The Ethics and Epistemology of Persuasion

Robin McKenna[1,2]

[1]Department of Philosophy, University of Liverpool, Liverpool, UK, and [2]The African Centre for Epistemology and Philosophy of Science, University of Johannesburg, Johannesburg, South Africa
Email: r.j.mckenna@liverpool.ac.uk

Abstract
What is persuasion and how does it differ from coercion, indoctrination, and manipulation? Which persuasive strategies are effective, and which contexts are they effective in? The aim of persuasion is attitude change, but when does a persuasive strategy yield a rational change of attitude? When is it permissible to engage in rational persuasion? In this paper, I address these questions, both in general and with reference to particular examples. The overall aims are (i) to sketch an integrated picture of the psychology, epistemology, and ethics of persuasion and (ii) to argue that there is often a tension between the aim we typically have as would-be persuaders, which is bringing about a rational change of mind, and the ethical constraints which partly distinguish persuasion from coercion, indoctrination, and manipulation.

In a pluralistic society, people disagree. Some of the time, we are happy to "live and let live." Other times, especially when the subject is ethics or politics, we try to persuade people to change their minds. This paper is about the ethics and epistemology of persuasion. Persuasion raises epistemological questions because, at least some of the time, the aim of persuasion is not just to change attitudes (or influence behavior) but to bring about a *rational* change of attitude—a change based on relevant evidence and reasons. Persuasion raises ethical questions because we can ask, of any persuasive strategy, whether it is *permissible* to utilize it.

My aim in this paper is not so much to answer these questions as to identify what I take to be some of the most important questions about the epistemology and ethics of persuasion. This is an exploratory paper; I hope that the reader who likes the questions I raise but is unsure about my initial attempts to answer them takes the paper as an invitation to elaborate on their own answers. I will however give some reasons for thinking that answering epistemological questions about persuasion will not necessarily help with answering ethical questions. That a persuasive strategy is liable to bring about a rational change of attitude in no way means that utilizing the strategy is ethically permissible. I will argue that this is so both in the context of interpersonal persuasion (one person attempting to persuade another) and "mass persuasion" (one person or group attempting to persuade a larger group). There is, therefore, a tension between the aim we typically have as would-be persuaders, which is bringing about a rational change of attitude, and the ethical constraints that partly distinguish persuasion from other, less benign ways of influencing attitudes, such as indoctrination and manipulation.

Hopefully this is enough to pique your interest. If it is not, let me advertise that the paper will also address the following issues:

-
-
-
-

Running through the paper is the following idea: developing a theoretical framework for thinking about persuasion will put us in a better position to consider the ethical and political questions raised in contexts where we might want to persuade. For example, should we try to persuade those who are vaccine-hesitant to change their attitudes toward vaccines and vaccine safety? If so, how? What about those who are skeptical about the risks posed by climate change? The driving idea behind this project is that we can only answer these questions once we have a systematic understanding of what persuasion is, and of its psychology, ethics, and epistemology.

1. Persuasion and Rhetorics of Persuasion

What is persuasion? How does it differ from coercion, indoctrination, and manipulation? A representative definition of persuasion comes from Daniel O'Keefe (2016). For O'Keefe, persuasion is "a successful intentional effort at influencing another's mental state through communication in a circumstance in which the persuadee has some measure of freedom" (p. 4). I'll work with this definition in this paper, so it will be helpful to break it down a bit.

First, while the definition is worded in such a way that one naturally thinks of persuasion in interpersonal communication (person A tries to persuade person B of something), we can talk about persuasion in a wide range of contexts, including mass communication contexts, where the "medium of persuasion" might be a speech, a video, a text, or something else that carries a message that the creator(s) intend to convince the audience of. In what follows I talk of "persuasive strategies," by which I mean acts of (attempted) persuasion that carry a content—a "persuasive message"—that is intended to persuade someone (or some group) of something, where that content may be any of a wide range of different things (an argument, a piece of evidence or information, a stance, a representation of the world as being a certain way) and it may be presented in a variety of different ways (e.g., in a work of philosophy or an op-ed, in a video, in an artwork).

Second, "persuasion" is a success term: if you try and persuade someone of something, but you fail, then we can say that you *attempted* to persuade but not that you *persuaded* them. Further, persuasion is *intentional*: if you say something that influences someone's (relevant) attitudes, but had no intention of doing so, we might say that you changed their attitudes but not that you persuaded them to do so. Persuasion is also *directional*: when you (attempt to) persuade someone of something, you intend to influence their (relevant) attitudes in a particular direction (cf. Coppock, 2022). Your attempt at persuasion is only successful if their attitudes move in the intended direction.

Finally, we only talk about persuasion in situations where the target of the persuasive strategy *freely* changes their attitudes. I do not have a theory of freedom but suffice it to say that this condition is meant to distinguish persuasion from *coercion*, *indoctrination*, and *manipulation*. Persuasion differs from coercion, indoctrination, and manipulation in that the coercer, indoctrinator, or manipulator attempts to influence through some form of communication but in a way that

circumvents their target's freedom. The simplest way to see what this amounts to is to compare coercion, indoctrination, and manipulation with *rational* persuasion. There is a clear difference between persuading someone to modify their behavior or attitudes by getting them to appreciate the force of relevant arguments, evidence, or reasons (that is: rational persuasion) and getting them to adopt the behavior or attitude in question by somehow tricking them into adopting it, or via a program of indoctrination (see DiPaolo & Simpson, 2016).

I want to say a bit more about the distinction between rational and non-rational persuasion. While I think there must be a satisfactory version of this distinction, it turns out to be more difficult to draw the distinction than many assume.

In a recent paper, Thomas Mitchell and Thomas Douglas give a helpful definition of rational persuasion:

> A rationally persuades B to adopt attitude α if (i) A brings it about that B adopts α, and (ii) A does so only by giving B reasons for adopting α, and (iii) B adopts α on the basis of recognising (some of) the reasons given by A, and (iv) A intends each of (i)-(iii) (Mitchell & Douglas, 2024, 3).

Thus, if I bring it about that you change your attitude towards higher taxes (you are now more in favor of them than you were before) by giving you some arguments for higher taxes, and you change your attitude because of those arguments, then I have rationally persuaded you to change your attitude. You have not just changed your attitude; you have changed your attitude because of the reasons I offered you.

