



Epistemic Libertarian Paternalism

Kengo Miyazono¹

Received: 17 February 2022 / Accepted: 5 February 2023
© The Author(s) 2023

Abstract

Libertarian paternalism is a weak form of paternalism that recommends nudges rather than bans, restrictions, or other strong interventions. Nudges influence people's choice by modifying contextual factors (the “choice architecture”). This paper explores the possibility of an epistemic analogue of libertarian paternalism. What I call “epistemic libertarian paternalism” is a weak form of epistemic paternalism that recommends “epistemic nudges” rather than stronger paternalistic interventions. Epistemic nudges influence people's beliefs and judgments by modifying contextual factors (the “epistemic choice architecture”). The main aim of this paper is to defend epistemic libertarian paternalism from the “irrationality problem”, which I take to be the most urgent problem for epistemic libertarian paternalism; given how epistemic nudges work (i.e. they typically co-opt psychological biases), nudged beliefs are irrational. In response to the irrationality problem, I admit that nudged beliefs are often (not always, though) irrational, but insist that there are conditions in which epistemic nudging can be justifiable nonetheless. I will propose two conditions that are jointly sufficient for justifiable epistemic nudging: “Veridicality Condition” (which says that nudged beliefs are more likely to be true than non-nudged beliefs) and “Not-More-Irrationality Condition” (which says that nudged beliefs are not more likely to be irrational than non-nudged beliefs).

1 Introduction

Libertarian paternalism (LP) is a weak form of paternalism that recommends nudges rather than bans, restrictions, or other strong interventions (Jolls et al., 1998; Sunstein & Thaler, 2003; Sunstein, 2020; Thaler & Sunstein 2003, 2008, 2021). Nudges influ-

✉ Kengo Miyazono
miyazono@let.hokudai.ac.jp

¹ Graduate School of Humanities and Human Sciences, Hokkaido University, Nishi 7 Chome, Kita 10 Jo, Kita-ku, Sapporo 060-0810, Hokkaido, Japan

ence people's choices by modifying contextual factors (the "choice architecture"). This paper explores the possibility of an epistemic analogue of LP¹. What I call "epistemic libertarian paternalism" (ELP) is a weak form of epistemic paternalism that recommends "epistemic nudges" rather than stronger paternalistic interventions. Epistemic nudges influence people's beliefs and judgments by modifying contextual factors (the "epistemic choice architecture").

LP is more freedom-preserving than stronger forms of paternalism, which is an important advantage of LP. Analogously, ELP is more freedom-preserving than stronger forms of epistemic paternalism, which is an important advantage of ELP (see Sect. 2.2). LP, however, faces a serious problem (which I call the "irrationality problem" for LP); nudged decisions are irrational because they are influenced by psychological biases that are co-opted in nudges (see Schmidt & Engelen 2020 for an overview). ELP faces an analogous problem (the "irrationality problem" for ELP); nudged beliefs are irrational because they are influenced by psychological biases that are co-opted in epistemic nudges.

I take the irrationality problem to be the most urgent problem for ELP. In response, I admit that nudged beliefs are often (although not always) irrational, but insist that there are situations in which epistemic nudging can be justifiable nonetheless. I will propose two conditions that are jointly sufficient for justifiable epistemic nudging: the "Veridicality Condition" (which says that nudged beliefs are more likely to be true than non-nudged beliefs) and the "Not-More-Irrationality Condition" (which says that nudged beliefs are not more likely to be irrational than non-nudged beliefs).

Before proceeding, let us consider an example of epistemic nudging. People's judgments are predictably influenced by the frame in which information is presented. For example, in a classic experiment (Tversky & Kahneman, 1981) where participants were asked to choose a policy to prepare for the outbreak of a hypothetical unusual disease, participants regarded an option as more attractive when the risk of the option was expressed in a positive frame in terms of how many people would be saved ("there is 1/3 probability that 600 people will be saved and a 2/3 probability that no people will be saved") than when the same risk was put in a negative frame in terms of how many people would die ("there is a 1/3 probability that nobody will die and a 2/3 probability that 600 people will die"). Psychological effects of this kind can be co-opted for the purpose of epistemic nudging, for example in the context of responding to groundless vaccination skepticism (e.g. McKenna 2020). In the context of public announcement of a vaccine, one might adopt a positive frame (e.g. in terms of how many people will be saved by it, how many medical professionals regard it as safe, etc.), which might help people to form accurate beliefs about the safety and health benefits of the vaccine, and avoid a negative frame (e.g. in terms of how many people will die due to its side effects, how many medical professionals regard it as unsafe, etc.), which might fuel groundless vaccination skepticism. I call this intervention "VACCINATION" and treat it as a possible example of epistemic nudging in the following discussions.

¹ Similar ideas have been discussed by Levy (2017, 2021), Grundmann (2021), McKenna (2020), and Meehan (2020).

There is a large body of empirical research of the framing effect in general (e.g. Kühberger 1998) as well as in health-related contexts (e.g. Rothman & Salovey 1997; Gallagher & Updegraff 2012) that motivates a serious philosophical discussion of VACCINATION and similar epistemic nudges. Recently, Altay and Mercier (2020) conducted a series of experiments on the influence of the positive and negative framing of vaccinations, in terms of side effects and medical consensus.

