus, while concealing that this constitutes an exaggeration.

Does this constitute deception? I believe it often does. To see this, consider that, in the non-ideal circumstances under consideration (§3.3), the epistemic benefits of the falsehood in question very often depend on the audience believing it. For example, in conditions where audience members misunderstand the epistemic significance of disagreement, a scientific panel hoping to impart truths about the reliability of climate science and the severity of climate change may need the audience to believe that there is less disagreement than there really is. If, as a result, they exaggerate the level of consensus to promote this last belief, then they are deceiving the audience: it is part of their intention that the audience form a belief they know to be false.

Or take the case of the ozone hole. We can, of course, imagine a version of this example where scientists refer to a "hole" in a transparently metaphorical sense, and do not intend for their audience to believe, counterfactually, that a literal hole exists. In this case, no deception occurs. Yet we can also imagine a version of this case where, due to the non-ideal communicative context, scientific communicators arguably need the audience to believe the misrepresentation. If, for example, the public knows little about atmospheric science, they may not realize the significance or imminence of the dangers associated with a thinning ozone unless they believe there is an actual hole going through the ozone layer. Suppose that, consequently, scientists hoping to impart truths about these ozone-related dangers tell the audience that there is a literal hole in the ozone, and conceal the fact that this is strictly speaking inaccurate. In this case, it seems plausible to think that the speakers *are* engaged in deception. Indeed, in this latter specification of the case, it is part of their plan that the audience believe the falsehood. Hence, in failing to be transparent about the inaccuracy of what they are communicating, they intend to induce a false belief.³⁸

38. The actual case is somewhat ambiguous between these two specifications of the ozone hole case. Grevsstuhl notes, for example, that the fact that reference to a hole was a misrepresentation was transparent to atmospheric scientists (who therefore often used quotation marks) but

One might object that this characterization is misleading. Even though (non-transparent) educational falsehoods intentionally cause false beliefs, their *ultimate* aim is to impart true beliefs, among other epistemic benefits, to the audience. Accordingly, one might deny—as Shlomo Cohen (2018: 494–496) has recently done—that deploying such falsehoods is deceptive.

I do not wish to rely on this response. The first reason is that, even if the aim of non-transparent educational falsehoods is to impart true beliefs and other epistemic benefits, they achieve this aim *by* creating false beliefs. Consequently, it seems more accurate to say that these falsehoods deceive as a means to an epistemically good end, rather than to say that they do not deceive at all.

To put this first point slightly differently, consider by contrast what Rebecca Brown and Michael de Barra have termed “altruistic lies.” An altruistic lie involves making “an untruthful statement to B with the assumption that B will not believe [this statement], and will instead come to believe the truth” (2023: 92). It is plausible to think that *this* is not deceptive, since it does not involve an intention to cause any false belief. But, to reiterate, the non-transparent educational falsehoods I am focusing on are not like this: they often aim to induce a false belief *as a means* to an epistemic gain. To claim that they are not deceptive *at all* is thus to deny this key (and *prima facie* morally important) difference between them and altruistic lies.

The second reason for not relying on Cohen’s response is dialectical. Even if Cohen’s response were correct, it remains controversial. To bolster my defense of educational falsehoods, I therefore wish to show that their permissibility does not depend on accepting this response. That is, educational falsehoods can be permissible even if we assume that they are deceptive.

I will establish this conclusion by scrutinizing three moral concerns associated with deception: its cost to personal autonomy (§4.2), its potential erosion of public trust (§4.3), and its impact on democratic accountability (§4.4).

4.2. Deception and Personal Autonomy

Deception is thought to be problematic partly due to its impact on the personal autonomy of those who are deceived. The threat it poses for autonomy is threefold.

The first threat concerns autonomy in the belief-forming process. The ability to form one’s beliefs autonomously—that is, to “govern” one’s belief-forming process—depends on the ability to reason effectively. That ability, in turn,

that the metaphorical status of the term later became less transparent, particularly when used in broader public discourse (2018: 78). But my point here does not depend on the second specification being historically accurate.

depends partly on how good the evidence at one's disposal is, and to what extent one has the opportunity to scrutinize this evidence.³⁹ Deception might seem problematic in both respects. By putting forward falsehoods, and presenting those falsehoods as true, deception risks impoverishing the quality of one's evidence, and making it harder effectively to scrutinize this evidence. The upshot, one might worry, is that beliefs induced by deception are less likely to be formed autonomously.

But the issue is not just that deception can lead us to form beliefs non-autonomously. In addition, the beliefs induced by deception can impair our subsequent practical decision-making. This is because successful deception leads to false beliefs, and false beliefs can make it more difficult accurately to determine which courses of actions will promote one's values or goals. So, deception can impair one's ability to direct oneself towards one's values or goals.⁴⁰

The final threat deception poses for autonomy concerns the relationship between the deceiver and the deceived. Often, deception turns those who are deceived into a mere means to the deceiver's ends. In other words, deception often involves using those who are deceived as an instrument or tool, rather than treating them as a self-directing autonomous agent (O'Neill 1989: 111; de Melo-Martín & Intemann 2018: 43).

