

Nudging for Judging that p ¹

Oscar Piedrahita and Matthew Vermaire

Note: This is the pre-print for an article forthcoming in The Philosophical Quarterly. Please cite the published version.

Recent work in social epistemology has begun to make use of the behavioral-scientific concept of the *nudge*, but without sustained attention to how it should be translated from behavioral to epistemic contexts. We offer an account of *doxastic nudges* that satisfies extensional and theoretical desiderata, defend it against other accounts in the literature, and use it to clarify ongoing discussions of how nudges relate to reason-giving, knowledge, and autonomy.

1. Introduction

With the epistemic perils of our societies looming large—public rejection of expert consensus, runaway conspiracy theorizing, the polarization of electorates around alternative sets of facts—well might concerned citizens look to their tools of belief management. What can be done about the bad views of our neighbors?² There's rational argument and careful fact-checking, of course; but we've grown hesitant to trust them for systematic results (see Walter et al. 2020; Ecker et al. 2022; Roozenbeek and van der Linden 2024, 110-111). There's government coercion: suppression of deviant literature, reeducation camps for the wrongheaded, mind-control via brainwashing or hypnosis; but these techniques have a coarse and unenlightened feel to them. If we want to take a lighter touch, it's natural to wish for further options, ones that don't rely on people's rational appreciation of evidence but that will still deserve the consideration of liberal democracies. Enter the *doxastic nudge*: the intervention popularized in behavioral economics (Thaler and Sunstein 2021/2008), now in epistemicized form.

¹ [REDACTED]

² See Funke and Flamini (2018) for an overview of different anti-misinformation strategies adopted by countries around the world, ranging from education campaigns and making it illegal to misinform, to deleting content, blocking media, imposing fines, and issuing prison threats. See also Roozenbeek and van der Linden (2024, 90-92).

It's that epistemicizing process that concerns us, in this paper. Authors have recently begun to explore the epistemology of nudges: the worries of epistemic injustice or epistemic laziness that may attend them (Riley 2017, Meehan 2020); the connection between nudges and epistemic autonomy or epistemic paternalism (Levy 2022, chs 5 and 6; McKenna 2020; Miyazono 2023); and the possibility of nudges leading to knowledge or justified belief (cf. Grundmann 2023, 2021; Levy 2021; Matheson and Joly Chock 2021). However—partly because of ambiguities in the original literature on nudges, and partly because of problems with the translation into epistemology—a serviceable account of **doxastic nudges**, nudges for beliefs, has been lacking. As we argue below, available proposals fail to pick out a useful category of epistemic interventions; and they therefore threaten to obfuscate rather than clarify further discussions about the ethics and epistemology of nudging.

We therefore stage a little epistemic intervention of our own. In the next section, §2, we introduce the standard view of nudging as it has been applied to behavior modification, and survey some attempts—faulty ones, we argue—to translate this view into epistemology. In §3, we lay out our alternative, the **Normative View of Doxastic Nudges**, designed to correct the failings of its predecessors. This view understands doxastic nudges as, roughly, non-rational but non-controlling influences on doxastic states. In §4 and §5, we show how our view can clarify two debated questions about doxastic nudges: first the question of whether they work via reason-responsive processes, and then the question of whether nudged beliefs can have the standing of knowledge.

Before we start, a few clarifications. Nudges are, broadly speaking, certain ways of influencing people, and one important source of interest in them is their potential for *benevolent* influence. From Thaler and Sunstein's work on, they have been associated with libertarian paternalism: through them, perhaps, citizens could be brought to act in their own interests, while retaining their free choice. Along with other authors, however, we distinguish nudging as such from nudging as a libertarian-paternalist tool (cf. Saghai 2013, 488, Hansen 2016, 156, Mongin and Cozic 2018). In our usage, a nudge need not be intended to benefit its targets: this allows us to say, for instance, that businesses and advertisers can nudge members of the public even without benevolent intent. In fact—a second clarification—we will not assume that an

instance of nudging must be intentional at all. Along with other ways of influencing people, like giving them reasons or leaving them without a choice, we may often do it quite inadvertently, though ethics (if not epistemology) will understandably have a special interest in the intentional cases.

2. The Nudge, Epistemicized

2.1 *The Standard View of Behavioral Nudges*

Thaler and Sunstein (2021) conceptualize and advocate nudges by drawing from work on behavioral sciences, which suggests that people's behavior can be reliably influenced by intervening in the circumstances of their choices without directly altering the benefits, restrictions, or constraints already salient to them. This characterization is supplemented by a large list of examples that circulate in the literature, such as the following.

Defaults: Enrolling people in a retirement savings plan or an organ donor registry by default (while leaving them free to opt out) makes them more likely to participate than if they are required to actively opt in (Thaler and Sunstein 2021, 16, 253 et passim).

Eye Level: Placing certain foods at eye-level in cafeterias makes it more likely that customers will pick out those products (Thaler and Sunstein 2021, 1-3).

Frames: Presenting a medical procedure as having a 90% survival rate, rather than a 10% mortality rate, might make doctors more likely to recommend it and patients more likely to consent to it (cf. Thaler and Sunstein 2021, 39).

Decoy: In a cinema, adding a \$6.50 medium popcorn option alongside a \$4 small option and a \$7 option large may increase orders for the large size, which appears to offer better value (cf. Hansen 2016, 166).

What do all these cases have in common? The standard, psychological view is that nudges operate on particular types of cognitive processes. The findings from behavioral science suggest

that people's choices are influenced less by reasoned, reflective assessment of their choices and information, and more by automatic and intuitive tendencies, which in some contexts deviate from expected economic rationality. A popular version of this standard view, which traces back to Thaler and Sunstein's seminal work, makes reference to Kahneman's (2011) dual process theory: nudges exploit System 1's fast, unconscious, and non-deliberative cognitive processes, it says, and so bypass the slow, conscious, and deliberative, cognitive processes of System 2 (cf. Thaler and Sunstein 2021, 40ff.; Sunstein 2014, ch. 1).

Another variation of this view, from Yashar Saghai (2013), is particularly useful for understanding epistemological discussions of nudging. Saghai's account does not directly invoke dual-process theory, but picks out "shallow" cognitive processes—that is, processes that work quickly, consume few cognitive resources, and are not fully deliberative. Saghai (2013, 491) then proposes the following definition:

Saghai's View of Nudging: "A nudges B when [i] A makes it more likely that B will φ , [ii] primarily by triggering B's shallow cognitive processes, while A's influence [iii] preserves B's choice-set and [iv] is substantially noncontrolling (i.e., it preserves B's freedom of choice)."