Let me explain three aspects of this definition. First, the requirement that A brings it about that B adopt the relevant attitude (condition i) corresponds to the point that "persuasion" is a success term, where success means "pushing" B's attitudes in the desired direction. The requirement that A intends that the other requirements of Mitchell and Douglas's account be met (condition iv) corresponds to the point that persuasion is intentional.

Second, the requirement that B adopts the attitude because of the reasons given to them by A (condition iii) is necessary because there could be a situation in which the other conditions are met but B adopts the attitude for some other reason. For instance, I might try to convince you of the need for higher taxes by giving you lots of reasons for raising taxes, but you end up agreeing with me not because of these reasons but because you were motivated to do some new research, which gave you some different reasons for being in favor of higher taxes.

Third, the requirement that A brings it about that B adopts an attitude by giving B reasons for adopting that attitude is meant to distinguish (attempted) rational persuasion from non-rational or *a* rational means of persuasion. Amelia Godber and Gloria Origgi (2023) argue that there is a difference between persuading by directly offering facts, evidence, and reasons (rational persuasion) and persuading by appealing to emotions or making cultural, personal, or political identities salient (non-rational persuasion). In the former sort of case, you give your audience the reasons in question; in the latter sort of case, you may succeed in persuading your audience, but you do not do it by giving them reasons. (Mitchell and Douglas take a similar view, but largely because of the dialectical aims of their paper).

The problem with this way of distinguishing between rational and non-rational persuasion, though, is that it seems to assume that the only way in which you can give your audience reasons is by presenting them explicitly. Imagine, for example, a video depicting the horrors of modern factory farming with little narration beyond what is needed to explain what is being depicted. It seems right to say that the video gives its audience reasons for thinking that factory farming is morally wrong—indeed, reasons for thinking that it is a moral horror. Part of the reason why this seems right is that the creators of the video presumably intended to draw the audience's attention to relevant moral reasons, and moreover, they intended to bring their audience to change their attitudes towards factory farming because of these reasons. However the moral reasons in question are not themselves

stated explicitly in the video. Indeed, part of the persuasive power of the video requires that they not be stated explicitly; this would detract attention from the visual images that are meant to prompt moral reflection.

What we need then is an account of rational persuasion which allows you can give someone reasons for adopting an attitude without simply stating those reasons. While such an account is presumably available, I will not myself try to develop it in this paper. This is for two reasons. The first is that, even if you have an account on which someone who produces a video depicting the horrors of factory farming is not engaged in rational persuasion, someone who responds to the video as intended—that is, by changing their attitudes towards factory farming—is responding rationally, indeed no less rationally than if they had been persuaded by an argument where those reasons were set out in the form of premises and a conclusion that follows from the premises. The lesson then is that we need to distinguish between the sort of persuasion a would-be persuader is engaged in (rational or non-rational) and the rationality of any change in attitude that results from the attempt at persuasion. Changes in attitude in response to non-rational forms of persuasion may themselves be rational.

The second reason is that, over and above the distinction between rational and non-rational persuasion, there is a distinction between two different *rhetorics* of persuasion. The first is a rhetoric of persuasion that not only emphasizes relevant facts, evidence, and reasons, but presents those reasons explicitly, unadorned, and—as the proponent of this rhetoric would put it— "without spin." This rhetoric of persuasion exemplifies what Vid Simoniti (2021) calls the "objective style" of persuasion. The objective style "separates the speaker's idiosyncratic position from the content of her arguments" (p. 564), eliminates "such features as wilful self-contradiction or lack of seriousness" (ibid.) and "makes no concessions to laziness, or to biases, or to being easily distracted, or indeed to the propensity to be moved by anything other than the force of the better argument" (pp. 564–5). Because it involves "perspicuously structuring arguments and impartially laying out evidence," the objective style is, at least in principle, accessible to all—it is easy for reasonable participants to follow (p. 564). Philosophical arguments, at least when set out in the form of contemporary analytic philosophy, are paradigm examples of this rhetoric of persuasion.

The second is a family of rhetorics of persuasion that, if they present facts, evidence, and reasons at all, do so by blending style with substance. These rhetorics all depart, in some way or other, from the "norm" prescribed by the objective style. Consider rhetorics of persuasion that foreground the speaker's "idiosyncratic position" as well as, or rather than, the content of their arguments (e.g. persuasive strategies that appeal to first-personal experience). Or consider more artistic rhetorics—rhetorics that embrace a degree of ambiguity and perhaps even self-contradiction (in the way in which great works of literature often do), rhetorics that are playful, such as satire (one of Simoniti's examples), or rhetorics that use the power of visual images (as in the example above of a video depicting factory farming). Finally, consider rhetorics of persuasion that in one way or another exploit the audience's biases (cf. McKenna, 2020). While such rhetorics of persuasion may involve the presentation of relevant facts, evidence, and reasons, they are not presented in the way that is typical of the objective style. As a result, they may not be equally accessible to all audiences, but it may be that they are designed to be accessible to a particular audience, or to fit with that audience's sensibilities.

In what follows I will discuss epistemological and ethical issues that arise both for rhetorics of persuasion that exemplify the objective style and for rhetorics that, in some way or other, depart from it. But let me make three further comments about this distinction between different rhetorics of persuasion.

First, it is possible to utilize the objective style in a way that disguises what you are really trying to do. Someone might engage in the objective style, claiming that they are simply setting out the facts, unadorned and without spin, but really be presenting a highly selective subset of the facts, designed to lead their audience to accept a particular conclusion. Consider the philosopher who gives a heavily slanted summary of the arguments for and against their favored philosophical view, or the

historian who argues for the virtues of their chosen figure or cause by selecting facts that cast that figure or cause in an overly positive light. This philosopher (or historian) may be said to be utilizing the objective style, and perhaps even to be engaging in rational persuasion, but they are clearly being manipulative, perhaps even deceptive. (It is also possible to engage in the non-objective style while being manipulative and deceptive, though the deception would not consist in pretending to be objective when one is not being objective).