Consider again, for example, Fauci's false claim about the ineffectiveness of masks. First, Fauci's intervention impoverishes the quality of our evidence about masks, and conceals the fact that it is doing so. Hence, it arguably impairs our ability to reason effectively, for ourselves, about their usefulness. We therefore form our beliefs about masks less autonomously than we may otherwise have done. In turn, the false belief encouraged by Fauci's deception (e.g., that masks are ineffective at protecting those who wear them) diminishes our ability to direct our lives in accordance with our ends. Suppose that what I most value is minimizing my personal exposure to COVID-19. Erroneously believing that masks are ineffective makes it more difficult for me to select the correct means to this end. Finally, on one interpretation of this case, the audience is being treated as a mere means. Suppose Fauci's principal goal is to maximize public health, whereas some members of the public aim to maximize their own health. These two ends can diverge. Public health could best be served by having all non-frontline workers refrain from wearing masks (to make the limited supply available for frontline workers). But the best for *me* could be to wear a mask, while other non-frontline workers refrain from doing so. In this case, the deceptive interference does not simply make it difficult for me to pursue my ends. It also uses me as a tool to promote an end that is not my own.

39. Carter (2021: 21–22) refers to this as “inward-directed” autonomy.

40. This second problem pertains to what Carter (2021: 21–22) calls “outward-directed autonomy.” For discussion of this problem, see also Bok (1978: 20–21) and Carson (2010: 98).

Put together, these three costs to autonomy help explain why the Fauci intervention seems so problematic—and why, more generally, many philosophers consider deception to be morally wrong. One might therefore worry that, if (non-transparent) educational falsehoods are deceptive, then, notwithstanding their epistemic benefits, deploying them is impermissible.

There are several things to say in response. First, even if educational falsehoods impose these autonomy-related costs, it does not necessarily follow that communicating such falsehoods is always wrong. After all, there could conceivably be cases where the benefits of deception are so great that they outweigh its moral cost to personal autonomy.⁴¹

Admittedly, however, this first response is limited. It establishes that educational falsehoods *could* be permissible despite their cost to autonomy. Yet this is compatible with thinking that, given how unpalatable these autonomy-related costs are, the circumstances under which educational falsehoods are permissible remain vanishingly rare.

But there is a second and more important response. Not all forms of deception are equally problematic for autonomy. And educational falsehoods, in particular, seem less problematic for autonomy than standard cases of deception. Consider the second autonomy-related cost of deception, namely that deception makes it harder for the deceived to direct themselves in accordance with their values. Educational falsehoods can actually have the opposite effect. By improving the audience's epistemic position, an educational falsehood can, as mentioned earlier, help audience members direct themselves according to their own values.

Take, for example, the case where audience members mistakenly view scientific disagreement as indicative of bad science. As discussed in §3, this background misperception can prevent us from registering significant truths (e.g., that the reality and severity of climate change is very strongly supported by the evidence). Failing to recognize these truths can make it more difficult to identify the course of action that will best promote our values (e.g., reducing one's carbon footprint if one values the environment; pursuing geoengineering if one values industrial solutions to problems; etc.). The implication is that, insofar as educational falsehoods (such as exaggerations of the climate consensus) help us appreciate truths that are relevant to our values, they can also empower us to direct ourselves according to these values.

For the same reason, educational falsehoods do not necessarily treat the audience as mere means. To reiterate: the educational falsehoods I am defending can enable the audience to achieve their own goals or values by placing them in

41. I am therefore rejecting the claim, often attributed to Kant, that the moral prohibition on lying or deception is absolute. As Noggle (2022: §3.1) observes, this claim is almost universally rejected.

an overall better epistemic position. Whether one cares more about the environment or the promotion of industry, knowing that anthropogenic climate change is real is relevant to determining what course of action one should adopt to satisfy one's concerns. Thus, insofar as educational falsehoods empower people to pursue their own values, it seems inaccurate to say that they treat people as mere means or instruments.⁴²

Can we say something similar about the first autonomy-related cost, namely undermining autonomy in the belief-formation process? One might try to alleviate this concern by pointing to the audience's pre-existing epistemic predicament. As explained in §3.3, my defense of educational falsehoods focuses predominantly on situations where people are uninformed about the relevant issue, and where the communicative environment is rife with misinformation. In such a situation, people's ability to reason about the relevant issue—and thus, their ability to engage in effective doxastic self-government—is already compromised to some degree. So, one might argue that, relative to the status quo, educational falsehoods come at little cost to autonomy in the belief-formation process.

But this response only goes so far. At most, it shows that the process through which (non-transparent) educational falsehoods induce beliefs may not be substantially worse, from the perspective of autonomy, than the process through which audience members would otherwise form their beliefs. Yet this should not distract us from the fact that educational falsehoods attempt to steer the audience's beliefs in a way that bypasses their rational scrutiny. Like Fauci's intervention, educational falsehoods achieve their positive effect (in this case, their epistemic benefits) by deploying falsehoods, and presenting them as true.

My conclusion is therefore not that (non-transparent) educational falsehoods come at *no* cost to personal autonomy. I believe they do: they realize their epistemic benefits in a way that bypasses the audience's rational scrutiny, and thus, their autonomous control over their belief-formation process.⁴³ But the crucial

42. There are stronger and weaker ways of interpreting what it means to treat someone not merely as means, but also as ends. At a minimum, Carson (2010: 84) notes, it requires acting in a way that is responsive to their goals or values. But many Kantian interpreters understand the requirement more strongly. On this stronger view, it also requires respecting others' rational agency (see, e.g., Fahmy 2018: 97). Importantly, the educational falsehoods I am defending can satisfy both interpretations. Not only do they help realize the audience's goals or values, but, by improving the audience's overall epistemic position, they help audience members *rationally pursue* their goals or values. Now, Korsgaard (1996: 139–142) suggests an alternative way of understanding what it means to respect someone's rational agency. On her view, doing so requires treating others in ways to which they could rationally assent. But educational falsehoods could pass this test too. I could rationally assent, in advance, to my doctor telling me something false, provided believing this falsehood would help me make a more informed decision. On the possibility of rationally assenting to deception, see Carson (2010: 81–82).