Condition (i) identifies the nudge as a species of influence on behavior, and condition (ii) picks out the psychological processes through which that influence operates. Conditions (iii) and (iv) in set nudging apart from other ways of influencing behavior, like financial disincentives or physical coercion, precisifying Thaler and Sunstein's characterization of nudging as working "without forbidding any options or significantly changing economic incentives" (2021, 8). Thus, placing preferred products at eye level in the cafeteria, as in the *Eye Level* case above, can constitute nudging customers to buy them; but removing alternative products from the shelves altogether will not. Likewise, playing Christmas music in a department store may nudge shoppers to make more purchases, but, as condition (iv) specifies, the influence must not be made too difficult to resist. If a behavioral intervention works via brainwashing or mind-control—or, less exotically, by exploiting a target's compulsions, addictions, or phobias—it will

be too controlling to count as a nudge. Nudges work through the un-compelled choices of the nudgee (cf. Bovens 2013, 495).

To take stock, the standard, psychological view of behavioral nudging says that nudges work by operating on certain kinds of cognitive processes: shallow, fast, cheap ones, whether incompletely deliberative or non-deliberative altogether. They differ, then, from forms of persuasion that work by targeting the nudgee's full-blown reflective capacities, such as giving an argument, demonstrating, or explaining. Nevertheless, they can plausibly be thought to preserve her freedom of choice, since their influence leaves her set of options in place, is easily resistible, and doesn't exclude the nudgee's self-determination.

In the rest of this section, we move to examining how views like Saghai's might serve as a blueprint for views of nudging doxastic states, more specifically. Is the standard, psychological view of behavioral nudging fit for epistemological translation?

2.2. *Doxastic Nudges*

It's not difficult to see why nudging would be of interest for epistemologists. Doxastic states can clearly be affected by nudges. In the *Frames* case, for instance, where the framing of statistics affects whether patients will consent to a procedure, it seems likely that it also affects whether they believe that the procedure is safe. (Indeed, this is plausibly *why* it affects whether they consent to the procedure.) Moreover, since so much depends on our doxastic states—in which category we include not only beliefs and disbeliefs but also suspension of judgment, credences, and even the lack of any of these attitudes—any systematic way of modifying them will be worth attending to. But how exactly should we think of such doxastic nudges? By focusing on three initial attempts to define the category, we will see some of the challenges of translating nudges into epistemology.

To start with, it might seem that an account of doxastic nudging follows directly from the existing account of behavioral nudging. For instance, consider this toy view of doxastic nudging, which simply fills in φ (from Saghai's definition above) with a doxastic value.

The Doxastic Result View: A doxastically nudges B iff A (i) makes it more likely that B will hold some doxastic state, (ii) primarily by triggering B's shallow cognitive processes, while A's influence (iii) preserves B's choice-set and (iv) is substantially noncontrolling.

Under this view, doxastic nudges are just nudges with doxastic outcomes. A simple picture; but it seems too weak to pick out an interesting category of interventions. After all, nudges will typically have plenty of downstream effects on the nudgee's doxastic states just as a matter of course. If the arrangement of the cafeteria nudges you (primarily through your shallow cognitive processes) to pick out the veggie option, because it's placed at eye level, you will very likely go on to believe that you've picked out the veggie option. But that belief will be prompted in a perfectly normal way by your experience of picking out the veggie option. It doesn't seem worth distinguishing from your other, non-nudged beliefs: for instance, it doesn't seem to circumvent your rational faculties, or to raise concerns about your lack of autonomy in forming that belief. Even if a nudge figures in the causal chain that leads to a belief, then, we might not want to call it a doxastic nudge.

A second problem for the Doxastic Result View is that the conditions of preserving the choice-set (iii) and being non-controlling (iv), which are integral to Saghai's definition, aren't clearly relevant in the epistemic domain. In this context, they suggest an unnecessary form of doxastic voluntarism such that beliefs are formed through choices under the free control of the believer. According to reigning epistemological orthodoxy, though, belief ordinarily isn't like that. Even if I make some belief positively irresistible to you—say, by giving you overwhelming evidence for it—it doesn't show I've coerced you or diminished your freedom; at least, not in any normatively important way. So, these clauses seem poorly motivated.

In light of these two problems, one might modify the Doxastic Result View by dropping ill-motivated clauses (iii) and (iv), and requiring that the doxastic state in question be *directly* produced by shallow cognitive processes, rather than simply being downstream from them.

The Shallow Processes View: A doxastically nudges B iff A makes it more likely that B will hold some doxastic state, primarily by triggering shallow cognitive processes in B that directly produce that state.

The Shallow Processes View blocks the veggie-option counterexample and avoids suggesting doxastic voluntarism, so it is an improvement over the Doxastic Result View. However, it fails to exclude very normal, paradigmatically rational ways of influencing beliefs, like showing someone something.

Obvious Phone. You've got the new iPhone. As you converse with your friend, you casually take it out of your pocket and hold it at eye level, causing her to believe that you have the new iPhone.³

This case satisfies the Shallow Processes View: you influence your friend's doxastic states by triggering her (fast, cheap, non-deliberative) tendency to believe what she sees. However, this intervention again seems to us the wrong kind of thing to distinguish as a nudge. Like the belief that you've picked the veggie option, in the earlier objection, your friend's belief that you have the new iPhone seems perfectly ordinary, rational, and above-board. By showing her the phone, you've simply given her ample evidence for that belief. But nudging is supposed to be a form of influence that is distinct from just giving people new information (cf. Hansen 2016, 157; Schmidt and Engelen 2020, sec. 2; Parmer 2022, 1203). So this view not only overgenerates doxastic nudges, but threatens to obscure the very concept of them.

These problems with the Doxastic Result and Shallow Processes Views indicate that there's a genuine question about how to articulate an account of doxastic nudging. Straightforward behavioral nudges like "anti-littering nudges" or "nudges to save money" could be categorized easily by plugging in a value for φ in Saghai's definition, but doxastic nudges will need more careful consideration. Thomas Grundmann has recently proposed an account of doxastic nudging that aims to fill this need, while using Saghai's definition as a template:

³ [REDACTED]

Grundmann’s View: A doxastically nudges B iff A (i) makes it more likely that B will hold some doxastic state (ii) by triggering the nudgee’s shallow cognitive processes, while her influence preserves both (iii) the represented facts and (iv) the nudgee’s rational control concerning the doxastic attitude (see Grundmann 2023, 210).⁴

In contrast to the other two views above, Grundmann’s View makes two important modifications to Saghai’s definition. On the one hand, Saghai’s “non-controlling” requirement is translated in clause (iv) into a requirement of “rational control,” which Grundmann understands as not removing “the nudgee’s ability to revise these states in the light of further evidence” (2023, 211). This distinguishes doxastic nudging from indoctrination and brainwashing without relying on doxastic voluntarism. On the other hand, instead of requiring the preservation of a choice-set, Grundmann holds that doxastic nudges preserve the facts represented by the nudgee prior to the intervention. He does not elaborate on this latter condition (iv), importantly, but it could be interpreted in a way that blocks the *Obvious Phone* case. Showing your friend a phone alters the facts she represents. Before you took it out of your pocket, she might not have had any evidence bearing on the question of whether you have the new iPhone; but now she does. So, she wouldn’t satisfy condition (iii), so construed. That condition could help distinguish doxastic nudging from simply providing or withholding information.