Second, the objective style articulates an ideal—a rhetoric of persuasion that consists purely in stating relevant facts, evidence, and reasons. When we talk of rhetorics of persuasion that depart from the objective style we are talking about rhetorics of persuasion that depart from that ideal, whether because they do not involve the explicit statement of facts, evidence, and reasons at all (they are implicit, implied or suggested, not explicit), or because they involve the explicit statement of a particular subset of (relevant) facts, evidence and reasons (a subset that the would-be persuader knows will be effective in the present context). This means that there is not necessarily anything that rhetorics of persuasion that depart from the objective style have in common, beyond the fact that they depart from the objective style.

Third, in describing the objective style as an ideal I am not intending to take a stance on the relative values of the objective style and the various non-objective styles. I am using the word "ideal" in what you might call a descriptive sense; in this sense, an ideal is simply that towards which someone or some practice aims. Whether the ideal is valuable depends on whether the practice that is organized around it is valuable. More generally, nothing I say in what follows should be read as implying that the objective style is in some sense *better than* rhetorics of persuasion that depart from it. Indeed, one of my aims is to show that rhetorics of persuasion that depart from the objective style can be used to bring about a rational change of attitude. Another one of my aims is to show that both rhetorics of persuasion that utilize the objective style and rhetorics that depart from it raise difficult ethical questions. Before getting to that, though, it will be helpful to look at the *psychology* of persuasion: what do we know about how persuasion works (when it works), and about what sorts of persuasive strategies are likely to work (and in what contexts)?

2. The Psychology of Persuasion

In the psychology literature there are a bewildering variety of models of persuasion (see e.g., Cialdini, 2001; Dillard & Shen, 2012; Maio et al., 2019; O'Keefe, 2016). There is also a large literature on the effectiveness of particular persuasive strategies in particular contexts, such as that of climate change skepticism (see e.g., Cook et al., 2018; Kahan, 2014; van der Linden et al., 2017). My aim in this section is not to give a systematic overview of these debates but rather to draw out some broad morals.

First, though, there is an important difference between the psychological and the philosophical literature on persuasion that is worth highlighting. Painting in broad strokes, philosophers typically take persuasive strategies to target *beliefs* (and perhaps *intentions*) whereas psychologists typically talk about *attitudes*. Attitudes in the psychologist's sense are, roughly, general evaluations of objects. Attitude objects can be concrete things (e.g., a car), people, or abstract entities (e.g., freedom, a policy).

A little less roughly, attitudes are "learned predisposition[s] to respond in a consistently favorable or unfavorable manner with respect to a given object" (Fishbein & Azjen, 1975, 42). Attitudes are typically affectively charged and can vary in their strength—your evaluation of an object can be strongly or weakly positive/negative. Most important for our purposes, attitudes are based on some combination of (i) beliefs about the attitude object, (ii) affect/emotions/feelings toward the attitude object, and (iii) actions/behavior involving the attitude object (Petty et al., 2003). So, for example, my attitude towards my car is based on my beliefs about it (it is reliable), my feelings about it (driving it makes me happy), or my behavior involving it (using it to drive to mountains).

One of the aims of this paper is to develop a framework for thinking about persuasion that integrates insights from both philosophy and psychology. So it is helpful to pause a second to consider how this psychological framework for thinking about persuasion fits with the sort of framework that is typically assumed in more philosophical discussions. As an initial attempt, we can say this: where philosophers typically think of persuasion as the attempt to influence (in a particular way) doxastic attitudes (beliefs) and practical attitudes (intentions), psychologists think of persuasion as the attempt to influence (in that same way) attitudes more generally. Moreover, the psychological literature gives us a helpful picture of how persuasive strategies work (when they work). Persuasive strategies change attitudes by changing the underlying basis of these attitudes. For example, you can change my attitude towards a policy by changing my beliefs about it, perhaps by giving me evidence that it will benefit everyone (cf. Coppock, 2022). But you can also change my attitude towards the policy by changing how I feel about the policy.

What do we know about how attitudes can be changed? Let us work with a relatively simple model of persuasion—the elaboration likelihood model (Petty & Cacioppo, 1986). The central idea behind the elaboration likelihood model is that there are two main variables that need to be considered when crafting a persuasive strategy: (i) how motivated and (ii) how likely the intended target of the message is to assess the central merits of the object, issue, or position in question.

When the target of the persuasive strategy is motivated and likely to put a lot of effort into assessing the central merits of the issue, we say that the "elaboration likelihood" is "high," and persuasion (change in attitude) is typically the result of slow, deliberative processes. For example, consider someone who changes their attitude toward a policy after carefully considering arguments in favor of the policy produced by one of its proponents. On the other hand, when the target is not motivated or likely to put a lot of effort in, we say that the elaboration likelihood is "low," and persuasion is typically the result of fast, automatic processes. For example, consider someone who changes their attitude towards a policy after following a simple decision rule—support the policies my party supports and oppose the policies my party opposes.

The elaboration likelihood model may not be an adequate model of persuasion—perhaps we need to add several more variables. But the important point is that the "right" persuasive method in any given context (the method most likely to bring about an attitude change in that context) depends on whatever variables feature in your favored model of persuasion. On the elaboration likelihood model, what matters is how high or low the elaboration likelihood is for the individual or group whose attitudes you want to change. A "high effort" persuasive strategy is more likely to succeed when elaboration likelihood is high than when it is low; a "low effort" strategy is more likely to succeed when elaboration likelihood is low than when it is high. On other models, different variables will need to be factored in. Still, the point stands that the way to craft a successful persuasive strategy or message is to design it around the relevant variables.

What factors influence elaboration likelihood? While there aren't going to be any "hard and fast" rules, we can expect the following factors to play a role:

- Perceived personal importance.
- Whether you will be held accountable for your attitude.
- Whether your existing attitudes are somewhat ambivalent.
- Whether you have encountered the persuasive message before.
- Whether you have the time to scrutinize the message.
- Various individual psychological differences.

Some brief comments (for details see Petty et al., 2003). While there are always going to be exceptions, it seems safe to say that *ceteris paribus* the more important you perceive something to be, the more motivated you are to think about it. Similarly, if you expect to be held accountable for your attitude on an issue (e.g., being asked about it, or asked to justify it) you are more motivated and more likely to think about it. More interestingly, it seems that you are more motivated to think

about an issue if your existing attitudes are somewhat ambivalent—if you are "of two minds." Less surprisingly, the more often you have encountered a persuasive message, and the more opportunities you have had to scrutinize it, the more likely you are to engage with it—through would-be persuaders need to be careful not to be too repetitive. Finally, there may be all sorts of individual psychological differences that impact elaboration likelihood, such as the need for cognition and ability to process certain kinds of information.