43. We can therefore understand (non-transparent) educational falsehoods as a form of epistemic paternalism. Like epistemic paternalism more generally, they interfere with an agent's

point remains that, in other respects, educational falsehoods seem less inimical to autonomy than standard cases of deception. Indeed, they may even enhance their audience's autonomy in meaningful ways. The epistemic benefits they induce can empower the audience better to pursue their own values. And, partly for this reason, they treat the audience not merely as instruments for others' ends.

The upshot is that, even if educational falsehoods are deceptive, considerations of autonomy are unlikely to constitute a decisive objection to their deployment. In some cases, educational falsehoods may increase the autonomy of their target more than they diminish it. And even when they do not, their cost to autonomy is substantially less than that of standard cases of deception—and so, it could plausibly be overridden by the value of their epistemic benefits, and of the downstream consequences of those benefits.

4.3. *Deception and Trust*

Nevertheless, one might worry that the deceptive character of (non-transparent) educational falsehoods makes them problematic for a different reason: simply put, it risks undermining trust in science.

The reason for this worry is straightforward. With all deception, there is a risk of detection. And, as Pamuk observes, the detection of deceit seems likely to undermine trust in the deceivers. Hence, resorting to educational falsehoods in science communication risks eroding the public's willingness to believe scientists. For Pamuk (2021b), this constitutes a decisive reason to steer clear of "noble lies." Instead, scientists should be honest and transparent, notably about "the uncertainty and limitations of their knowledge." Prasad (2020) likewise concludes that, given the risks deception poses for trust, "the safest path is to always and only present the truth."⁴⁴

The risk of eroding trust is a real problem. What is worse, there are reasons to think that this problem is *especially* severe in the non-ideal circumstances I am focusing on. In uncooperative communicative environments, scientific claims are likely to be closely scrutinized by ill-motivated actors who are intent on delegitimizing them. This scrutiny, in turn, increases the odds that educational

freedom or autonomy to conduct inquiry, to promote their epistemic good. Now, it is commonly agreed that epistemic paternalism *can* in principle be justified (see, e.g., Ahlstrom-Vij 2013). But epistemic paternalism varies extensively in terms of the methods it deploys (e.g., nudging, information selection, deception) and its domain (e.g., medical, legal, scientific). So the fact that it is in principle permissible leaves it open whether particular forms of epistemic paternalism are permissible. In this context, my aim in §4 is to show that an especially controversial form of epistemic paternalism (which deploys deception when communicating about science in contexts involving vigorous democratic debate) can be permissible.

44. See also Bok (1978: 28), Pamuk (2021a: 93), Director (2023: 957–958), and Willhelt (2023: 201).

falsehoods will be exposed. So, the same conditions that make educational falsehoods epistemically necessary (§3.3) may also make such falsehoods more dangerous for trust.

There are several things we can say to soften the force of this objection. The first is that *educational* falsehoods seem less vulnerable to this worry than many other kinds of falsehoods. The extent to which finding out that one has been deceived warrants distrusting the deceiver depends importantly on the rationale for the deception—and, relatedly, on the extent to which one is likely to find this rationale acceptable. Indeed, finding out that you deceived me to steal my life savings plausibly gives me a far stronger reason to distrust you than finding out that you deceived me to surprise me for my birthday (a rationale that, let us assume, I wholeheartedly endorse). The former act of deception thus seems, on the face of it, to present greater risks for trust than the latter.⁴⁵

It therefore matters that educational falsehoods are intended to improve their targets' epistemic situation—and that, as we saw in §4.2, such falsehoods consequently do not treat people as a mere means to another's ends, but rather aim to empower them to pursue their own ends. It is not clear that being deceived for *these* purposes provides a warrant, or at least a strong warrant, for distrust. After all, being deceived for these purposes is something one might plausibly and rationally assent to.⁴⁶ Accordingly, finding out, after the fact, that you deceived me for these reasons seems less likely to destroy my trust in you—including my trust in you as an epistemically reliable source—than finding out that you deceived me simply for your own gain, or in a way that did not improve my epistemic situation. Returning to the objection at hand, the point is that, even if it is true that educational falsehoods risk being exposed, exposure may be considerably less detrimental to trust than it would be with most other forms of deception.

Yet one might think that this first response, too, is insufficiently sensitive to the challenges posed by uncooperative communicative environments. Even if the rationale underpinning educational falsehoods were acceptable to the public (such that appreciating it would mitigate the trust-related costs of exposure), the response assumes that scientists can make this rationale clear to the public. But doing so may be very difficult in non-ideal circumstances, particularly if ill-motivated actors actively promote misinformation about this rationale, for example

45. There is some indirect empirical support for this in the context of environmental risk communication. Peters et al. (1997), for example, find that perceptions of trustworthiness depend, not just on perceived openness and honesty, but *also* on perceptions of care and concern. Deception motivated by care and concern for the deceived thus seems *prima facie* less threatening for trust than malicious deception.

46. See fn 50 for further discussion of the point that educational deception needn't warrant distrust.

by ascribing sinister motives to scientists who use educational falsehoods. This, to be clear, does not wholly negate the first response—it is not strictly speaking *impossible* successfully to communicate the rationale underpinning educational falsehoods. But it does mean that the extent to which this response can alleviate the trust-related worries at hand remains limited.