If this “represented facts” condition blocks the *Obvious Phone* case, though, it does so at great cost: practically no doxastic interventions will leave the facts represented by the nudgee totally unaltered. Take the case of *Decoy*, above: if you introduce an overpriced medium-sized popcorn option at your concession stand, in hopes of customers’ believing that the large option is good value, you’ll be changing the facts they represent: now they’ll know that there’s a medium popcorn option. Similarly, take nudges that trade on positive associations and the affect heuristic:

⁴ Strictly speaking, Grundmann offers these conditions for what he terms “brute intentional nudges.” But the distinctions between brute and non-brute nudges and between intentional and non-intentional nudges aren’t significant for our discussion, and we therefore omit this qualification.

Celebrity Endorsement: A politician's team hires a famous and well-liked actor, a real Tom Hanks type, to read lines in a campaign ad, rather than having an unknown volunteer do it.

We think this casting decision, like many advertising techniques, nudges the ad's viewers to believe its content. The nudge does not leave its targets representing the very same facts they would otherwise have had, though: most obviously, they will now notice that the well-liked actor appears in the politician's campaign ad. Indeed, that's how the strategic move is supposed to work: the alleged efficacy of celebrity endorsements and decoy effects lies precisely in a change in the facts represented by their targets. Indeed, this is the case in Grundmann's own central example of doxastic nudging: Alicia persuades John's supporters that he has committed murder by framing the story of his trial in an unsympathetic light. Although we will discuss this case more below, we can presume that framing his trial unsympathetically involves some alteration of the facts that are represented to them—even if it is only such facts as *John had a pattern of bad-tempered behavior* or *there is an unflattering picture of John in the newspaper*. Clearly, then, Grundmann's "represented facts" condition is too restrictive, at least as we have interpreted it; but apart from this interpretation of it, his view is vulnerable to the *Obvious Phone* case and its ilk. Caught between overly permissive and overly restrictive views, we still have no account of doxastic nudging suitable for addressing epistemological questions.

3. The Normative View of Doxastic Nudges

3.1. Psychological and Normative Strands in the Nudging Literature

As we have argued, we need to keep nudges distinct from other sorts of interventions (publishing nutritional information, say, or using clear photographic evidence in a courtroom trial) that simply provide agents with information relevant for their beliefs and behavior. But how can we do this, without requiring that a nudge provide agents with *no* new information? The natural response, we propose, is to focus on the notion of *relevance* just invoked. What makes nudges special, as a method of persuasion, isn't that their effects are brought about without changing the information their targets have, but that those effects aren't explained by

the relevance of that information to what an agent should believe or do. In other words, nudges (*qua* nudges) do not achieve their effects by virtue of giving agents *reasons* to believe or act. (It may be true, nevertheless, that they do give some reasons—more on this below). Thus, the way forward, we claim, is to see that the concept of a nudge is in fact a normative one: we can make sense of it only with reference to talk of reasons.

This claim puts us at odds with most treatments of nudging (whether doxastic or behavioral) since Thaler and Sunstein. As we saw in section 2.1, those treatments have tried to delineate their subject matter in strictly psychological terms: nudging, they say, is what you get when effects are achieved through particular cognitive processes. The arguments given in the previous section, however—which we develop more fully in other work⁵—raise doubts for this non-normative psychological strategy: the boundaries of a useful “nudge” category will need to be sensitive not just to the cognitive mechanisms that interventions trigger but to the rationalizing force they do or don’t provide; and when these two factors come apart, we think the boundaries of the category should be fitted to normative features rather than to cognitive-psychological ones. Moreover, this sensitivity to normativity isn’t just a quirk of intuitions independent from the rest of the nudge literature. It accords with the strong normative overtones and underpinnings that the concept of nudging has had since it debuted; for, whatever the definitions of nudges have involved, their interest to theorists has long been connected to their distinctness from *rational* ways of influencing belief.⁶

Thaler and Sunstein’s book gives pride of place to dual process theory, treating nudges as working through fast and automatic ways of thinking rather than slow and deliberate ones (see 2021, ch. 1). The normative component shows up, in their work, less in how nudges are defined than in how they are characterized: nudges are the sorts of things that change what

⁵ [REDACTED]

⁶ For definitions of nudges, see Thaler and Sunstein (2021), Saghai (2013), Hansen (2016), Engelen and Nys (2020), Congiu and Moscati (2021), and Parmer (2023). Hansen’s (2016) definition combines cognitive and normative concerns by describing nudges as operating via “cognitive boundaries, biases, routines and habits ... posing barriers for people to perform rationally in their own declared self-interest” (158). Our proposal, like Hansen’s, contrasts the influence of nudges with that of reasons and rationality (cf. Hansen 2016, 162-169). Our view is simpler and more purely normative, though, since we drop the requirement that nudges work by means of any particular cognitive process or feature of cognition.

humans do, but “would be ignored by Econs” —i.e., by an imaginary race of idealized beings who are “not influenced by such ‘irrelevant’ factors as the order in which options are displayed” (Thaler and Sunstein 2021, 12; cf. Hansen 2016, 160-163). Similarly, though Saghai’s (2013) definition of nudges includes no clearly normative conditions, he contrasts nudges with methods of “rational persuasion” (490–91) as with a non-overlapping category. Grundmann, too, introduces nudges with normative talk —they typically influence people “without giving reasons” and “by using non-rational cognitive mechanisms” (2023, 208)—while declining to incorporate the notion of non-rationality into his definition.