In short: choosing a successful persuasive strategy requires keying your strategy to features of the intended target audience. In the elaboration likelihood model, the central features are those that influence elaboration likelihood. On other models, they will be whatever features influence the things that the model says determine the impact of persuasive strategies. The important point is that successful persuasive strategies are typically *targeted* and based on a degree of understanding of the target audience. A corollary of this is that we should expect *different* persuasive strategies to be appropriate for different audiences and in different contexts. If your audience regards an issue as of paramount importance, persuasive strategies that require them to process a good deal of information may be effective; if they do not regard it as particularly important, different strategies are called for, strategies that require a good deal less effort.

So far, I have been talking about the psychology of persuasion and attitude change in general. One area where persuasion and attitude change have been extensively studied is science communication. There is an important debate within the literature on science communication about the relative merits of two kinds of persuasive strategies. The first kind of strategy consists, roughly speaking, of simply providing relevant arguments, data, and information. A lot of science communication uses this kind of strategy. It provides relevant figures and comparisons, consensus reports, simple explanations of mechanisms, and so on. Critics of these "simple information strategies" claim that they are often ineffective, particularly in contexts with high degrees of misinformation and polarization (Kahan et al., 2011; Lewandowsky & Oberauer, 2016; van der Linden et al., 2017). But some think these criticisms are overblown and that, utilized properly, simple information strategies can be effective (see Coppock, 2022; Gerken, 2022).

The second kind of strategy consists, also roughly speaking, of appealing to the cultural or political values shared, or presumed to be shared, by the intended target of the persuasive strategy (see Kahan, 2014; Kahan et al., 2011). This might involve a direct appeal to these values, as when an issue is framed in a way designed to appeal to someone with a certain set of values (e.g., framing climate change as an opportunity for enterprise and innovation), or when a spokesperson is chosen on the grounds that they exemplify certain values (e.g., using religious figures to make the case for the need to combat climate change). Critics of these "value-based reporting strategies" do not tend to take issue with their potential effectiveness but rather with whether they are necessary (Coppock, 2022) or desirable (Gerken, 2022).

This distinction—between simple information reporting and value-based reporting strategies—nicely parallels the distinction I drew in the previous section between rhetorics of persuasion that utilize the objective style (simple information strategies) and rhetorics that depart from the objective style (value-based reporting). In the next two section, I will address the debate between proponents of these different strategies, albeit indirectly, by looking at epistemological and ethical issues raised by persuasion. My aim is not so much to argue for one "side" of the debate over the other (both kinds of persuasive strategies can be useful) as to discuss epistemological and ethical issues pertaining to both kinds of strategies.

3. Epistemology of Persuasion

When you attempt to persuade someone of something, you intend to get them to change their (relevant) attitudes in a particular direction–the direction of the persuasive message. For certain purposes, it may be that a persuasive attempt is successful so long as it brings about a change of attitude in the right direction. You might think it is important to change the attitudes of climate

change skeptics because democratic norms require public support for climate change mitigation policies, in which case it may be enough to make attitudes less skeptical (cf. Anderson, 2011).

However, there may be situations where a little more is required. There is a difference between changing attitudes towards climate change (e.g., viewing it as more of a threat than you did before) and changing attitudes *for the right reasons* (e.g., because you now appreciate reasons why climate change is a real threat). One reason why you might want to bring about a change in attitude for the right reasons rather than a simple change in attitude is that a change in attitude for the right reasons is more *robust* (e.g., it may be less susceptible to being undermined by new evidence). Another reason is that you might be concerned about the epistemic status of people's attitudes, whether for their own sake (maybe you think it is good for people to have attitudes based on good reasons) or for the sake of some other goal, such as enhancing public debate and deliberation (cf. Ahlstrom-Vij, 2013). If you have these or related concerns, then simply bringing about a change of attitude in the desired direction is not enough for your persuasive attempts to succeed. The change of attitude must also be based on the right sorts of reasons.

One way of integrating epistemology into the psychology of persuasion is to use the tools of epistemology, especially theories of epistemic rationality, to evaluate persuasive strategies along epistemological dimensions. Just as some persuasive strategies may be more likely to bring about a change of attitude (in certain contexts) than other persuasive strategies, some persuasive strategies may be more suited to bringing about *rational* changes of attitude (in those contexts) than others.

It is here that we can draw some further connections between the psychology of persuasion and our earlier discussion of the objective style. It is perhaps natural to assume that rhetorics of persuasion that utilize the objective style, like simple information strategies, are more suited to bringing about epistemically rational changes of attitude than rhetorics of persuasion that depart from the objective style. After all, simple information strategies involve putting forward arguments, data, and information designed to "push" attitudes in the desired direction. If these attempts succeed, the change of attitude is likely based on the arguments, data, and information provided, and so is epistemically rational.

It is also perhaps natural to think that rhetorics of persuasion that depart from the objective style, like value-based reporting, are less well-suited to bringing about an epistemically rational change of attitude. These rhetorics of persuasion involve directly appealing to what seem like non-epistemic considerations (e.g., shared values), or appealing to a curated set of arguments, data, and information designed to appeal to the target audience. If these attempts succeed, the change of attitude is, you might think, not epistemically rational.

Natural or not, I think both these lines of thought are mistaken. First, there is a crucial distinction between the content of a persuasive message (or intent behind a persuasive strategy) and how that message is processed by the recipient. From the fact that a persuasive message contains arguments or reasons you cannot infer that it receives uptake because of the recipient's assessment of those arguments. It could be that it receives uptake because of something else entirely, such as the fact that the content of the message aligns with the recipient's cultural or political values. Or it could be that it receives uptake because of the recipient's assessment of the weight of arguments contained in the persuasive message, but that assessment is itself slanted by the recipient's cultural or political values, in such a way that they afford more weight to the argument than they would have if they had not had those values (Carter & McKenna, 2020).

Second, modes of persuasion that depart from the objective style can bring about a change of attitude that is epistemically rational. Here are two examples to illustrate the point. The first example is drawn from Dan Kahan (2014). Kahan argues for climate change communication strategies that frame climate change and particular mitigation policies in ways designed to appeal to the political values of the intended audience. For example, when addressing a conservative audience, you should frame the problems posed by climate change as an opportunity for businesses to innovate and develop new technologies, rather than (putting things somewhat crudely) requiring the overthrow of capitalism.