Still, these remaining worries about trust needn't constitute a decisive objection to educational falsehoods. One reason for this—and this is my second response—is that they do not necessarily constitute a *comparative* problem for educational falsehoods. The objection at hand emphasizes that, in non-ideal and uncooperative communicative environments, deploying educational falsehoods can erode trust. But, in these same circumstances, a strict commitment to telling the truth can *also* undermine trust in scientists and their findings.⁴⁷

Consider again the case of scientific disagreement. Pamuk recommends that, instead of masking their disagreement, scientists should publicize their dissenting opinions, in order to initiate a process of public deliberation on the topic. Yet, as we saw in §3.3, this recommendation can sometimes do more to damage trust in science than to sustain it. Where people falsely believe that disagreement is a mark of bad science, publicizing disagreement could lead them to lose trust in scientists.⁴⁸

Of course, ideally, we would dispel these misconceptions about scientific disagreement. This is precisely why Pamuk (2021a: 88–90) recommends that scientific advice be integrated within a robust process of democratic deliberation on scientific findings. But dispelling such misconceptions takes time, and, as Pamuk notes, it may require the media to play a supportive role. The problem is that, in non-ideal conditions, these conditions commonly do not obtain. Time may be in short supply if, say, scientists are communicating about vaccine safety amidst a deadly pandemic. Moreover, in uncooperative communicative environments, prominent media sources may actively fuel the misconception that disagreement means bad science, and opportunistically highlight scientific disagreement to erode trust in epistemic institutions (Kovaka 2021: 2368). This last worry is far from hypothetical. As Russell Muirhead and Nancy Rosenblum (2020: chs. 5–6)

47. This is true even if Director (2023: 958) is right that that, on the whole, dishonesty breeds distrust “more often than not.” For Director’s claim is compatible with thinking that, in a meaningful subset of (non-ideal) cases, dishonesty is no worse than honesty.

48. Both Moore & MacKenzie (2020: 5) and Pamuk (2021a: 94) deny that this is supported by the evidence, citing research by van der Bles et al. (2020). However, as Gustafson and Rice’s (2020) literature review shows, van der Bles et al. (2020) focus, not on the effects of publicizing disagreement between scientists, but rather on the effects of publicizing “technical uncertainty” (reporting quantitative results in terms of a range or probability) and “deficient uncertainty” (emphasizing a gap in knowledge). Gustafson and Rice’s (2020: 626–27) review further indicates that publicizing disagreement between scientists *does* consistently have negative effects on trust, even though publicizing technical uncertainty and deficient uncertainty tends not to.

have shown, contemporary disinformation often uses dissent to create a climate of disorientation, where the public no longer know whom to trust. And while it is possible to counteract such disinformation, doing so is typically a long-term task, which may not always be achievable in particularly pressing cases.⁴⁹

The upshot of this second point is that in the kinds of non-ideal circumstances I am focusing on, scientists often face a trust-related dilemma, where both educational falsehoods *and* telling the truth risk eroding trust.⁵⁰ So, even if educational falsehoods do come with trust-related risks (notwithstanding their distinctive rationale), these risks needn't always give us a reason to favor the communication of truth over educational falsehood.

An important qualification is needed here. The different claims considered in this section—about the likely effects on trust of different categories of falsehoods and of truth-telling—are at bottom empirical claims. Ideally, therefore, we would simply appeal to empirical evidence to determine whether or not using educational falsehoods in uncooperative communicative environments always has an overall negative effect on trust, relative to truth-telling; and, if it does, how significant this effect is likely to be. However, existing evidence tends to be insufficiently fine-grained to settle these questions. While there *is* evidence, on the one hand, that perceptions of dishonesty are often detrimental to trust, and on the other, that telling the truth can also damage scientific trust in some circumstances,⁵¹ existing evidence tends not to distinguish suffi-

49. The fact that, in the long term, we should be trying to correct the background misconceptions that make educational falsehoods necessary has implications for *which* educational falsehoods we should deploy. Specifically, educational falsehoods should be designed so that, even when they exploit background misconceptions (e.g., that disagreement necessarily indicates bad science), they do not *reinforce* those misconceptions. So, in the masking disagreement case, communicators should avoid saying “There is a scientific consensus that *p*. This matters because *all good science involves consensus*.” Instead, it would be better to say “There is a scientific consensus that *p*. This matters because scientific consensus is one indicator—though not the only possible indicator—that *p* is supported by strong evidence.”

50. Wilholt (2023) hints at a further trust-related worry prompted by this last point: namely, that exploiting commonly accepted “fictions” about scientific practice *in order to* maintain trust is in tension with the “pursuit of *trustworthiness*” (2023: 201, emphasis added). Yet this seems too quick. Using such falsehoods to sustain trust in science does mean that the audience will end up having inaccurate views as to *why* scientific testimony is trustworthy, or “warrants” trust. But this practice nonetheless seems compatible with scientific testimony in fact being trustworthy. After all, in the case of educational falsehoods, accepting the falsehood is likely to *improve* one’s overall epistemic position. According to John, therefore, a scientist who uses epistemically useful deception may well be a more “effective” informer than a scientist who adheres rigidly to honesty (2018: 83). And, insofar as this is the case—insofar as they *are* effective informers—it seems *prima facie* plausible to think that audiences are epistemically warranted in placing trust in them (even if they are mistaken as to *why* they are warranted in doing so).

51. On the positive connection between honesty and trust, see, e.g., Peters et al. (1997) and Intemann (2023: 351). On evidence that telling the truth can also negatively impact trust, see fn 48 and John (2018: 80–82).

ciently finely between different kinds of communicative contexts (e.g., cooperative and uncooperative), and different types of falsehoods (e.g., educational and non-educational).