Positive Frame: 999 people out of 1,000 don't have any severe side effects after being injected.

Negative Frame: (Only) 1 individual out of 1,000 has some severe side effects after being injected with a vaccine.

Positive Frame: 90% of medical scientists think that vaccines are safe.

Negative Frame: (Only) 10% of medical scientists don't think that vaccines are safe.

Participants (pro-vaccination individuals in the United States and United Kingdom) were exposed to positively and negatively framed descriptions of vaccines such as the ones above, and responded to questions concerning how plausible the information is, how well they can remember the information, how willing they are to transmit the information to others, etc. Overall, the results suggest that the framing effect does influence participants' attitude toward the safety of vaccination; those exposed to positively framed descriptions exhibited more positive attitude toward vaccination than those exposed to negatively framed descriptions.

Still, more empirical research is needed to examine whether epistemic nudges like VACCINATION really work not only for pro-vaccination people but also for anti-vaccination people, whether they really work not only in experimental settings but also in real-life settings, whether they have not only momentary and short-lived effects but also stable and long-term effects (e.g. Meehan 2020), whether, even if they are effective, they really influence beliefs (e.g. on the safety of vaccines), or they directly influence behaviors (e.g. of getting vaccinated) without changing beliefs², etc. These are empirical issues that I do not address in this paper (see Mertens et al. 2022 for a recent meta-analysis of the effect of nudge-style interventions in general). It suffices to say for the purpose of this paper that epistemic nudges like VACCINATION deserve a serious philosophical discussion given the current empirical research and evidence.

2 Clarifying Epistemic Libertarian Paternalism

2.1 Libertarian paternalism

Let us start by clarifying what LP is. LP is a weak form of paternalism that recommends nudges rather than stronger interventions (Jolls et al., 1998; Sunstein &

² In the latter case, VACCINATION turns out to be a non-epistemic nudge rather than an epistemic nudge. In thank an anonymous referee for raising this possibility.

Thaler, 2003; Sunstein, 2020; Thaler & Sunstein 2003, 2008, 2021). LP is said to be “libertarian” in the sense that “choices are not blocked or fenced off” (Sunstein & Thaler, 2003, 1162) by nudgers, and it is said to be “paternalistic” in the sense that nudgers “are self-consciously attempting to move people in welfare-promoting directions” (Sunstein & Thaler, 2003, 1162).

A nudge is characterized as an intervention in the “choice architecture”: “A nudge [...] is any aspect of the choice architecture that alters people’s behavior in a predictable way without forbidding any options or significantly changing their economic incentive” (Thaler & Sunstein, 2008, 6)³. The choice architecture is characterized as the “context in which people make decisions” (Thaler & Sunstein, 2008, 3). Here are some classic examples from Thaler and Sunstein (2008).

CAFETERIA An example of nudging is to arrange the food display in a school cafeteria so that healthy items (such as salads or fruits) are displayed at eye level and unhealthy items (such as fatty chocolate brownies) are displayed at a higher or lower level. This intervention is expected to increase students’ consumption of healthy food items and decrease the consumption of unhealthy ones because of the psychological tendency to pick items at eye level more often than items at different levels.

RETIREMENT Another example of nudging is to introduce default retirement savings programs where people in an organization are automatically enrolled on a retirement savings program by default, but there is an easy opt-out option. This intervention is expected to increase the number of subscribers to retirement savings programs because of the psychological tendency to stay with the default option (the status quo bias).

Both CAFETERIA and RETIREMENT are “libertarian” in the sense that “choices are not blocked or fenced off”. CAFETERIA, for example, does not block the option of eating chocolate brownies. Students can still have them if they really want to. RETIREMENT does not block the option of not enrolling on the retirement savings program. People can easily opt-out if they really want to. Both CAFETERIA and RETIREMENT are “paternalistic” in the sense that nudgers “are self-consciously attempting to move people in welfare-promoting directions”⁴. CAFETERIA aims to promote the welfare of students by nudging them to consume more healthy food items and avoid unhealthy ones. RETIREMENT aims to promote the welfare of people in an organization by nudging them to enroll on a retirement savings program⁵.

³ There is, however, an ambiguity here; a nudge can be understood either as “any aspect of choice architecture” (Thaler & Sunstein, 2008, 6) or as an intentional intervention into choice architecture. I adopt the latter conception in this paper. The difference between the two conceptions matters in the context of evaluating the so-called inevitability argument for nudging (Grill, 2014; Hausman & Welch, 2010; Vallier, 2016).

⁴ One might think that nudges such as CAFETERIA and RETIREMENT are “paternalistic” only in a weak sense because they do not involve interfering with the liberty or autonomy of agents (Dworkin, 2020). In response, one might insist that nudges do interfere with autonomy (see Engelen & Nys (2020) for an overview). Alternatively, one might deny the interference with liberty and autonomy as a necessary condition for paternalistic acts (e.g. Ryan 2016, 2018). I do not discuss these conceptual issues in this paper; I will be neutral on whether nudges are “paternalistic” in a genuine sense.