I think that someone who is persuaded by this way of framing things is being epistemically rational. If someone comes to think that climate change is a serious problem—or a more serious problem than they previously thought—because the issue is framed in a way that appeals to their political values, what has essentially happened is that a "block" that might have prevented them from recognizing the reasons for thinking that climate change is a serious problem has been removed or circumvented. Kahan's point is that we can choose to frame things in such a way as to nullify the impact of cultural and political values that prevent some people from recognizing the force of relevant scientific evidence. As he puts it:

> It would not be a gross simplification to say that science needs better marketing. Unlike commercial advertising, however, the goal of these techniques [such as framing] is not to induce public acceptance of any particular conclusion, but rather to create an environment for the public's open-minded, unbiased consideration of the best available scientific information (2010, p. 297).

By utilizing framing effects, we can reach people who are otherwise unable to make use of the best available scientific information. But, once you reach them, any change of attitude that is brought about is epistemically rational because it is based on an appreciation of relevant reasons and evidence (cf. McKenna, 2023, chap. 4).

The second example is drawn from Neil Levy (2021) and his discussion of nudging. Briefly: nudges (in the sense of Thaler & Sunstein, 2008) are meant to influence our attitudes, behavior, and beliefs by modifying the "choice architecture"—the background against which we interact with our environment. For example, an employer might automatically enroll new employees in the pension scheme as a way of getting more employees to enroll, or a supermarket might make healthier foods more visible as a way of encouraging healthy eating habits. There is plenty of debate about the extent to which nudges influence behavior, and indeed whether nudges influence behavior at all (see Mertens et al., 2022; Maier et al., 2022). But I will set this to one side, as my interest is not in whether nudges are effective, but in the rationality of responding to them in the way the "nudger" (the designer of the choice architecture) intended.

Levy argues that, far from bypassing our reasoning capacities, nudges—or at least a certain kind of nudge—provide evidence, and therefore any changes in attitude and behavior that they prompt are epistemically rational. For Levy, understood properly, nudges provide reasons and so, to the extent that they influence us, that influence is epistemically rational. As he puts it:

> Nudges don't simply manipulate us by bypassing our capacities to reason. Instead, they provide us with evidence, which we typically weigh appropriately. Nudges don't tend to provide arguments or evidence that fit our paradigms, but that's because our paradigms are of *first-order* evidence. We neglect *higher-order* evidence, but higher-order evidence is genuine evidence (p. 135).

Levy's view is that, while nudges may not provide evidence that directly bears on the issue at question (being automatically enrolled in a pension scheme is not evidence that enrolling is good for your long-term financial security), they are still a form of evidence. I'm not sure if Levy is right that they provide higher-order evidence (cf. Dutilh Novaes, 2023). But his basic idea seems to be that they provide evidence given certain assumptions about how the world works. For example, given the assumption that your employers aren't trying to take advantage of you, the fact that they recommend enrolling in the pension scheme is evidence that enrolling in the pension scheme is a good idea. To the extent that you have evidence that your company is not trying to screw you over, automatic enrolment in the pension scheme provides a kind of evidence, and responding by enrolling (or, rather, not opting out of enrolment) is a rational response to this evidence.

In my view, Levy is right that some nudges can be viewed as providing reasons and evidence in this way. Following Levy, nudges like automatic enrolment in a pension scheme function like implicit recommendations, and they provide reasons in the same way that recommendations provide evidence (at least, when we make certain assumptions about the person doing the recommending; see above). Someone who responds appropriately to a recommendation—that is, by doing what is recommended, assuming you trust the recommender—is responding to these reasons, even if they are not conscious of themselves as doing so. Viewed like this, responses to nudges are often rational in the same way that responses to testimony are often rational. When we believe what someone tells us, we are not necessarily conscious of ourselves as responding to the reasons to believe the thing in question provided by the testimony. But what makes believing testimony rational (when it is rational) is that it manifests responsiveness to these reasons.

It may well be that Levy goes too far in saying (or at least implying) that *all* nudges work like this. Consider a nudge like putting the image of a fly into urinals to reduce splattering (one of Thaler and Sunstein's own examples). It is hard to see how this functions as a recommendation to not make a mess, implicit or otherwise, because it provides no reason why splattering is bad or undesirable. It is simply an attempt to modify the environment in such a way that a certain form of behavior becomes less likely. It may well be that, as critics of nudges argue, some kinds of nudges (like the image of the fly in urinals) manipulate us and undermine the rationality of our attitudes and choices by bypassing our critical capacities entirely (see e.g., Bovens, 2009; Hausman & Welch, 2010; Riley, 2017). But, equally, I think Levy is right in saying that other kinds of nudges (like automatic enrolment in a pension scheme) do not bypass our critical capacities. They engage our critical faculties, but not necessarily at the level of conscious reflection.

I have provided two examples of rhetorics of persuasion that depart from the objective style yet seem to lead to changes of attitude that are epistemically rational. (At least, if we assume Levy's view of how nudges work; if you do not like this view, you can focus on the first example). You might object that these changes of attitude would be *more epistemically rational* if they had been achieved via a rhetoric of persuasion that utilizes the objective style. I think there is something to this, but I'm not convinced that the difference here is one of rationality. A more plausible way of putting the difference would be that, if you change your attitudes via a conscious, deliberative engagement with a wide set of relevant reasons, then you have a better chance of achieving other epistemic goods, such as understanding the issue(s) in question, than if you change your attitudes in response to nudges, framing strategies, or other value-based reporting strategies that operate below the level of conscious deliberation.

You might also object that rhetorics of persuasion which utilize the objective style are *more likely* to bring about epistemically rational changes of attitude than rhetorics that depart from it. Whether this is true in the context of science communication, particularly when it comes to scientific issues that are intertwined with political ones, is a topic of lively debate in the empirical literature. Until recently, the consensus seemed to be that it is not true in this particular context (Kahan, 2014), but this is now changing (Coppock, 2022). Rather than speculate about how this debate will turn out, let me just reiterate a point made earlier: different persuasive strategies are appropriate for different audiences and in different contexts. It may simply be that, in some contexts, we should utilize the objective style whereas, in other contexts, we should utilize a rhetoric of persuasion that departs from it, such as value-based reporting.