Having said this, even if existing evidence cannot definitively settle the specific questions outlined above, it can still give us grounds for skepticism towards the most pessimistic hypotheses concerning the impact of falsehoods on trust. (Non-transparent) educational falsehoods are not new. For example, scientists, as we have seen, already practice masking disagreement. And they do so, often, in real-world contexts involving hostile media environments. So if such falsehoods were highly destructive of trust in science, all things considered, we might expect this trust to be low already. Yet trust in science remains, on the whole, quite high—including in areas of science, like climate science, that are highly politicized (Gundersen et al. 2022: 6–7; Cologna et al. 2023: 3–4; Wong 2024). Now, there is only so much we can infer from this evidence, for the reasons stated in the previous paragraph. But it does provide tentative grounds for thinking that, in real-world settings, including in settings involving uncooperative communicators, the use of educational falsehoods may not be as destructive of trust as it is sometimes feared to be.

4.4. Deception and Democratic Accountability

Over the course of my defense, I have invoked and relied upon several key qualifications. The permissibility of communicating non-transparent educational falsehoods depends, firstly, on the precise social context in which we find ourselves: educational falsehoods are more justifiable, as we have seen, in contexts involving pressing and complex issues, an uninformed public, and an uncooperative communicative environment (§3.3). And their permissibility depends, secondly, on the type of falsehood we are communicating: the falsehood in question must be such that the epistemic benefits of believing it are likely to outweigh its epistemic costs (§3.2).

One might therefore worry that my defense places excessive faith in scientists and other scientific communicators. It requires them to show restraint, and to exercise keen judgment (relating to descriptive *and* evaluative matters, as discussed in §3.4), when assessing when to deploy falsehoods, and what falsehoods to deploy. This, one might think, is problematic. Given my focus on non-ideal circumstances, we should not adopt an overly idealized vision of scientists. Real-world scientists are imperfect, and thus, accidentally or deliberately, they may fail to communicate the right kinds of educational falsehoods, in the right kinds of conditions (Bok 1978: 24–30; Director 2023: 956–962).

What this concern shows is that those tasked with communicating falsehoods must somehow be held accountable to the public. In other words, oversight mechanisms are needed to ensure that, as far as possible, science communicators deploy educational falsehoods in ways that conform to the conditions I have outlined. This seems problematic for my defense, because holding scientists and other scientific communicators accountable is likely to be very difficult.

What can we say about this worry? One might firstly deny that the problem at hand is really about educational falsehoods. *Any* attempt at integrating scientific expertise within democracy faces the problem of how scientists and other experts can be held accountable. Given the informational asymmetry between experts and laypeople, it is extremely difficult to assess whether expert communications are true or false, or whether they result from good epistemic practice (see, e.g., Anderson 2011; Pamuk 2021a). Thus, one might think that the problem stems not from anything specific to educational falsehoods, but rather from the informational asymmetry between scientists and laypeople.

But this initial response may be too quick. A skeptic might insist that the difficulty of holding scientists accountable is amplified when they deceptively communicate falsehoods. One reason for thinking so concerns the epistemic consequences of the communication. Communicating falsehoods about something is likely to cause false beliefs about it, and one might worry that the resulting ignorance in turn makes it difficult to hold communicators accountable.

Yet this first concern seems inapplicable to the educational falsehoods I am defending. Even if these produce false beliefs, they thereby put the audience in an overall *better* epistemic position than they would otherwise have been. Accordingly, if ignorance erodes accountability, this is conceivably an argument *for* educational falsehoods, rather than against them.

Still, there is a second reason for thinking that the deceptive communication of educational falsehoods makes it more difficult to hold scientific communicators accountable. The problem here is not so much about the epistemic consequences of the communication. Rather, it is about the lack of transparency regarding the reasons that underpin this communication. The falsehoods I am defending are not and often cannot be fully transparent: a scientific communicator cannot disclose that they are communicating a falsehood in order to produce epistemic benefits because, as discussed in §3.3, the epistemic benefits of the falsehoods I am defending may well depend on the audience believing those falsehoods. The worry is that this opacity makes it more difficult for the public to scrutinize scientific communications. If scientists are not transparent about the fact that some of what they are saying is false, this plausibly makes it harder to determine whether the falsehoods they communicate conform to the requirements I have outlined.⁵²

52. For this relationship between non-transparency and accountability, see Williams (2002: 207).

This second concern is more worrying than the first. But, on closer inspection, it does not constitute a decisive objection to educational falsehoods. The issue of transparency, and its relationship to accountability, is in fact a familiar one in democratic theory. In particular, deliberative democrats have influentially argued that it may sometimes be useful for public speakers to debate the merits of a policy in secret, and later express support for (or opposition to) that policy, without necessarily disclosing the reasons that led them to support (or oppose) it. Doing so is useful, in part, because it allows speakers to change their minds in response to these reasons, when doing so would otherwise be politically costly (e.g., Gutmann & Thompson 1996: ch. 3; Chambers 2004; Kogelmann 2021).⁵³

Crucially, deliberative democrats widely agree that this lack of transparency about the reasons underpinning one's communication *can* be reconciled with the demands of democratic accountability. This is because it is possible to devise institutional mechanisms that reduce the likelihood that non-transparency will lead to communicators behaving unaccountably. Numerous institutional proposals have been advanced (see Kogelmann 2021 for an overview). Some focus on the internal composition of the deliberating group. For instance, Simone Chambers (2004: 408) recommends that the composition of the secret deliberative body should mirror, as far as possible, that of the broader population, such that it remains responsive to diverse values, concerns, and interests. Other proposals involve external oversight measures. Here, one possibility would be to charge a committee of policy advisers, or even a small assembly of randomly selected citizens, with the task of assessing whether the group in question is making responsible use of its secrecy.⁵⁴