⁵ Some examples of nudging by Thaler and Sunstein do not aim to improve the welfare of those nudged; instead, they aim to improve the welfare of other individuals in society; e.g. a new organ donation system where people agree to be donors by default (“DONATION”), which aims to improve the welfare of other

Nudges (e.g. placing chocolate brownies at a high level in the food display) are distinguished from incentives (e.g. taxing chocolate brownies). Nudges play a central role in LP, while incentives play a relatively minor role. Incentives modify the *content* of choice, while nudges modify the *context* of choice. More precisely, incentives modify the “choice set”, which is constituted by options (e.g. salads, chocolate brownies, etc.) together with their features that are relevant to the choice (e.g. taste, price, ingredients, calories, etc.). Nudges, in contrast, modify the “choice architecture”, which is constituted by irrelevant contextual factors (e.g. the physical position in the food display)⁶.

Taxing chocolate brownies (“TAX”) is an incentive because it modifies the choice set by changing the price of chocolate brownies. Placing chocolate brownies at a high level in the food display (CAFETERIA) is a nudge because it does not modify the choice set; instead, it modifies the choice architecture by changing the physical position of chocolate brownies.

There are, however, two problems with this distinction between nudges and incentives. First, nudges can modify the choice set. As Thaler and Sunstein (2008) admit, for example, CAFETERIA slightly modifies the choice set by increasing the cognitive cost of choosing chocolate brownies. Thaler and Sunstein agree that CAFETERIA modifies the choice set but insist that the modification is negligible. CAFETERIA only generates a low and negligible cost of choosing chocolate brownies. This response seems to imply that the incentive/nudge distinction is a matter of degree: “it is more precise to define a nudge as an initiative that affects people’s behavior without imposing significant material burdens or offering significant material benefits. As an economic incentive approaches zero, it starts to look more like a nudge” (Sunstein, 2020, 6).

Second, incentives can modify the choice architecture. As Grill points out, “removing one option may cause some other options to become more salient, or it may change chooser perspective on the entire choice set” (Grill, 2014, 140). It is not obvious that we can simply solve this problem by saying that removing one option in the choice set only causes a negligible impact on the salience of other options. Perhaps a better response to these problems is to distinguish the primary aim of intervention and its side-effects. For instance, the primary aim of nudges, such as CAFETERIA, is to modify the choice architecture, and they can have some side-effects on the choice set (e.g. increasing the cognitive cost of choosing some options). In contrast, the primary aim of incentives, such as TAX, is to modify the choice set, and they can have some side-effects on the choice architecture (e.g. increasing the salience of some options).

individuals in society (e.g. recipients of donated organs) rather than of nudged individuals themselves (e.g. organ donors). This seems to show that there are two kinds of nudges (Hagman et al., 2015): pro-self nudges (e.g. RETIREMENT) and pro-social nudges (e.g. DONATION). I will be neutral on whether pro-social nudges are “paternalistic” in a genuine sense.

⁶ This characterization of the choice architecture is not very informative unless the meaning of the term “contextual factor” is clarified. A possible proposal is that “contextual factors” are the ones that causally but not rationally relevant to the choice (which is consistent with Thaler and Sunstein’s (2008) claim that perfectly rational economic agents are not responsive to nudges). However, this proposal is not theory-neutral; it terminologically rules out the possibility that some nudges are in fact rationally relevant (Gigerenzer, 2015; Levy, 2019a, 2021). See Sect. 3.4 for relevant discussions.

2.2 Epistemic Libertarian Paternalism

Let us now consider ELP, which is an epistemic analogue of LP. The latter is contrasted with stronger forms of paternalism, while the former is contrasted with stronger forms of epistemic paternalism.

A strong form of epistemic paternalism is what I call “epistemic access paternalism” (EAP) (Ahlstrom-Vij, 2013, 2018; Goldman, 1991; Pritchard 2013). EAP interferes with freedom of inquiry (Ahlstrom-Vij, 2013, 2018). For instance, EAP restricts jurors’ access to some information about defendants (such as the information about their past crimes, their personalities and character traits, etc.) to prevent the jurors from being biased by such information. In the context of vaccinations, EAP might recommend restricting people’s access to some information about the safety of vaccinations (such as the information of rare tragic side-effects) to prevent them from being biased by such information. Let us call this intervention “VACCINATION2”.

ELP is a weak form of epistemic paternalism that recommends epistemic nudges. In the context of vaccination, ELP recommends VACCINATION where the information about the safety of vaccination is put in a particular frame. ELP is “epistemically libertarian” in the sense that no inquiries are banned or blocked; e.g. VACCINATION does not restrict people’s access to the information of rare tragic side effects⁷. ELP is “epistemically paternalistic” in the sense that epistemic nudges are self-consciously attempting to move people in epistemic-goods-promoting directions⁸. ELP can take different forms depending on different interpretation of “epistemic goods”. An interpretation, called “veritism”, is that truth is the ultimate epistemic good, and other epistemic standings, such as understanding, are merely instrumentally valuable. Another interpretation, which can be called “non-veritism”, is that other epistemic standings, such as understanding, can also be ultimate epistemic goods (Pritchard, 2013).