4. Ethics of Persuasion

I said at the beginning that we engage in persuasion because we want to change attitudes. Of course, there are many ways in which you might try to change attitudes. You might spread misinformation, engage in propaganda, cultivate myths, or what have you. In this section, I set aside ethical questions about methods of attitude change that are clearly problematic and focus on something that you

might think is less fraught: the ethics of *rational* persuasion. What could be wrong with trying to change someone's attitude by offering them evidence and reasons?

In this section, I discuss two issues in the ethics of (rational) persuasion. The first issue concerns autonomy and some ways in which even rational forms of persuasion might infringe on it. The second concerns a basic assumption that is made in many discussions of persuasion in mass communication contexts, which is that the "job" of someone tasked with persuading the public (e.g., to get vaccinated) is simply to "insert" true beliefs in the heads of members of the public. My aim in this section is not to resolve these issues, but to urge the importance of thinking carefully about them, and to identify some important lessons for how we think about different rhetorics of persuasion.

Let me start with persuasion and autonomy. It is tempting to think that there is a key difference between rhetorics of persuasion that utilize the objective style and rhetorics that depart from it. Persuasion is meant to differ from more coercive ways of influencing attitudes in that it does not infringe on the freedom of the intended target to make up their own mind. When a would-be persuader utilizes the objective style, they offer something—an argument, some evidence, a reason—that does not compel the intended target to change their mind, though—if it is a good argument —may make it rational for them to change their mind. In contrast, when they utilize a rhetoric of persuasion that departs from the objective style, they may bring certain arguments and reasons to the attention of their intended target, but they do so in a way that restricts the ability of the intended target to make up their own mind. Perhaps the would-be persuader frames the issue in a way designed to appeal to the intended target, or they appeal to values which they presume the intended target to have, in a way designed to push their thinking in a particular direction.

While I share these concerns about autonomy, the point I want to make here is that is less of a difference between rhetorics of persuasion that utilize the objective style and rhetorics that depart from it than there might initially appear. One of the conceits behind the objective style is that it is possible to separate the persuasive power of an argument from the persuasive power of the person giving that argument. This is especially problematic when we consider persuasion in interpersonal contexts, where it is very difficult to achieve this sort of separation. (This may also be true in mass communication contexts, but let us keep things simple). As George Tsai (2014) has highlighted, persuasive strategies that consist of simply presenting relevant arguments, evidence, and information can still disrespect autonomy precisely because we cannot separate the rational pressure of arguments from the rational pressure one exerts when one presents those arguments. He gives the example of a parent (Peter) who tries to persuade his daughter (Claire) to go to law school rather than to grad school to study philosophy. Peter does this by bombarding Claire with arguments why law school is the better option. Of this example Tsai says:

> when others offer us reasons to persuade us at the wrong time or in the wrong way, they make it harder for us to be able to engage more purely and directly with the reasons most centrally tied to the choice-worthiness of our options. When our deliberations are *distorted* in this way, this potentially alters the self-determining and self-expressive aspects of our decision … the point is that even the rational pressure of Peter's *reason-giving* (as distinguished from the rational pressure of the *reasons* themselves) might potentially alter the nature of Claire's deliberations in a way that results in a sense of loss for Claire … Insofar as the timing of Peter's attempt at rational persuasion precludes Claire from having the purer, more direct engagement with the reasons most centrally relevant to her deliberative situation, this limits her exercise of epistemic agency (95–6).

Tsai's point is that, even if Claire would be rational to be persuaded by Peter's arguments, his intervention disrespects her autonomy and epistemic agency precisely because Peter's giving those arguments to Claire, in an interpersonal context where power dynamics are in play (Peter is Claire's father), exerts pressure on Claire in a way that undermines her autonomy and inevitably alters,

perhaps even distorts, her deliberations. But Peter's persuasive strategy is a paradigm example of persuasion in the objective style.

Here's another example that makes much the same point. Imagine a patient, Craig, discussing treatment options for his chronic medical condition with his doctor. His doctor is a strong advocate of assistive technologies for treating Craig's condition. He spends a lot of time detailing why it would be a good idea for Craig to start using a particular device. Because the doctor does not himself have Craig's particular condition, he cannot say much about the potential downsides of using the device (cost, inconvenience, how well it would fit into Craig's lifestyle, etc.). Even if it would be rational for Craig to be persuaded by his doctor's arguments (they are good arguments, after all), in pressing these arguments the doctor disrespects Craig's autonomy and agency in much the same way that Peter disrespects Claire. His giving these arguments, in a context where the power dynamics are just as complicated (Craig's doctor has the power to decide which medicines and treatments Craig can access), exerts pressure on Craig in a way that undermines his autonomy and inevitably alters, perhaps even distorts, his deliberations. But, again, the doctor's persuasive strategy is a paradigm example of persuasion in the objective style.

I take these examples to show that rhetorics of persuasion that utilize the objective style can also run into troubles with autonomy. Moreover, they run into trouble for a reason which one would think also applies to rhetorics of persuasion that depart from the objective style: they manifest a lack of respect for the capacity of autonomous agents to deliberate for themselves. This is particularly problematic in interpersonal contexts where agents are deliberating about how to live their lives (what to do at grad school, how to manage their medical condition). The ethics of persuasion in interpersonal contexts is fraught, and using rhetorics of persuasion that utilized the objective style need not making it less fraught. Notice that it does not make it any less fraught to say that there is nothing *irrational* in changing your attitudes in response to reasons that you are offered. The point is not that Claire or Craig would be irrational to accept the arguments they are given; the point is that the act of giving them alters their deliberations in a way that infringes on their autonomy. Going back to nudging, one reason why I suspect there has been so much resistance to it is that its proponents seem to simply assume that, if they can show that there is nothing *irrational* in responding to a nudge as the nudger intended, they have thereby shown that there is nothing wrong with nudging. This is a mistake for the same reason that it would be a mistake to think that the fact that Craig would not be irrational to go along with his doctor's advice means that his doctor has done nothing wrong.