Although these proposals were not designed with educational falsehoods in mind, they may conceivably apply to this particular case of non-transparent communication. For example, we might ensure that scientific panels are internally diverse, so that, when they decide whether to publicly communicate falsehoods, they are more likely to do so in a way that is responsive to an appropriate range of values, concerns and perspectives (Moore & MacKenzie 2020: 3–4). In addition, or alternatively, we might ask a committee composed of scientifically literate policy advisers, or randomly selected citizens who have undergone issue-specific training, to monitor prominent scientific panels such as the IPCC.⁵⁵ The committee would examine the panel's findings and communications, and question the panel where needed, in order to assess whether their communications

53. Pamuk (2021a: 94–95) herself acknowledges the usefulness of non-transparency, and therefore insists that her proposal “does not involve the disclosure of internal deliberations or the immediate release of meeting minutes.”

54. The first proposal is inspired by Christiano (2012), the second by MacKenzie & Warren (2012).

55. This suggestion draws on Pamuk's (2021a) proposal for a randomly selected “science court.”

distort the findings—and where they do, whether it is plausible to think that these falsehoods serve an appropriate purpose. In cases where they blatantly do not, the committee could then blow the whistle.

To be clear, these proposals are tentative. There may well be better institutional mechanisms for promoting accountability, and it is beyond the scope of my present argument to identify the best such mechanism. But this should not detract from my central point: namely, that the issue of non-transparency—and more specifically, of its potential impact on accountability—is not unique to educational falsehoods. On the contrary, it is a familiar issue, shared by other communicative practices, which are nonetheless widely considered to be permissible and compatible with respecting the value of democratic accountability. The implication, and what the above proposals are meant to highlight, is this: just as we can devise strategies to make non-transparent deliberative practices accountable, so too we can devise strategies to hold communicators of non-transparent educational falsehoods accountable.

5. Conclusion

I have argued that communicating (non-transparent) falsehoods about scientific matters can play an indispensable role in promoting greater understanding of those matters, and that, even when this communicative practice constitutes a form of deception, it can still be permissible in real-world communicative environments.

Yet, as should be clear from the foregoing argument, this defense of educational falsehoods does not give scientists a license simply to deliver “noble lies” as they please. Indeed, the justification of educational falsehoods depends on at least three sets of conditions. First, it depends on whether the epistemic benefits of the falsehoods outweigh their epistemic costs—and so, it depends on the extent to which the falsehoods actually help simplify, highlight, and render familiar significant features of the matter at hand. Second, the justification also depends on whether there are alternative ways of producing these epistemic benefits. Educational falsehoods are easier to justify when such alternatives are lacking. In practice, whether or not this is the case depends on how pressing and complex the matter at hand is, how much the public knows about it, and how (un)cooperative the broader communicative environment proves to be. Lastly, the task of judging whether the first two sets of conditions are satisfied should not be left wholly to scientific communicators. Like other non-transparent communicative practices, the practice of deploying educational falsehoods must be subject to internal and external accountability mechanisms.

These conditions are not easily satisfied. Accordingly, real-world science communications may often fail to satisfy them. But nor are these conditions

exceedingly rare or hopelessly demanding. Many of these conditions are commonplace in real-world democracies (e.g., the uncooperative communicative environment). And others, though demanding, seem no more demanding than familiar practices aimed at reconciling secrecy and democratic accountability.

The broader upshot is that this defense of educational falsehoods does more than simply tell us that deploying scientific falsehoods is sometimes permissible. It also helps us understand and identify the various ways in which communicating falsehoods can and does go wrong. In doing so, it provides us with a clearer map for assessing and improving science communication in contemporary democracies.

Acknowledgements

For comments on previous drafts and/or helpful discussion, I am grateful to Rufaida Al-Hashmi, Giulia Bistagnino, Rebecca Brown, Joanna Demaree-Cotton, Jamie Draper, Corrado Fumagalli, Stephen John, Ruairí Maguire, Temi Ogunye, Zeynep Pamuk, Suzanne Whitten, Federico Zuolo, two anonymous reviewers for *Ergo* (whose extensive comments significantly improved the paper), Mark Steen (for his exceptionally helpful copyediting), and participants in the University of Genoa's Workshop on New Trends in Non-Ideal Democratic Theory. I am especially indebted to Jacob Barrett for his immensely helpful (and patient) reflections. This research was generously supported by a Leverhulme Research Fellowship (RF-2022-001: "Felicitous Fictions: Legitimate Falsehoods in Public Discourse").

References

- Ahlstrom-Vij, Kristoffer (2013). *Epistemic Paternalism: A Defence*. Palgrave Macmillan.
- Anderson, Elizabeth (2011). Democracy, Public Policy, and Lay Assessments of Scientific Testimony. *Episteme*, 8(2), 144–164.
- Asplund, Therese (2011). Metaphors in Climate Discourse. *Journal of Science Communication*, 10(4), 1–8.
- Bayes, Robin, Toby Bolsen, and James N. Druckman (2020). A Research Agenda for Climate Change Communication and Public Opinion. *Environmental Communication*, 17(1), 1–19.
- Beatty, John (2006). Masking Disagreement Among Experts. *Episteme*, 3(1–2), 52–66.
- Beatty, John and Alfred Moore (2010). Should We Aim for Consensus? *Episteme*, 7(3), 198–214.
- Berstler, Sam (2019). What's the Good of Language? On the Moral Distinction between Lying and Misleading. *Ethics*, 130(1), 5–31.
- Bhakthavatsalam, Sindhuja (2019). The Value of False Theories in Science Education. *Science & Education*, 28, 5–23.