In fact, then, normative and psychological strands have long been tangled together in accounts of nudging. This tangle has been obscured by an implicit assumption that it is only through deliberative reasoning—a particular kind of psychological phenomenon—that we manifest rationality. At least, this assumption is normally only implicit; Grundmann (2023, 210) is upfront about using “reasoning” interchangeably with “rationality.”⁷ Implicit or explicit, though, this kind of bridge principle between psychology and normativity doesn’t survive scrutiny. Arpaly and Schroeder (2012) have argued that our rational behavior, our responsiveness to reasons, must extend beyond the results of deliberation, since deliberation itself (even if unpremeditated) can be rational or irrational. Levy (2019; 2021) has pressed similar points in the nudging literature, specifically, and the discussion in the previous section helps to dramatize them: in cases like Obvious Phone, our “shallow” or “fast” cognitive mechanisms are precisely the mechanisms that enable us to be responsive to the reasons we have. The easy identification of shallow mechanisms with non-rationality is perhaps an artifact of the heady turn of the millennium, when work in behavioral psychology had made a vivid case for the fallibility of our cognitive heuristics. While this might be an important lesson, however, it should not be overlearned. Perhaps “fast” ways of thinking present especially good opportunities for non-rational influences on people’s beliefs. Often, though, there’s nothing non-rational about believing what comes easy to believe. The psychological and normative

⁷ Maybe Grundmann’s use of “reasoning” is meant to include more than only explicit deliberative reasoning, but it is hard to imagine that word’s extension including, for instance, the formation of perceptual beliefs.

strands of past treatments need to be kept distinct, therefore; and we'll do best, we propose, if our view of nudging hews to the normative strand.

3.2. *The Normative View Presented*

Let that suffice for preamble. Here is our proposal.

The Normative View of Doxastic Nudges: A doxastically nudges B to have doxastic state D iff

- (i) A makes B more likely to have D by intervening in B's circumstances,
- (ii) A preserves B's rational control over D,
- (iii) the effect of A's intervention is not explained by its rational force.

By and large, the first two clauses of the *definiens* echo previous definitions. First, condition (i), the **probabilifying condition**, marks out the *definiendum* as a species of doxastic influence that, like a behavioral nudge, works by altering the target agent's context. Next, condition (ii), the **control condition**, distinguishes the target species of intervention from such coercive measures as brainwashing, hypnosis, or alien mind-control. If you can induce a belief directly by administering electrical pulses to the cerebral cortex, it won't be a doxastic nudge. The key to the normative view of doxastic nudges, though, its chief innovation, is naturally its **normative condition**, condition (iii). Whereas the control condition marked nudges off from coercive, brute-force methods of belief modification, the normative condition distinguishes them from the very different kind of intervention that is the providing of reasons.

Some care is called for here, however. We think it would be a mistake to characterize doxastic nudges as giving agents no reasons at all for the doxastic states they aim at. Reasons of *some* degree of strength are just too easy to come by: see Schroeder (2007, 95–96) on the reasons you have to eat your car. It's only natural that since, *pace* Grundmann, nudges can change the information presented to an agent, they can do something to modify the reasons they possess, and so to make doxastic states more or less rational. Consider again *Celebrity Nudge*, where a well-liked actor is hired to read lines in a campaign ad. We said in the last section that this

nudge changes the represented facts for the audience; they will now know, for example, that the actor appears in the politician's campaign ad. It would seem dogmatic, too, to think that this change of information has *no effect at all* on the reasons the ad's audience has for belief. That this well-liked celebrity reads the lines is a new piece of evidence, which may plausibly count in favor of the politician and her message. If the actor is well-liked and respected, it's some (weak) evidence in favor of his being a trustworthy person; and that he would be willing to deliver the lines in the ad is some (weak) evidence that they're true. But these justifications feel a bit tiresome and peripheral—and that's the point. The real story, we may presume, is elsewhere. The rationalizing force of the nudge is not what explains the magnitude of the effect it will have, if it works as the campaign team hopes. They hope for it to work better, for instance, than just reliably informing audiences that the actor was hired to read those lines. Plausibly, the team is trying to marshal further, non-rational effects to its cause: the comforting, reassuring tones of that sonorous voice, say; the association of the actor with the kind, upright characters that made him famous. Doxastic nudges may in fact contribute something to the rationality of the doxastic states they aim at; but, insofar as they really are nudges, it's not *by* making doxastic states rational that they help to produce them (cf. Hansen 2016, 164–65). This is what our condition (iii) says.

3.3. *Extensional and Theoretical Advantages*

Our approach to the Normative View of doxastic nudges has been primarily via the method of cases: the psychological definitions considered above, we argued, deliver the wrong verdicts for cases like *Obvious Phone*, *Decoy*, or *Celebrity Nudge*. Not that we're relying on robust folk intuitions about the extension of "nudge," let alone "doxastic nudge"; rather, such cases help to show that the categories picked out by these definitions are not theoretically interesting. The Normative View, though, passes the tests they failed. Showing your friend your new phone is determined not to be a nudge, just as we want, since it so amply rationalizes your friend's belief that you have a new phone. It fails to satisfy our normative condition. On the other hand, hiring a beloved actor to read your political ad, or introducing a decoy option, will plausibly satisfy

the normative condition, even though it changes the facts represented by the agent and even (to some extent) her reasons for the intended doxastic states. Our extensional problems have been addressed.

Now that the account has been put on the table, though, we would like to point out that its extensional advantages are not merely accidental, but closely related to other theoretical strengths of our view. First, as we have emphasized, our account does not rely on particular divisions of cognitive mechanisms, like that posited between System 1 and System 2 processes. That's a strength, because the psychological aptness of such categorizations may be questioned (see Levy 2019, 285, n. 4; Gigerenzer 2015, sec. 4.2). It's nice not to be held hostage to the results of such empirical challenges. However they shake out, it is exceedingly plausible that people's doxastic states can be affected in ways not fully explained by the force of the reasons they are given. That's not a posit of flash-in-the-pan pop psychology: polities have been concerned with the possibility since Athens condemned Socrates for "making the weaker argument stronger" (*Apology* 19b); and it will be worth theorizing about long after dual-system theory has been decisively confirmed or rejected.

That leads us to another theoretical advantage. As we argue at greater length in [REDACTED], it is a normative conception of nudging that carves at the joints, cleanly picking out a category of interventions that operates neither coercively nor strictly in accord with reason—a kind of gray area for behavior- and belief-modification. Because of this, the Normative View positions us well to address the distinctive kinds of ethical questions that arise in that shadowy realm. Through the control condition, we rule out the most obviously objectionable kinds of interventions in doxastic states. We're not worried, here, about jackbooted goons force-feeding you doxastic states in the style of *1984*. (Its protagonist is tortured and drugged until he really believes that his interrogator is holding five fingers up, rather than four.) But the normative condition distinguishes our category, too, from the unobjectionable benevolence of a government that publishes reliable information and otherwise leaves you to yourself. As interventions that operate non-rationally, nudges raise special worries about autonomy. Do they make their targets too much the objects of another's will, failing (in Kantian terms) to treat rational agency as an end in itself? Do they put the people of

democratic states too much in the place of children, to be managed and maneuvered by their enlightened minders rather than respected as equals? (See Grüne-Yanoff 2012, Hausman and Welch 2010, MacKay and Robinson 2016, Wilkinson 2013, White 2013.) These are the questions that arise naturally for a kind of intervention that circumvents reason-giving. In contrast, they aren't so clearly relevant to some intervention just because the cognitive processes it triggers are faster or less resource-intensive than others, or evolved earlier in the Pleistocene era.