Let me now turn to persuasion in mass communication contexts. One might worry that frameworks for thinking about persuasion in mass communication contexts—think, for example, of science communication, which I discussed above—typically assume a picture on which the goal of "mass persuasion" is simply to insert true beliefs (or, more generally, desired attitudes—intentions, preferences, desires, etc.) into the "heads" of the intended audience. Thus, when a science communicator devises a strategy for convincing a skeptical public that a new vaccine is safe, their goal is essentially to get members of that public to have true beliefs about the safety of the new vaccine. The task for science communication studies is to figure out how to achieve that goal. Now, this picture clearly disrespects the autonomy of the intended targets of mass persuasion; it simply disregards it, and views people as simple repositories for true beliefs. But there is another issue here, which is that, even if you think it might sometimes be justified to view people as simple repositories for true beliefs, there may be serious unintended consequences of doing so.

Stephen John's work on persuasion in science communication is worth considering here, as it has several shortcomings that illustrate the point I want to make (see John, 2018a, 2018b, 2019). Here is a brief outline of John's basic picture. The central norm governing science communication is simple —science communicators should only communicate scientific claims that are well-established by the standards of the relevant scientific discipline (John, 2018a, 84). Call this The Central Norm.

John defends The Central Norm on the grounds that it promotes a particular goal: it is likely to lead to a situation where members of the public, not to mention policymakers, have beliefs about

scientific issues that are well-established by the standards of the relevant discipline. He argues that, at least in the case of climate science communication, The Central Norm overrides various ethical and political considerations. On his view, it can be entirely legitimate to communicate well-established scientific claims in ways that aren't entirely honest (maybe they ignore complexities and uncertainties liable to confuse the audience), sincere (maybe the science communicator does not personally accept the claim), or transparent (maybe they ignore the compromises and fudges in the scientific process that led to the claims being established). As John puts it:

> If a scientist knows that reporting a point estimate without adding further qualifications is likely to lead a policy-maker to some conclusion which it is in her epistemic interests to believe—such as that 'climate change will lead to ice-sheet collapse'—whereas a more 'honest' estimate is unlikely to lead to such belief—the policy-maker will disregard her advice as too complex—then she may be justified in making the first, spuriously precise estimate (2018a, 83–84).

Now, it may be that John is right in the particular case he has in mind, which is climate science communication. Perhaps the existential risk posed by climate change is such that the benefits of laypersons and policymakers accepting well-established claims about climate change are so great as to over-ride any concerns about honesty and the like. But what about the more general view that John gestures towards, which is that, so long as science communicators ensure their claims are well-established, it does not matter that much if they are—presumably within certain limits—less than fully honest, sincere, or transparent?

There are two problems with the more general view. First, the logic of John's argument is essentially consequentialist. In the case of climate change and climate change communication, the risks are perhaps such that The Central Norm and the goal it serves (ensuring that people have beliefs about relevant scientific issues that are well-established) trumps any other concerns, such as concerns about honesty and transparency. More generally, in the case of climate change, the risks might be so high as to justify adopting a framework for thinking about mass climate communication on which we simply view people as repositories for true (strictly speaking, for John well-established) beliefs. But this means that John's argument is not going to generalize to any scientific issue where the consequences of the public not having well-established beliefs are likely to be less disastrous.

Second, and more importantly, John's thought is that a bit of dishonesty or lack of transparency is justified so long as it means that well-established claims are more likely to receive uptake. This would make sense if dishonesty and lack of transparency were typically not barriers to scientific claims receiving uptake. But if dishonesty and lack of transparency, or indeed mere suspicions about lack of honesty or transparency, were barriers then that would be a problem. Interestingly, John's own analysis of science skepticism suggests that suspicions about lack of honesty and transparency are key drivers of science skepticism. (This is not to say there aren't *other* key drivers). On John's analysis, science deniers do not take issue with the scientific method or process itself but rather with the grounds on which scientists (and science communicators) make scientific claims. They typically think that the best explanation for the claims that scientists make is not that they are well-established but that they serve the scientists' interests.

If this analysis of science skepticism is along the right lines—and I think it is—then it points to a serious problem with John's views about science communication. One obvious risk of science communication strategies that ignore the complexities and uncertainties in the scientific process, or the compromises and fudges that lead to claims being accepted by the scientific community, is that this may feed the suspicions of science deniers by exacerbating—and in some ways justifying—distrust in scientists and scientific institutions. The result may be entrenchment of skepticism. This is perhaps a plausible explanation of some strains of climate science skepticism. But it seems a particularly plausible explanation of a prominent strain of science skepticism that has sprung up around the COVID-19 pandemic. A perceived lack of transparency on the part of scientists and

science communicators (public health officials, scientists in public health roles, working scientists) has been cited in support of narratives on which political interests are the primary drivers of public health messaging. Even if these narratives are wrong, their existence, and their effectiveness within particular communities, serve to highlight the risks of a science communication strategy that downplays values like honesty and transparency and the ways in which such a strategy can itself be an obstacle to the uptake of scientific testimony if lack of honesty and transparency leads to increasing distrust in scientists and scientific institutions.

Of course, there are many obstacles to the uptake of scientific testimony, and many reasons why people might lack the trust in scientific institutions required for that testimony to receive uptake. It may well be that insufficiently careful science communication strategies are not the main obstacle. I myself doubt that they are the main obstacle: analyses of common forms of science skepticism such as vaccine hesitancy typically identify lack of trust in scientific institutions as the main obstacle, and this lack of trust is often due to serious failings—sometimes *very* serious failings—on the part of these institutions (Goldenberg, 2021). But, if John's analysis of science skepticism is correct, then misguided science communication strategies *contribute* to this lack of trust, even if they are not the main cause of it. If scientists and scientific institutions are going to hold on to (or regain) public trust, a lot needs to change. One thing that needs to be abandoned is the idea that the job of the science communicator is simply to get the public to believe true things.

5. Concluding Remarks

Let me tie some of the strands together. In this paper, I have (i) discussed what persuasion is and what distinguishes it from coercion, indoctrination, and manipulation, (ii) distinguished between rhetorics of persuasion that utilize the objective style and rhetorics that, in one way or another, depart from the objective style, (iii) integrated the psychology of persuasion with the epistemology of persuasion by using the tools of epistemology to address the prospects of different rhetorics of persuasion bringing about rational changes of attitude, and (iv) highlighted some ethical issues that, at least in broad outline, apply to rhetorics of persuasion that utilize the objective style as well as rhetorics that depart from it. While this has hardly been an exhaustive treatment of these issues, I hope that I have at least demonstrated the value of thinking about persuasion in a systematic way, and the importance of bringing debates about the epistemology and ethics of persuasion into contact with each other, and with the literature on the psychology of persuasion.