- Birch, Jonathan (2021). Science and Policy in Extremis. *European Journal for Philosophy of Science*, 11(90), 1–27.
- Bok, Sissela (1978). *Lying: Moral Choice in Public and Private Life*. Vintage.
- Brighouse, Harry (2009). Moral and Political Aims of Education. In Alex Broadbent (Ed.), *Oxford Handbook of Philosophy of Education* (35–51). Oxford University Press.
- Brown, Rebecca and Michael de Barra (2023). A Taxonomy of Non-honesty in Public Health Communication. *Public Health Economics*, 16(1), 86–101.
- Carson, Thomas (2010). *Lying and Deception*. Oxford University Press.
- Carter, Jason Adam (2021). Epistemic Autonomy and Externalism. In Kirk Lougheed and Jonathan Matheson (Eds.), *Epistemic Autonomy* (21–40). Routledge.
- Chambers, Simone (2004). Behind Closed Doors: Publicity, Secrecy, and the Quality of Deliberation. *Journal of Political Philosophy*, 12(4), 389–410.
- Chen, Xiang (2012). The Greenhouse Metaphor and the Greenhouse Effect. In Lorenzo Magnani and Ping Li (Eds.), *Philosophy and Cognitive Science: Western and Eastern Studies* (105–114). Springer.
- Chinn, Sedona and P. Sol Hart (2022). Can't You All Just Get Along? Effects of Scientific Disagreement and Incivility on Attention to and Trust in Science. *Science Communication*, 44(1), 108–129.
- Christensen, David (2009). Disagreement as Evidence: The Epistemology of Controversy. *Philosophy Compass*, 4(5), 756–767.
- Cohen, Shlomo (2018). Manipulation and Deception. *Australasian Journal of Philosophy*, 96(3), 483–497.
- Cologna, Viktoria, et al. (2023). Trust in Climate Science and Climate Scientists: A Narrative Review. *PsyArXiv Preprint*. Retrieved from <https://osf.io/preprints/psyarxiv/hj2xk>
- de Melo-Martin, Inmaculada and Kristen Intemann (2018). *The Fight Against Doubt*. Oxford University Press.
- Dethier, Corey (2022). Science, Assertion, and the Common Ground. *Synthese*, 200(30), 1–29.
- Douglas, Heather (2009). *Science, Policy, and the Value-Free Ideal*. University of Pittsburgh Press.
- Director, Samuel (2023). Public Health Officials Should Almost Always Tell the Truth. *Journal of Applied Philosophy*, 40(5), 951–966.
- Duffy, B. (2022) *Public Perceptions on Climate Change*. Retrieved from <https://www.kcl.ac.uk/policy-institute/assets/peritia-climate-change%E2%80%8BB.pdf>
- Elgin, Catherine (2017). *True Enough*. MIT Press.
- Elgin, Catherine (2020). The Ends of Education. In Prakash Iyer (Ed.), *Conceptualizing Education, and Related Issues* (58–77). Azim Premji University.
- Elliott, Kevin (2022). *Values in Science*. Cambridge University Press.
- EPA (2023). Basic Ozone Layer Science. *Environmental Protection Agency*. Retrieved from <https://www.epa.gov/ozone-layer-protection/basic-ozone-layer-science>
- Fahmy, Melissa (2018). Kantian Perspectives on Paternalism. In Kate Grille and Jason Hanna (Eds.), *Routledge Handbook of the Philosophy of Paternalism* (96–107). Routledge.
- Fallis, Don (2009). What is Lying? *Journal of Philosophy*, 106(1), 29–56.
- Flottum, Kjersti and Oyvind Gjerstad (2017). Narratives in Climate Change Discourse. *Climate Change*, 8(1), 1–15.
- Fraser, Rachel (2021). Narrative Testimony. *Philosophical Studies*, 178(12), 4025–4042.
- Grevsmuhl, Sebastian (2014). The Creation of Global Imageries: The Antarctic Ozone Hole and the Isoline Tradition in the Atmospheric Sciences. In Birgit Schneider and