An epistemic nudge is an intervention in what I call the “epistemic choice architecture”, which is the context in which people make judgments and form beliefs. VACCINATION, for example, is an intervention in the framing of relevant informa-

⁷ There could be some interventions that interfere with freedom of inquiry in a subtle way; e.g. changing the search engine algorithms such that some information is very hard to find (“ALGORITHM”). This is analogous to a version of CAFETERIA where unhealthy food items are placed in a more obscure place so that they cannot easily be found. This version of CAFETERIA, according to Thaler and Sunstein, is not a nudge; it generates some non-negligible cognitive cost of finding and choosing the unhealthy food items. It does not satisfy the condition that “nudges count as such, and qualify as libertarian paternalism, only if any costs are low” (Thaler & Sunstein, 2008, 8). Something similar can be said about ALGORITHM; it is not an (epistemic) nudge because it generates some non-negligible cognitive cost of finding and evaluating the relevant information.

⁸ As we have seen above (the footnote 4), there is a conceptual issue as to whether LP is a genuine example of paternalism. Interfering with liberty or autonomy, which is part of Dworkin’s (2020) definition of paternalism, is lacking in LP. There is an analogous issue as to whether ELP is a genuine example of epistemic paternalism. Interfering with freedom, which is part of Ahlstrom-Vij’s (2013) definition of epistemic paternalism, is lacking in ELP. VACCINATION2 interferes with freedom of inquiry, but VACCINATION does not. I do not discuss these conceptual issues here; I will be neutral on whether ELP is “epistemically paternalistic” in a genuine sense.

tion, which is part of the context in which people make judgments about the safety of vaccination⁹.

Here are some clarificatory remarks on ELP and its theoretical details.

(1) Epistemic LP and Non-Epistemic LP: One might think that there is nothing particularly epistemic about VACCINATION; it is just a standard example of (non-epistemic) LP that aims at promoting welfare. VACCINATION does influence what people believe, and it does so in order to get more people vaccinated and thereby promote their welfare. This is an instance of a general worry that epistemic paternalism collapses into non-epistemic paternalism (e.g. Bullock 2018) in the sense that the former is reduced to the latter and thus there is nothing particularly epistemic about the former. This paper does not address this issue, which is about epistemic paternalism in general rather than about ELP in particular. A possible option would be to introduce a fine-grained distinction between different versions of epistemic paternalism, including the one that collapses into non-epistemic paternalism and the one that does not (e.g. Pritchard 2013).

(2) Epistemic Nudges and Epistemic Incentives: There is a distinction in LP between nudges and incentives. And there is a corresponding distinction in ELP between epistemic nudges and epistemic incentives¹⁰. Epistemic incentives modify what I call the “epistemic choice set”, while epistemic nudges modify the epistemic choice architecture. The epistemic choice set is constituted by opinions (or the propositions that express the opinions) together with their relevant evidence that is available to epistemic agents. In contrast, the epistemic choice architecture is constituted by irrelevant contextual factors.

Denying access to information about vaccinations (VACCINATION2) is an epistemic incentive because it modifies the epistemic choice set by changing the evidence people have about the safety of vaccinations. Framing information in a particular way (VACCINATION) is an epistemic nudge because it does not modify the epistemic choice set; instead, it modifies the epistemic choice architecture in which people make judgments (but see Sect. 3.4).

Just like in the context of (non-epistemic) nudges/incentives, we can distinguish the primary aim of epistemic nudges/incentives and their side-effects. The primary aim of epistemic nudges, such as VACCINATION, is to modify the epistemic choice architecture, and they can have some side-effects on the epistemic choice set. In contrast, the primary aim of epistemic incentives, such as VACCINATION2, is to modify the epistemic choice set, and they can have some side-effect on the choice architecture.

⁹ For this reason, providing an argument (e.g. for the effectiveness of vaccination) is not an epistemic nudge. It is an intervention not into the context in which people make judgment (the “epistemic choice architecture”) but rather into the information that is directly relevant to judgment (the “epistemic choice set”). See below for the distinction between the epistemic choice architecture and the epistemic choice set.

¹⁰ The term “incentive” might be misleading, especially when it gives the false impression that epistemic incentives are motivational. I use the term “incentive” for epistemic incentives only to highlight the analogy between nudges (interventions into the choice architecture) / incentives (interventions into the choice set) in LP and epistemic nudges (interventions into the epistemic choice architecture) / epistemic incentives (interventions into the epistemic choice set) in ELP.

VACCINATION2 is an epistemic incentive (rather than an epistemic nudge); it primarily aims to modify the epistemic choice set (rather than the epistemic choice architecture)¹¹. Here are some other examples of epistemic incentives in the recent literature. Rini (2017) discusses the idea of putting a “disputed” tag to problematic Facebook posts as a solution to the problem of fake news. This intervention, which I call “TAG”, is an epistemic incentive. By putting the “disputed” tag to a problematic post, TAG primarily aims to change the evidence people have about the reliability of the post. Levy (2019b) discusses the idea of not giving a platform to those with offensive views. This intervention, which I call “NO-PLATFORMING”, is another epistemic incentive. NO-PLATFORMING primarily aims to change the evidence people have about the plausibility of the offensive views in question; “provision of a platform itself provides evidence that rational agents ought to take into consideration in forming their beliefs” Levy 2019b, 487).