Our account of doxastic nudges is not designed to settle such ethical questions. (It is a “normative view” because it makes reference to rationality, not because it assigns a uniform normative status to every nudge.) They simply aren't to be settled at the general level, we suspect (cf. Schmidt and Engelen 2020). We think they are the right questions to ask, however, and the Normative View cleanly identifies the set of interventions for which they are appropriate. A normative treatment of our topic best serves the normative interests that bring it to prominence in the first place.

A preliminary argument for that last claim has now been given. Substantiating it more fully will have to involve putting the Normative View to work, demonstrating its theoretical serviceability. Accordingly, in the next two sections, we make a start at applying our picture: we use it to clarify two debated questions about the rationality and epistemic potential of doxastic nudges.

4. Non-Rational Influence through Reason-Responsive Mechanisms

We hold that it is the non-rational nature of a nudge's influence, and *not* the particular class of cognitive mechanisms by which that influence may operate, that is essential to it as a nudge. This reverses the usual ordering: as we have seen, treatments of nudging have typically put the emphasis on particular psychological processes, while perhaps assuming (wrongly, we said) that, because these processes are non-deliberative, they are non-rational. But our view also brings us into direct conflict with some alternative accounts, which—far from assuming that nudges will operate non-rationally—seek to show that their influence is as rational as that of any other epistemic intervention. Neil Levy has argued particularly forcefully that “nudges

work by providing reasons to agents” (2021, 44). Our Normative View rules out this kind of rehabilitation project from the start, but we don’t wish to deny the insights of its proponents. Rather, we hope to clarify the ongoing debate by expanding on two of the points we’ve emphasized: first, that the rationality of a doxastic state is no more than weakly correlated with the sort of cognitive mechanisms that produced it; and second, that there’s a real difference between just giving *some* reason for a doxastic state and giving *enough* of a reason to account for the intervention’s effects as a rational response.

On the first point, it’s fundamental to our approach that the status of some intervention as a nudge doesn’t depend just on the kind of cognitive mechanism through which it takes effect. The very same process through which, in one case, a nudge brings about an irrational doxastic state may in another case be the process enabling a perfectly rational response to one’s evidence. We’re not on the hook, therefore, for the claim that nudgeable mechanisms aren’t reasons-responsive. That’s Grundmann’s view: nudging works through cognitive heuristics, rather than other processes, “and is thus unresponsive to reasons” (2021, 34). If you form a belief or make a decision by deferring to default assumptions or the conversational common ground (2023, 210), imitating the behavior of your peers (2023, 213), or using your emotional reactions as shortcuts around more difficult questions (2023, 213), he thinks you just can’t be responding to your reasons: processes like these, he’s confident, involve “neither reasons nor reason-responsive mechanisms” (2023, 213). Likewise, use of the “availability heuristic” for estimating frequencies can’t be a way of responding to evidence, for “one cannot infer that something is more frequent than something else from the fact that it is more easily available [to memory]” (2021, 29). This seems to us a revealing mistake. Of course, the fact that one more easily retrieves instances of one type than another from memory doesn’t *entail* that instances of that type are more frequent; but this hardly shows that it constitutes no evidence at all. On the contrary, in probabilistic reasoning, by Bayesian logic, there is a very straightforward evidential relation here. (All else equal, one would expect that more frequent instances would be more frequently experienced, and so more easily accessible in memory; and so, all else equal, greater accessibility is confirmation of greater frequency.) In a similar spirit, Levy (2021) argues that the beliefs and actions of our peers, and our own affective responses, are perfectly eligible to serve

as rational factors in our judgments: “consensus is strong evidence in favor of a claim,” and “relying on our gut feelings is typically a way of getting things right” (Levy 2021, 46). As with forming simple perceptual beliefs, so here: there’s often nothing unreasonable about believing what comes easy to believe.

But, if the mechanisms by which nudges operate are typically responding to reasons, this doesn’t show that they’ll systematically result in justified beliefs and rational actions. Here’s where the second point we’ve highlighted comes in: just having *a* reason for something comes cheap. (Recall the reasons you have to eat your car.) Even if your doxastic state is “reasons-responsive” in a weak sense—it’s produced in response to some rational considerations or other (as eating your car might be a response to the need for iron in your diet)—that doesn’t get you far toward a more demanding kind of reasons-responsiveness, a rational sensitivity to the whole *set* of reasons you have. That’s a matter of finding an appropriate balance, assigning to many competing considerations the right weights for them in the context. We shouldn’t focus only on the question that has so far been centered in this debate, then, the question whether nudges typically “provide us with reasons” (Levy 2021, 45), or instead “bypass the nudgee’s reasoning capacity” (Grundmann 2023, 208). If a nudge exercises non-rational influence, it may be not because it operates in a totally reason-free zone but because there is a significant *mismatch* between the degree of its effect and the strength of the reasons it provides. That’s the sort of structure we imagined in *Celebrity Nudge*: if the beloved actor appears in the politician’s ad, it’s *some* evidence for its truth, sure enough; but its effect on the audience might well be out of proportion to its scant evidential force. That, at least, is what the people running the ad campaign will hope for.

Levy’s optimism about nudges seems to extend to this question of proportionality: not only do nudges typically give reasons, but “agents (typically again) respond to these reasons in ways that reflect their reason-giving force” (2021, 44). However, that stronger claim is hard to establish, and there is some cause for suspicion about it. If nudges function as testimony, as on Levy’s picture, and if their effect is in proportion to their reason giving-force, we would expect them to work no better than explicit testimony or official recommendations. But nudges that trade on defaults, at least, don’t pattern this way: default, opt-out organ-donation policies, for

example, have done more than official pleas for more donors have to increase enrollment; defaulting energy consumers into green-energy contracts can be far more effective than asking them if they want to opt in (Thaler and Sunstein 2021, 254, 306). Plausibly, inertia is just a more powerful force than reason, sometimes (cf. Thaler and Sunstein 2021, 305). That doesn't mean that the psychological mechanisms involved in nudges are, as Grundmann claims, "non-rational heuristics rather than reason-responsive processes" (2021, 31)—again, we don't approve of classing psychological mechanisms as one or the other. It does suggest, though, that there's room for predictable gaps between the strength of the reasons a nudge gives and the effect it has on its target. This in turn suggests that the category picked out by our Normative View has a significant extension.