- Thomas Nocke (Eds.), *Image Politics of Climate Change* (29–54). Columbia University Press.
- Grevsmuhl, Sebastian (2018). Revisiting the “Ozone Hole” Metaphor: From Observational Window to Global Environmental Threat. *Environmental Communication*, 12(1), 71–83.
- Gundersen, Torbjorn, et al. (2022). A New Dark Age? Truth, Trust, and Environmental Science. *Annual Review of Environment and Resources*, 47, 5–29.
- Gustafson, Abel and Ronald Rice (2020). A Review of the Effects of Uncertainty in Public Science Communication. *Public Understanding of Science*, 29(6), 614–633.
- Gutmann, Amy and Dennis Thompson (1996). *Democracy and Disagreement*. Belknap Press.
- Intemann, Kirsten (2023). Science Communication and Public Trust in Science. *Interdisciplinary Science Reviews*, 48(2), 350–365.
- John, Stephen (2018). Epistemic trust and the ethics of science communication: against transparency, openness, sincerity, and honesty. *Social Epistemology*, 32(2), 75–87.
- John, Stephen (2019). Science, Truth and Dictatorship. *Studies in History and Philosophy of Science*, 78, 64–72.
- Keohane, Robert, Melissa Lane, and Michael Oppenheimer (2014). The ethics of scientific communication under uncertainty. *Politics, Philosophy & Economics*, 13(4), 343–368.
- Kitcher, Philip (2001). *Science, Truth, and Democracy*. Oxford University Press.
- Kogelmann, Brian (2021). Secrecy and transparency in political philosophy. *Philosophy Compass*, 16(4), 1–10.
- Korsgaard, Christine (1996). *Creating the Kingdom of Ends*. Cambridge University Press.
- Kovaka, Karen (2021). Climate Change Denial and Beliefs about Science. *Synthese*, 198(3), 2355–2374.
- Lepoutre, Maxime (2024). Mobilizing Falsehoods. *Philosophy & Public Affairs*, 52(2), 106–146.
- MacKenzie, Michael and Mark Warren (2012). Two Trust-Based Uses of Minipublics in Democratic Systems. In Jane Mansbridge and John Parkinson (Eds.), *Deliberative Systems* (95–124). Cambridge University Press.
- Mandler, Jean and Nancy Johnson (1977). Remembrance of Things Parsed. *Cognitive Psychology*, 9(1), 111–151.
- Mazur, Allan and Jinling Lee (1993). Sounding the Global Alarm: Environmental Issues in the US National News. *Social Studies of Science*, 23(4), 681–720.
- Moore, Alfred and Michael MacKenzie (2020). Policy-Making During Crises. *BMJ*, 371, 1–7.
- Moser, Susanne and Lisa Dilling (2004). Making Climate HOT. *Environment*, 46(10), 32–46.
- Muirhead, Russell and Nancy Rosenblum (2020). *A Lot of People Are Saying*. Princeton University Press.
- Noggle, Robert (2022). The Ethics of Manipulation. *Stanford Encyclopedia of Philosophy* (Summer 2022 Edition), Edward N. Zalta (Ed.). Retrieved from <https://plato.stanford.edu/entries/ethics-manipulation/>
- Nussbaum, Martha (2009). Tagore, Dewey, and the Imminent Demise of Liberal Education. In Harvey Siegel (Ed.), *Oxford Handbook of Philosophy of Education* (52–65). Oxford University Press.
- O’Neill, Onora (1989). *Constructions of Reason*. Cambridge University Press.

- Pamuk, Zeynep (2021a). *Politics and Expertise*. Princeton University Press.
- Pamuk, Zeynep (2021b). 'Noble Lies': We need processes that check experts' bad behavior. *The Chronicle of Higher Education*, 68(7), 17–20.
- Pielke, Roger (2007). *The Honest Broker*. Cambridge University Press.
- Potochnik, Angela (2017). *Idealization and the Aims of Science*. Chicago University Press.
- Prasad, Vinay (2020). Why Did Fauci Move the Herd Immunity Goal Posts? *MedPage Today*. Retrieved from <https://www.medpagetoday.com/opinion/vinay-prasad/90445>
- Rancourt, Benjamin (2017). Better Understanding Through Falsehood. *Pacific Philosophical Quarterly*, 98(3), 382–405.
- Ratcliffe, Mary and Marcus Grace (2003). *Science Education for Citizenship*. Open University Press.
- Roche, Darragh (2021, June 2). Fauci Said Masks 'Not Really Effective in Keeping Out Virus,' Email Reveals. *Newsweek*. Retrieved from <https://www.newsweek.com/fauci-said-masks-not-really-effective-keeping-out-virus-email-reveals-1596703>
- Schroeder, Stephen (2022). An Ethical Framework for Presenting Scientific Results to Policy-Makers. *Kennedy Institute of Ethics Journal*, 32(1), 33–67.
- Shiffrin, Seana (2014). *Speech Matters: On Lying, Morality, and the Law*. Princeton University Press.
- Slater, Matthew (2008). How to Justify Teaching False Science. *Science Education*, 92(3), 526–542.
- Steel, Daniel (2010). Epistemic Values and the Argument from Inductive Risk. *Philosophy of Science*, 77(1), 14–34.
- Strevens, Michael (2016). How Idealizations Provide Understanding. In Stephen Grimm, Sabine Ammon, and Christoph Baumberger (Eds.), *Explaining Understanding* (37–49). Routledge.
- Ungar, Sheldon (1998). Bringing the Issue Back In: Comparing the Marketability of the Ozone Hole and Global Warming. *Social Problems*, 45(4), 510–527.
- van der Bles, Anne Marthe, et al. (2020). The Effects of Communicating Uncertainty on Public Trust in Facts and Numbers. *Proceedings of the National Academy of Sciences*, 117(14), 7672–7683.
- Veit, Walter, et al. (2021). In Science We Trust? Being Honest About the Limits of Medical Research During COVID-19. *American Journal of Bioethics*, 21(1), 22–24.
- Watson, Lani (2016). The epistemology of education. *Philosophy Compass*, 11(3), 146–159.
- Wilholt, Torsten (2023). Harmful Research and the Paradox of Credibility. *International Studies in the Philosophy of Science*, 36(3), 193–209.
- Williams, Bernard (2002). *Truth and Truthfulness*. Cambridge University Press.
- Winsberg, Eric and Stephanie Harvard (2024). *Scientific Models and Decision-Making*. Cambridge University Press.
- Wong, Carissa (2024). Largest Post-Pandemic Survey Finds Trust in Scientists is High. *Nature*, 626, 704.