In summary, ELP is a weak form of epistemic paternalism that recommends epistemic nudges (e.g. VACCINATION) rather than stronger epistemic paternalistic interventions (e.g. VACCINATION2). ELP is “epistemically libertarian” in the sense that no inquiries are banned or blocked. ELP is “epistemically paternalistic” in the sense that epistemic nudgers are self-consciously attempting to move people in an epistemic-goods-promoting directions.

3 Defending Epistemic Libertarian Paternalism

3.1 The Irrationality Problem

What are the advantages and disadvantages of ELP compared to other forms of epistemic paternalism such as EAP? What are the advantages and disadvantages of VACCINATION (which is recommended by ELP) compared to VACCINATION2 (which is recommended by EAP)?

ELP has an important advantage, which is inherited from LP in general, that it is freedom-preserving. CAFETERIA, which is recommended by LP, preserves freedom of choice (e.g. does not block the option of choosing chocolate brownies). Analogously, VACCINATION, which is recommended by ELP, preserves freedom of inquiry (e.g. does not deny people’s access to the information of rare tragic side-effects). As McKenna points out, epistemic nudges like VACCINATION do not “involve outright deception or the straightforward withholding of information (as when a judge withholds information about a defendant’s criminal record from the jury)”; rather they only “involve presenting information in ways that are designed to make it more likely that the audience will react in a desired way” (McKenna, 2020, 97).

Unlike EAP, ELP does not interfere with freedom of inquiry. This is not to say, however, that EAP is never justified because of its interference with freedom of

¹¹ EAP recommends VACCINATION2, which is an epistemic incentive. But note that EAP recommendations and epistemic incentives are not co-extensional. For instance, TAG is an epistemic incentive, but it is not what EAP recommends; it does not interfere with freedom of inquiry.

inquiry. I do not rule out the idea that there are some conditions that are sufficient for justifying interfering with freedom of inquiry (e.g. Ahlstrom-Vij 2013, 2018). My claim is only that, first, freedom of inquiry has at least some value (even if it could be overridden by some other considerations in some cases) and, second, other things being equal, an intervention that does not interfere with freedom of inquiry has an advantage over an intervention that does interfere with freedom of inquiry.

However, there seems to be a serious disadvantage of ELP: nudged beliefs are epistemically defective. ELP seems to have the epistemically problematic consequence that nudged beliefs are inherently irrational. Suppose that Ken, nudged by VACCINATION, forms a belief about the safety of vaccinations. Ken's belief is irrational; it is influenced by how the information is framed, which is an irrelevant contextual factor. Other forms of epistemic paternalism do not seem to face this problem. Suppose that Naomi, influenced by VACCINATION2, forms a belief about the safety of vaccinations. Naomi's belief can be rational; it is influenced by the amount of available information about the safety of vaccinations, which is a relevant factor. Naomi does not have access to information about every possible side-effect, but she can still form a rational belief on the basis of other evidence. Alternatively, she can rationally suspend her judgment if she thinks that she does not have sufficient evidence.

Several clarifications are in order.

(1) The Irrationality Problem and Biases: The irrationality problem arises for those nudges that co-opt psychological biases, such as VACCINATION (which co-opts the framing effect) or RETIREMENT (which co-opts the status quo bias). Although co-opting psychological biases is a salient feature of many nudges (Bovens, 2009), it is not an essential or definitional feature of nudges in general (Sunstein, 2018). For instance, some nudges, such as disclosing information about a complicated pricing structure of credit card fees ("DISCLOSURE"), do not exploit any psychological biases¹². Following Barton and Grüne-Yanoff (2015), one can distinguish three types of nudges: (i) nudges that co-opt some biases ("heuristics-triggering nudges"; e.g. VACCINATION and RETIREMENT); (ii) nudges that block some harmful biases ("heuristics-blocking nudges"); and (iii) nudges that provide some information ("informing nudges"; e.g. DISCLOSURE). Our focus in the following discussion is on the nudges of the first type.

(2) The Irrationality Problem for LP and for ELP: The irrationality problem has already been discussed in the literature, but mostly in the context of (non-epistemic) LP (see Schmidt & Engelen 2020 for an overview). Suppose that Ken, nudged by RETIREMENT, signs up for a retirement plan. Ken's choice is influenced by the default option, which is an irrelevant contextual factor. Making a choice under the influence of such an irrelevant contextual factor is irrational. It would be fair to say that the irrationality problem for ELP is more serious than the irrationality prob-

¹² One might think that DISCLOSURE, as an epistemic intervention, counts as an (epistemic) incentive rather than a (epistemic) nudge because it changes the epistemic choice set by providing people with additional information or evidence. However, it is not clear that the additional information provided constitutes additional evidence. Perhaps the additional information in DISCLOSURE is a consumer-friendly summary of complicated terms and conditions that were already available to consumers. In that case, it is not clear that the additional information gives any new evidence; perhaps it is only a summary of pre-existing evidence. The epistemic choice set is not modified.

lem for (non-epistemic) LP. For example, one might be able to bite the bullet in the context of (non-epistemic) LP; one can admit that nudged choices are irrational but insist that this cost is easily outweighed by their practical benefits. For instance, in RETIREMENT, Ken's choice is certainly influenced by an irrational status quo bias, but overall RETIREMENT is a good thing to do because it significantly improves Ken's well-being. In contrast, the biting-the-bullet response is less attractive in the context of ELP. In light of the very purpose of *epistemic* paternalism, we cannot simply accept the consequence that nudged beliefs are irrational.