In any particular case, to be sure, it may be up for debate whether a given doxastic intervention counts as a nudge or instead fully rationalizes the effect it produces. And we can even imagine that, in the far future, the balance of psychological evidence and normative epistemology will vindicate the rehabilitation project of Levy et al., more generally: that, as it turns out, it's very hard to influence belief and behavior except in strict proportion to the provision of rational support for them. In that case, the category of nudges that we have delineated would include very few actual phenomena. We are willing to risk our view against such an outcome, however. Besides its seeming unlikely, we also think that, if that's how things really are, the concept of the nudge will be of little interest in any case. As we've argued above and in [REDACTED], it's the normative character of the nudge that makes it worth thinking about in the first place.

5. Epistemic Nudges, without Compromises

5.1. Posing the Question

As a final extension of the Normative View, we consider how it might illuminate a question raised by Grundmann (2023) about the epistemic standing of the doxastic states that doxastic nudges induce. Can those doxastic states constitute justified belief (or justified disbelief or suspension), or knowledge, or understanding? We will focus our discussion on knowledge.

Nudges that produce it, if they are possible, will be in Grundmann’s terms “epistemic nudges” (2023, 210); but there is some *prima facie* reason for suspicion about them. To nudge someone toward a belief seems to be an operation of mere rhetoric. That might be thought a poor foundation for knowledge, as this exchange from Plato’s *Gorgias* (454e) suggests.

SOCRATES: And which sort of persuasion does rhetoric create in courts of law and other assemblies about the just and unjust, the sort of persuasion which gives belief [*pisteuein*] without knowledge [*eidennai*], or that which gives knowledge?

GORGIAS: Clearly, Socrates, that which only gives belief.

Of course, if anyone in such an assembly is positioned to know about the just and unjust independently—a separate Platonic question (see the *Meno*, esp. 99b–e)—then we needn’t think the rhetoric she hears would *remove* that knowledge. Likewise, it should be obvious that being nudged to believe *p* is compatible with knowing *p* (e.g., in cases where you have plenty of evidence for that belief anyways). The more promising question, and the one that will concern us in this section, is whether even a belief that is held by its subject only *because* of a doxastic nudge (persuasion that rhetoric *creates*, in Socrates’ language) could constitute knowledge.

Notice that the interest of this question depends on one’s view of doxastic nudges. If we adopt the Doxastic Result View, for instance, from section 2, it will be trivially easy to come up with epistemic nudges: if you believe that you’ve picked the veggie option, say, as a result of having been nudged to order veggies, you’ll surely know it. Suppose we adopt the Shallow Processes View instead. Then the Obvious Phone case, originally given as a counterexample to the view, will serve just as well to show how easily we get epistemic nudges: your friend can certainly know you have the new iPhone when you show it to her, even though the perceptual processes that produce this belief are shallow and non-deliberative. On Levy’s picture of nudges, too, we can expect the question to resolve itself at once. Nudges are for him sources of implicit testimony (2021, 44); and testimony is a paradigm source of knowledge. Case closed.

It might seem at first that the question of epistemic nudges is no more interesting on our Normative View. Aren't we bound to deny straightaway that there could be such things? For us, the influence of a nudge (*qua* nudge) is not just characteristically but essentially non-rationalizing. Wouldn't a belief produced through such influences have to be irrational, and so fall short of knowledge? Not so, we now argue. Far from foreclosing the question of epistemic nudges, the Normative View can help us find new ways of generating them—and without making the compromises that other strategies have made.

5.2. *Epistemic Nudges without Compromises*

According to Grundmann himself, the question of epistemic nudges comes to rest on the reliability of the nudger's activities. In his example case (214), Political Loyalty, Alicia persuades John's political supporters to believe (truly) that he is guilty of murder, by such nudgy means as printing very unflattering pictures of him when she reports on the trial. Fortunately, Alicia is well-informed and epistemically benevolent. She would only employ such methods to advance the cause of truth. Grundmann therefore judges that the people she succeeds in persuading will meet the conditions for justified belief and knowledge—if these statuses require only method-reliability and method-safety, *and if* methods are externally individuated in a sufficiently fine-grained way. Apart from those conditions, Grundmann is more pessimistic: “[i]f a belief must be based, e.g. on sufficient evidence or on agential virtues for it to count as justified, then nudging to justified beliefs is out of reach” (215). Nudging to knowledge would then be impossible as well. To secure epistemic nudges, it seems, we'll have to settle for pretty low epistemic standards. Evidentialists and virtue theorists need not apply.

Matheson and Joly Chock (2021) respond to Grundmann with a more optimistic take, and offer the evidentialist two ways of making room for epistemic nudges. First (38), she could adopt a non-causal, doxastic account of the epistemic basing relation, on which basing your belief on good evidence for it only requires believing, of that evidence, that it supports your belief. A doxastic nudge could then result in properly based, justified belief: it just needs to produce in you not only the belief that *p*, but also a meta-belief about the evidential support for

p. (In particular, the meta-belief needn't be what *causes* the lower-order belief.) Second, though, suppose the evidentialist sticks with a causal account of the basing relation: a justified subject's belief, she insists, must be causally connected to her having good reasons for it. That connection, Matheson and Joly Chock admit, will be weaker in beliefs triggered by non-rational influences like nudges. However, they maintain, the connection might still be strong *enough*. After all, we surely can't require that every last causal influence on a belief be a good one, for it to be justified; human knowledge isn't pure enough for that. The nudgee's belief can still have the justification required for knowledge, then, "even when a non-epistemic (or even epistemically suspect) factor plays the final triggering role in bringing about the belief in question" (38). If reasons-responsiveness gets you pretty close to believing some propositionally justified proposition, it's okay for non-rational influences to push you across the finish line.

To us, each of these strategies seems dubious. Grundmann's judgment in his Political Loyalty case depends on allowing that a belief formed through blatantly irrational biases can constitute knowledge, so long as the people manipulating those biases dependably lead their marks to the truth. That's counterintuitive, to us. Our doubts aren't much soothed, either, if we add that the bias-manipulators also bestow the kind of meta-beliefs that Matheson and Joly Chock's doxastic basing relation would call for. Suppose, for instance, that a juror's (true) belief in the defendant's guilt is explained simply by her racist prejudice, skillfully exploited by the prosecution; but this prejudice also causes her to think that the courtroom evidence supports that verdict. That's no way to *know* the defendant's guilt, we think, even if the evidence for guilt was actually strong, and even if, say, these prosecutors only ever use racist appeals against guilty parties.