Robin McKenna is a Senior Lecturer in Philosophy at the University of Liverpool. He is also a Senior Research Associate at the African Centre for Epistemology and Philosophy of Science at the University of Johannesburg. His research focuses on applied and social epistemology, which he approaches from the perspective of non-ideal theory, as outlined and defended in his 2023 book *Non-Ideal Epistemology*, published by Oxford University Press.

References

Ahlstrom-Vij, K. (2013). *Epistemic paternalism: A defence*. Palgrave-Macmillan.

Anderson, E. (2011). Democracy, public policy, and lay assessments of scientific testimony. *Episteme*, 8(2), 144–164.

Bovens, L. (2009). The ethics of nudge. In T. Grüne-Yanoff, & S. O. Hansson (Eds.), *Preference change: Approaches from philosophy, economics and psychology: Theory and Decision Library* (pp. 207–219). Springer Netherlands.

Carter, J. A., & McKenna, R. (2020). Skepticism motivated: On the Skeptical import of motivated reasoning. *Canadian Journal of Philosophy*, 50(6), 702–718.

Cialdini, R. B. (2001). *Influence: Science and practice* (4th ed.). Allyn and Bacon.

Cook, J., van der Linden, S., Maibach, E. H., & Lewandowsky, S. (2018). *The Consensus Handbook*. http://www.climatechangecommunication.org/all/consensus-handbook/.

Coppock, A. (2022). *Persuasion in parallel: How information changes minds about politics*. University of Chicago Press.

Dillard, J. P., & Shen, L., (Eds.). (2012). *The SAGE handbook of persuasion* (2nd ed.). Sage Publications.

DiPaolo, J., & Simpson, R. M. (2016). Indoctrination anxiety and the Etiology of belief. *Synthese*, 193 (10), 3079–3098.

Dutilh, C. (2023). The (higher-order) evidential significance of attention and trust—Comments on Levy's bad beliefs. *Philosophical Psychology*, 36(4), 792–807.

Fishbein, M., & Azjen, I. (1975). *Belief, attitude, intention, and behavior: An introduction to theory and research*. Addison-Wesley.

Gerken, M. (2022). *Scientific testimony: Its roles in science and society*. Oxford University Press.

Godber, A., & Origgi, G. (2023). Telling propaganda from legitimate political persuasion. *Episteme*, 20(3), 778–797.

Goldenberg, M. J. (2021). *Vaccine hesitancy: Public trust, expertise, and the war on science*. University of Pittsburgh Press.

Hausman, D. M., & Welch, B. (2010). "Debate: To nudge or not to nudge." *Journal of Political Philosophy* 18(1), 123–136.

John, S. (2018a). Epistemic trust and the ethics of science communication: Against transparency, openness, sincerity and honesty. *Social Epistemology*, 32(2), 75–87.

John, S. (2018b). Scientific deceit. *Synthese*, 198(1), 373–394.

John, S. (2019). Science, truth and dictatorship: Wishful thinking or wishful speaking? *Studies in History and Philosophy of Science Part A*, 78, 64–72.

Kahan, D. (2014). Making climate-science communication evidence-based--all the way down. In M. Boykoff, & D. Crow *Culture, politics and climate change* (pp. 203–220). Routledge.

Kahan, D., Jenkins-Smith, H., & Braman, D. (2011). Cultural cognition of scientific consensus. *Journal of Risk Research*, 14(2), 147–174.

Levy, N. (2021). *Bad beliefs: Why they happen to good people*. Oxford University Press.

Lewandowsky, S., & Oberauer, K. (2016). Motivated rejection of science. *Current Directions in Psychological Science*, 25(4), 217–222.

Linden, S., Leiserowitz, A., Rosenthal, S., & Maibach, E. (2017). Inoculating the public against misinformation about climate change. *Global Challenges*, 1(2), 1600008. https://doi.org/10.1002/gch2.201600008.

Maier, M., František Bartoš, T. D. S., Shanks, D. R., Adam, J. L. H., & Wagenmakers, E.-J. (2022). No evidence for nudging after adjusting for publication bias. *Proceedings of the National Academy of Sciences*, 119(31), e2200300119. https://doi.org/10.1073/pnas.2200300119.

Maio, G. R., Haddock, G., & Verplanken, B. (2019). *The psychology of attitudes and attitude change* (3rd ed.). Sage Publications.

McKenna, R. (2020). Persuasion and epistemic paternalism. In G. Axtell & A. Bernal (Eds.), *Epistemic paternalism: Conceptions, justifications, and implications* (pp. 91–106). Rowman and Littlefield.

McKenna, R. (2023). *Non-ideal epistemology*. Oxford University Press.

Mertens, S., Herberz, M., Ulf, J. J. H., & Brosch, T. (2022). The effectiveness of nudging: A meta-analysis of choice architecture interventions across Behavioral domains. *Proceedings of the National Academy of Sciences*, 119(1), e2107346118. https://doi.org/10.1073/pnas.2107346118.

Mitchell, T., & Douglas, T. (2024). Wrongful rational persuasion online. *Philosophy and Technology*, 37(1), 1–25.

O'Keefe, D. J. (2016). *Persuasion: Theory and research*. Sage Publications.

Petty, R. E., & Cacioppo, J. T. (1986). The elaboration likelihood model of persuasion. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (Vol. 19, pp. 123–205). Academic Press.

Petty, R. E., Wheeler, S. C., & Tormala, Z. L.. (2003). Persuasion and attitude change. In *Handbook of psychology* (pp. 353–382). John Wiley & Sons, Ltd. https://doi.org/10.1002/0471264385.wei0515.

Riley, E. (2017). The beneficent nudge program and epistemic injustice. *Ethical Theory and Moral Practice*, 20(3), 597–616.

Simoniti, V. (2021). Art as political discourse. *British Journal of Aesthetics*, 61(4), 559–574.

Thaler, R. H., & Sunstein, C. R. (2008). *Nudge: Improving decisions about health, wealth, and happiness*. Yale University Press.

Tsai, G. (2014). Rational persuasion as paternalism. *Philosophy and Public Affairs*, 42(1), 78–112.