(3) Different Versions of the Irrationality Problem: Different authors discuss different versions of the irrationality problem (for LP). The first version of the problem is that “nudges ‘exploit’ or ‘take advantage of’ behavioral biases” (Sunstein, 2018, 64), where the emphasis is on the fact that nudges co-opt irrational biases. The second version is that “nudges are based on a belief that human beings are ‘irrational’, which is both insulting and false” (Sunstein, 2018, 64), where the emphasis is on the pessimistic picture of human nature that is assumed in LP. This version seems to be what Gigerenzer has in mind when he writes: “I do not argue against nudging per se. But I do object to the justification of such techniques on the basis of people's lack of rationality by libertarian paternalists such as Thaler and Sunstein” (Gigerenzer, 2015, 363). The third version is that “there is something objectionable about the government treating people as irrational” (Schmidt, 2019, 512), where the emphasis is on nudgers treating people as irrational or failing to respect their rational agency. The fourth version is that “when something bypasses our reasoning, as nudges are supposed to, our autonomy as agents is threatened” (Levy 2019a, 282), where the emphasis is on nudges threatening autonomy by bypassing rational capacities. The fifth version is that “exploiting decision-making foibles will ultimately diminish people's autonomous decision-making capacities” (Hausman & Welch, 2010, 135), where the emphasis is on the long-term negative consequences of nudges on rational decision-making capacities. The first version (about co-opting irrational biases) is directly relevant to our discussion, while others are only indirectly relevant.

(4) Rationality: The irrationality problem is open to different interpretation of rationality/irrationality¹³. I adopt the so-called standard picture of rationality (Stein, 1996) which defines rationality in terms of rules of logic and mathematics. Nudges often co-opt psychological biases that are regarded as irrational according to the standard picture. The framing effect, which is co-opted in VACCINATION, is irrational according to the standard picture; for example, the probability of a harmful side-effect is the same when the information of the side-effect is put in a positive frame and when it is put in a negative frame. Again, the status quo bias, which is co-opted in RETIREMENT, is irrational according to the standard picture; the expected utility of a retirement plan is the same when signing up to the plan is the default option and when it is not¹⁴.

It is far from obvious, however, that the irrationality problem gets off the ground when we accept a different conception of rationality, such as the ecological concep-

¹³ See Engelen (2019) for a discussion of nudging and different types of rationality.

¹⁴ See Samuels and Stich (2004) and Samuels, Stich, and Bishop (2002) for psychological biases and rationality.

tion of rationality. The ecological conception of rationality characterizes rationality in terms of cognitive success in the relevant environment or, in other words, the “fit between structures of information-processing mechanisms in the mind and structures of information in the world” (Todd & Gigerenzer, 2007, 170). Assuming the ecological conception of rationality, Gigerenzer (2015) has argued that the kind of psychological biases that are utilized in LP, such as the framing effect, are not irrational. These biases can be ecologically rational in the sense that they can be useful in some ecological contexts in which these biases are supposed to operate.

However, the ecological conception of rationality is still controversial. One of the objections to the ecological conception is that it conflates (biological) adaptiveness and (epistemic) rationality (e.g. Stanovich 2004); for instance, Ken’s nudged belief in the case of VACCINATION might be biologically adaptive in a particular ecological context, but it is not “rational” in the epistemically relevant sense of the term. I do not try to settle the debate between the standard picture and the ecological conception here (for relevant discussions, see Samuels & Stich 2004; Samuels, Stich, & Bishop 2002). For the purpose of this paper, I simply set aside the ecological conception (but I will discuss Gigerenzer’s interpretation of some particular nudges in Sect. 3.4) and focus on the standard picture (for nudging and ecological rationality, see Gigerenzer 2015; Grüne-Yanoff & Hertwig 2016; Schmidt 2019).

3.2 Solving the Irrationality Problem

A strong response to the irrationality problem is to deny the idea that nudged beliefs, such as Ken’s belief in the case of VACCINATION, are irrational. For instance, one might argue that Ken’s nudged belief is rational because it is rational to be responsive to the framing effect (Gigerenzer 2015; Levy 2019a, 2021; McKenzie 2004; Sher and McKenzie 2006, 2008; see also Bermúdez 2022). I will discuss the strong response to the irrationality problem in Sect. 3.4. My response to the irrationality problem is weaker; I do think that nudged beliefs are, or can be, irrational, but insist that there are some cases in which epistemic nudges can be justifiable nonetheless. In particular, I propose two conditions as jointly sufficient for justifiable epistemic nudging.