What about Matheson and Joly Chock's second route to epistemic nudges? We agree with them that a justified belief may well be epistemically "impure," so to speak; some slight irrational influences may have contributed to its basis. Plausibly, though, these irrational influences can be washed out when a believer's *good* reasons are more than sufficient both for producing the belief and for justifying it. For example, you might have an irrational tendency to blame it on thievery, when your possessions go missing; but even so, you can still know someone has pickpocketed you when you watch the security footage and see it happen plain as

day. Under those circumstances, you would have believed it quite independently of your silly bias. Mildly irrational believers, then, can still be credited with plenty of justified beliefs. But that doesn't license Matheson and Joly Chock's deeper concession to irrationality. It doesn't show that a belief could be rational even when irrational factors play the necessary "final triggering role" for it, bumping an undecided subject into knowledge.

Our main point isn't that these strategies fail, though. Rather, we deny that they are needed. It isn't true, as Grundmann (2023, 216) thinks, that epistemic nudging "is possible only under specific meta-epistemological assumptions, namely that the relevant methods are externally individuated, and that knowledge, or justified belief, requires nothing but a modal correlation with truth (e.g. safety or reliability)"; nor does it depend on unusual accounts of basing, or the willingness to award justification to beliefs formed because of biases. Guided by the Normative View, we lay out a clearer path to nudge-induced knowledge.

5.3. *Enabling Nudges*

According to the Normative View, nudges change a subject's circumstances to probabilify doxastic states, while preserving her rational control, but without exerting their influence rationally. What can now be pointed out is that these changes are in principle reversible by other nudges. Suppose that, without changing the reasons you have or depriving you of rational control, Ada nudges you from circumstances C1 to C2, and that this makes you more likely to hold doxastic state D. But now introduce another nudger, Zuzu, who (again without changing your reasons or violating autonomy) simply puts you back in C1, instead. Zuzu thus makes you less likely to hold D; or, as we could also say, makes you more likely to hold D⁻ (the doxastic state of not holding D). Zuzu's intervention, returning you to your initial state, will also count as a nudge; and so we have a recipe for what we may call the **de-nudge**: a nudge that serves to cancel a different nudge.

The de-nudge in turn provides a template for nudges that enable their targets to know. For it seems clear, at least, that a nudge could be responsible for *preventing* a target from knowing. Maybe you're quite capable of arriving safely at the knowledge that *p*, in normal

cases, but despite your good evidence you're easily thrown off by certain irrelevant factors. Ada nudges you *not* to believe p , by introducing such an irrelevant factor. But, before this has had a chance to trip you up, the epistemically benevolent Zuzu de-nudges you by removing the irrelevant factor from your cognitive environment, and so makes you more likely to believe p after all. If you then do so, and on the basis of your evidence for p , there is no reason in principle to doubt that you know p . You are of course lucky to know the truth, lucky that Zuzu intervened on your behalf. But this is not the kind of luck that destroys knowledge. (It's more like the luck of someone who narrowly avoids hypoxia, and so can do their math homework unimpaired.)⁸ In such a case, Zuzu's nudge has resulted in your knowledge. It is an epistemic nudge.

This strategy for generating epistemic nudges works by moving beyond the psychological conception of nudges common to our interlocutors, with its corollary that successful nudges will produce beliefs through some nudge-specific cognitive process. If, whenever a nudge resulted in a belief, it had to be manufactured by a special "nudge apparatus" in the brain, then the question we've taken up would probably just depend on whether the nudge apparatus works according to reasons. Thus, Levy (who thinks it does) will easily countenance epistemic nudges, while our other interlocutors (who think it doesn't) struggle to accommodate them. Resisting the psychological conception of nudges, as we've tried to do throughout, helps us see another approach. Instead of giving us new psychological bases for our beliefs, good or bad, nudges can get us to base our beliefs on the reasons we *already* possess (see Grundmann 2021, 32). That's what happens in the case of the successful de-nudge: it enables you to form your belief properly, in the way you would have anyway if Ada's interference hadn't thrown you off track. And the de-nudge is just one convenient dramatization of this approach. More broadly, doxastic nudges can enable knowledge when

⁸ On the difference between luck that excludes knowledge and luck that doesn't, see Pritchard (2005, chs. 5-6).

they make your cognitive environment more forgiving of epistemic weakness, disposing of the red herrings that would otherwise throw imperfect inquirers off the scent.⁹

A doxastic nudge can be more, then, than an impingement on its target's epistemic autonomy. It can aspire to a nobler vocation as, in Kant's phrase, "a hindering of the hindrances to freedom" (*Metaphysics of Morals*, 6:396) — a way of making us more responsive to the reasons we have, rather than less. Such nudges, whether doxastic or behavioral, could be called **enabling nudges**.¹⁰ We present them as a category of special interest for the ethics of nudging. To be sure, there is no guarantee that an enabling nudge will get a moral green light: to intervene on another's behalf, even to facilitate their rational attitudes, may be in some cases objectionably patronizing; or it may remove opportunities for them to develop important skills and virtues (Meehan 2022). Nor are we claiming that *only* enabling nudges will be morally permissible; again, we expect the ethics to be more particularistic than that. We do claim, though, that enabling nudges have a leg up on non-enabling nudges. Like other nudges, they treat their targets as less than fully rational; but, instead of exploiting their irrationalities (paternalistically or otherwise), they work to ameliorate them.

6. Conclusion: The Place of Doxastic Nudges

We have spent most of this paper considering doxastic nudges, specifically, arguing that our Normative View offers better prospects than its rivals for bringing the nudging literature to bear in epistemology. As we bring our discussion to a close, though, it may help to comment on the relationship between doxastic nudges and the more general category of behavioral nudges that has chiefly occupied theorists. Where does the doxastic nudge fit into the broader picture?

Different theories of nudges will suggest different taxonomies. If Levy is right to say that nudges in general work by giving testimony, for instance, it will be natural to suppose in effect

⁹ We therefore agree with Levy (2019, 298) that, while "badly designed environments subvert autonomy by feeding bad reasons to us," good nudges can promote our rational agency. But this optimism does not require us to accept the further claim that nudges work *by* providing reasons.

¹⁰ See Engelen and Nys (2020, 147 and ff.), who argue that behavioral nudges can sometimes promote a nudgee's ability to achieve their own settled ends when such success is hindered by psychological or environmental factors.

that *all* nudges are doxastic nudges, operating first and foremost to change the nudgees' doxastic states. Any behavioral effects are to be found downstream of the nudged beliefs that rationalize them: "We nudge behavior by changing people's mental states, and that almost always entails changing their attitude toward some proposition" (Levy 2021, 43).