The jointly sufficient conditions for an epistemic nudge, N , targeting a nudgee, X , to be justifiable are:

- (1) **Veridicality Condition (VC)**: X is more likely to form a true belief when X is nudged by N than when X is not.
- (2) **Not-More-Irrationality Condition (NMIC)**: It is not the case that X is more likely to form an irrational belief when X is nudged by N than when X is not.

VC is met when, for example, Ken is more likely to form a true belief about the safety of vaccinations when he is nudged by VACCINATION than when he is not. Ken forms the belief that P when he is nudged by VACCINATION and he forms the belief that Q when he is not nudged; and the former belief is more likely to be true than the latter belief. Note that I am not saying that VACCINATION actually satisfies VC. The purpose of this section is to provide jointly sufficient conditions

for justifiable epistemic nudges rather than making any substantive claims about particular epistemic nudges such as VACCINATION. Whether VACCINATION satisfies VC depends on how VACCINATION actually works, which is an empirical issue. If VACCINATION is not effective at all, then Ken's nudged belief that P might just be identical with his non-nudged belief that Q . Alternatively, even if VACCINATION is effective ($P \neq Q$), it is conceivable that the nudged belief that P is not more likely to be true than the non-nudged belief that Q . These possibilities are not ruled out. However, these are empirical issues about the effects of VACCINATION, which I will not discuss in this paper.

NMIC is met when, for example, it is not the case that Ken is more likely to form an irrational belief about the safety of vaccinations when he is nudged by VACCINATION than when he is not. It is not the case that Ken's nudged belief that P is more likely to be irrational than his non-nudged belief that Q . His nudged belief that P is irrational because it is influenced by the framing effect. But his non-nudged belief that Q might also be influenced by the framing effect. After all, the non-nudged belief that Q is based on some information, which is inevitably put in some frame. In that case, the nudged belief that P and the non-nudged belief that Q are equally influenced by the framing effect. It is not the case that the former is more irrational than the latter (e.g. Engelen 2019). This idea is related to the inevitability (or unavoidability) argument for nudges: "a great deal of nudging is inevitable. So long as government has offices and websites, it will be nudging. If the law establishes contract, property, and tort law, it will be nudging, if only because it will set out default rules, which establish what happens if people do nothing. (Much of the law is a form of choice architecture). If the government provides information, it will be nudging" (Sunstein, 2018, 62).

Here is a clarification about the inevitability argument. A common objection to the inevitability argument is that it conflates the act of nudging, which is an intentional intervention, and the choice architecture, which is the context of choice; the latter is certainly inevitable but the former is not (Grill, 2014; Hausman & Welch, 2010; Vallier, 2016). But this response to the inevitability argument is largely irrelevant to our discussion. What is crucial for NMIC is that Ken is subject to the framing effect both when he is nudged by VACCINATION and when he is not. It does not matter whether the framing effect takes the form of an intentional intervention (in the former case) or an unintended part of the choice architecture (in the latter case).

Another clarification is that I am not committed to the view that all nudges are inevitable; e.g. Sunstein admits that "some nudging is optional" (Sunstein, 2018, 63). My claim is a conditional one: *if an epistemic nudge is inevitable*, then it may be justifiable. More precisely, *if an epistemic nudge is inevitable*, then it will likely to satisfy NMIC, which is one of the jointly sufficient conditions for justifiable epistemic nudging.

Is VACCINATION really inevitable? This depends on whether the framing effect is really inevitable. Some might think that there can be some neutral frames, which are neither positive nor negative¹⁵. I do not have a strong view on this issue. As I said

¹⁵ The conjunction of the positive frame and the negative frame might serve as a neutral frame. See, for instance, Druckman's (2001) experiment with conjunctive frames.

above, my purpose in this section is to provide jointly sufficient conditions for justifiable epistemic nudges rather than to make any substantive claims about particular epistemic nudges such as VACCINATION and particular psychological effects such as the framing effect. But I am inclined to think that the framing effect cannot easily be avoided¹⁶. Being influenced by a neutral frame, if there is such a thing, is not the same as being insensitive to the framing effect. The former does not imply the latter. For example, when Ken forms a belief in response to the neutrally framed information, he can be still described as being sensitive to the framing effect in a counterfactual sense; i.e. he would have judged differently if the information had been framed differently¹⁷.

VC in itself is too weak for justifying epistemic nudges. Suppose that an epistemic nudge satisfies VC but not NMIC; e.g. it modifies the contextual factors that robustly influence judgment but can easily be avoided¹⁸. It is not easy to justify this nudge as an epistemic paternalist intervention¹⁹. Certainly, this nudge does have an epistemic advantage; the nudged belief is more likely to be true than the non-nudged belief. But it also has an epistemic disadvantage; the nudged belief is more likely to be irrational than the non-nudged belief (because the relevant contextual factor is easily avoidable)²⁰.