We offer a different categorization, however. Just as we think it is possible to affect a person's beliefs in ways disproportionate to her (epistemic) reasons for them, so too we think it is possible to affect a person's actions in ways disproportionate to her (practical) reasons for them. We can then unify the phenomenon of nudging under a general Normative View (see [REDACTED]), while at the same time recognizing the different ways in which these interventions can be made. Many cases of nudges, like those involving framing effects or the order in which evidence is presented, seem to route through the nudgee's doxastic faculties, even if they are meant to terminate finally in altered behavior. Other examples, though, seem classifiable as "direct" behavioral nudges:¹¹ an opt-out organ donation policy, say, like other kinds of "default" nudges, plausibly influences the nudger's behavior without going to the trouble of inculcating any mediating doxastic states in her.

We cannot expand further on this division of categories here. By way of farewell, however, we suggest that the difference between doxastic and (direct) behavioral nudging will come along with a general difference in normative profile. Purely behavioral nudges, affecting actions directly, may make relatively slight contact with the rational powers of their targets, who are at risk of being maneuvered puppet-like by others' will—a threat akin to that of coercion. When doxastic nudges influence behavior indirectly, on the other hand, the nudgee may retain full practical self-governance, but under informational conditions that have been chosen for her—a threat to autonomy more like that of deception.

References

- Arpaly, N., and Schroeder, T. (2012). "Deliberation and Acting for Reasons." *The Philosophical Review* 121(2), 209–39.

¹¹ [REDACTED]

- Bovens, L. (2009). 'The Ethics of Nudge'. In *Preference Change: Approaches from Philosophy, Economics and Psychology*, edited by Till Grüne-Yanoff and Sven Ove Hansson, 207–19. Theory and Decision Library. Dordrecht: Springer Netherlands.
- Congiu, L., and Moscati, I. (2021) Message and Environment: A framework for nudges and choice architecture. *Behavioural Public Policy*, 4(1), 71–87
- Ecker, U.K.H., Lewandowsky, S., Cook, J. et al. (2022). The psychological drivers of misinformation belief and its resistance to correction. *Nat Rev Psychol* 1, 13–29.
- Engelen, B. and Nys, Th. (2020). Nudging and Autonomy: Analyzing and Alleviating the Worries. *Review of Philosophy and Psychology*, 11, 137–56.
- Funke, D., and Flamini, D. (2018). A Guide to Anti-misinformation Actions around the World. *Poynter*. www.poynter.org/ifcn/anti-misinformation-actions/
- Gigerenzer, G. (2015). "On the Supposed Evidence for Libertarian Paternalism." *Review of Philosophy and Psychology* 6: 361–83.
- Grundmann, T. (2021). "The Possibility of Epistemic Nudging: Reply to My Critics." *Social Epistemology Review and Reply Collective* 10 (12): 28-35.
- Grundmann, T. (2023). 'The Possibility of Epistemic Nudging'. *Social Epistemology* 37 (2): 208–18. <https://doi.org/10.1080/02691728.2021.1945160>.
- Grüne-Yanoff, T. (2012) Old wine in new casks: Libertarian paternalism still violates liberal principles. *Social Choice and Welfare*, 38(4), 635–645.
- Hansen, P. G. (2016) The definition of nudge and libertarian paternalism: Does the hand fit the glove? *European Journal of Risk Regulation*, 7(1), 155–174
- Hausman, D. M., and Welch, B. (2010). Debate: To nudge or not to nudge*. *Journal of Political Philosophy*, 18(1), 123–136.
- Kahneman, D. (2011) *Thinking, fast and slow*. Farrar, Straus and Giroux.
- Kant, I. (1996). *The Metaphysics of Morals*. In *Kant's Practical Philosophy*. Tr. Gregor, M. Cambridge University Press.
- Levy, N. (2019). 'Nudge, Nudge, Wink, Wink: Nudging Is Giving Reasons'. *Ergo, an Open Access Journal of Philosophy* 6. <https://doi.org/10.3998/ergo.12405314.0006.010>
- Levy, N. (2021). "Nudging is giving Testimony: A Response to Grundmann." *Social Epistemology Review and Reply Collective* 10 (8): 43-47.
- Levy, N. (2022). *Bad Beliefs*. Oxford University Press.
- McKenna, R. (2020). Persuasion and Epistemic Paternalism. *Epistemic Paternalism: Conceptions, Justifications and Implications*, 91-106.
- MacKay, D., and Robinson, A. (2016). The ethics of organ donor registration policies: Nudges and respect for autonomy. *The American Journal of Bioethics*, 16(11), 3–12.
- Matheson, J, and Joly Chock, V. (2021). 'The Possibility of Epistemic Nudging: Reply to Grundmann'. *Social Epistemology Review and Reply Collective* 10 (8): 36–42.
- Meehan, D. (2020). 'Epistemic Vice and Epistemic Nudging: A Solution?' *Epistemic Paternalism: Conceptions, Justifications and Implications*, 247–59.
- Miyazono, K. (2023). 'Epistemic Libertarian Paternalism'. *Erkenntnis*, <https://doi.org/10.1007/s10670-023-00664-9>.
- Mongin, P, and Cozic, M. (2018). 'Rethinking Nudge: Not One but Three Concepts.' *Behavioural Public Policy* 2: 107-24.

- Parmer, W. J. (2023). Nudges, Nudging, and Self-Guidance Under the Influence. *Ergo an Open Access Journal of Philosophy*, 9(0). <https://doi.org/10.3998/ergo.2913>
- Plato (1967). *Plato in Twelve Volumes: With an English Translation*. Tr. Lamb, W. R. M. Harvard University Press.
- Pritchard, D. (2005). *Epistemic Luck*. Oxford University Press.
- Riley, E. (2017). The Beneficent Nudge Program and Epistemic Injustice. *Ethical Theory and Moral Practice*, 20(3), 597–616. <https://doi.org/10.1007/s10677-017-9805-2>
- Saghai, Y. (2013). ‘Salvaging the Concept of Nudge’. *Journal of Medical Ethics* 39 (8): 487–93.
- Schmidt, A. T., and Engelen, B. (2020). The ethics of nudging: An overview. *Philosophy Compass*, 15(4). <https://doi.org/10.1111/phc3.12658>
- Schroeder, M. (2007). *Slaves of the Passions*. Oxford University Press.
- Sunstein, C. R. (2014). *Why Nudge?* Yale University Press
- Thaler, R.H. and Sunstein C.R. (2021). *Nudge. The Final Edition*. Yale University Press.
- Walter, N., Cohen, J., Holbert, R. L., and Morag, Y. (2020). Fact-Checking: A Meta-Analysis of What Works and for Whom. *Political Communication*, 37(3), 350–375.
- Wilkinson, T. M. (2013). Nudging and Manipulation. *Political Studies*, 61(2), 341–355.
- White, M. D. (2013). *The manipulation of choice: Ethics and libertarian paternalism*. New York: Palgrave Macmillan.