Again, NMIC in itself is too weak for justifying epistemic nudges. Suppose that an epistemic nudge satisfies NMIC but not VC; e.g. it modifies the contextual factors that are unavoidable but has little impact on judgment. It is not easy to justify this nudge as an epistemic paternalist intervention. Certainly, this nudge does not have an epistemic disadvantage; it is not the case that the nudged belief is more likely to be irrational than non-nudged belief (because the relevant contextual factor is unavoid-

¹⁶ One might think that, if the framing effect is really inevitable, there is a sense in which the nudged belief is “innocent” or “blameless” despite its irrationality. This idea is closely related to (“No Alternative” condition of) what Bortolotti (2020) calls “epistemic innocence” of irrational beliefs.

¹⁷ Still, there can be a further debate as to whether the “sensitivity” in the counterfactual sense is relevant to the inevitability of the framing effect. One might think, for instance, that it is too strong to assume that in order for X to avoid the framing effect, X needs to be “insensitive” to the framing effect in the counterfactual sense. This issue seems to depend on what “avoiding the framing effect” actually means, which I cannot address here.

¹⁸ A possible example is the nudge that relies on the anchoring effect (Tversky & Kahneman, 1974); e.g. giving a random low number as an “anchor” in order to mitigate the tendency (of some people) to overestimate the number of vaccination-related accidents in the past. This anchoring nudge would satisfy VC (assuming that the anchoring effect is effective in the cases like this) but not NMIC if giving a random number as an anchor is avoidable (although it is conceivable that people are always surrounded and influenced by some anchors in everyday life).

¹⁹ Of course, it is possible to defend these nudges on non-epistemic grounds; e.g. because of these nudges, Ken correctly judges that vaccinations are safe and useful, and he actually gets vaccinated. But this is not to defend them as *epistemic paternalist interventions*. Non-epistemic defense of these nudges collapses into non-epistemic paternalism (Bullock, 2018).

²⁰ One might think, however, that VC is sufficient for justifying epistemic nudges at least for those who are committed to a strong veritist epistemic paternalism where rationality, understanding, knowledge, etc. are all instrumental. This is certainly true, but this paper seeks a more neutral justification of epistemic nudges that does not presuppose strong veritism. See Pritchard (2013) for a related discussion of veritist and non-veritist approaches to epistemic paternalism.

able). But it does not have an epistemic advantage either; it is not the case that the nudged belief is more likely to be true than the non-nudged belief.

In contrast, VC and NMIC jointly justify epistemic nudges. Suppose that an epistemic nudge satisfies VC and NMIC; e.g. it modifies the contextual factors that robustly influence judgment and cannot easily be avoided. This nudge can be justified as an epistemic paternalist intervention. It does have an epistemic advantage; the nudged belief is more likely to be true than the non-nudged belief. And it does not have an epistemic disadvantage; it is not the case that the nudged belief is more likely to be irrational than non-nudged belief (because the relevant contextual factor is unavoidable).

In the remainder of this paper, I will discuss two residual worries about my response to the irrationality problem. My overall strategy is to admit that epistemic nudges are irrational but at the same time insist that they can be justified in some cases. One worry about my strategy is that once I admit that epistemic nudges are irrational, they are less attractive than other paternalistic interventions that are not irrational (Sect. 3.3). Another worry is that I am conceding too much when I admit that epistemic nudges are irrational (Sect. 3.4).

3.3 Epistemic Nudges vs. Epistemic Incentives

The first worry about my proposal is that, even if some epistemic nudges satisfy VC and NMIC, it does not change the fact that epistemic nudges, which are irrational, are less attractive than other epistemic interventions, such as epistemic incentives, that are not irrational²¹. When Ken is influenced by VACCINATION, which is an epistemic nudge, he forms the belief that P , which is irrational due to the framing effect. When, in contrast, Naomi is influenced by VACCINATION2, which is an epistemic incentive, she forms the belief that R (possibly $P=R$), which is rational because it is responsive to the available evidence. In this case, it is hard to resist the conclusion that, other things being equal, VACCINATION2, which produces a rational belief, is more desirable from an epistemic paternalist point of view than VACCINATION, which produces an irrational belief. Even if VACCINATION satisfies VC and NMIC, it cannot be justified when there is an obviously better option (namely VACCINATION2).

One possible response to this worry is to deny the other-things-being-equal assumption. It is certainly true that if the only relevant difference between VACCINATION and VACCINATION2 is that the former produces an irrational belief and the latter produces a rational belief, then the latter is more desirable than the former from an epistemic paternalist point of view. But the antecedent of this conditional can be challenged. There can be some important differences between VACCINATION and VACCINATION2. As I have noted already, an obvious difference between

²¹ A traditional example of a rational intervention is education. As Gigerenzer points out, “the true alternative to nudging is education: making children and adults risk savvy” (Gigerenzer, 2015, 375–376). Other rational interventions include: (1) the “boost” approach (Grüne-Yanoff & Hertwig, 2016); (2) the “think” approach (John et al., 2009). Similar proposals include: (3) the “nudge plus” approach (Banerjee & John, 2021); (4) the “type 2 nudge” approach (Hansen & Jespersen, 2013); and (5) the “nudges to reason” approach (Levy, 2017).

them is that, unlike VACCINATION2, VACCINATION does not restrict freedom of inquiry. Even setting this difference aside, there can be some epistemic differences between VACCINATION and VACCINATION2. It is far from obvious that VACCINATION and VACCINATION2 are equally likely to produce true beliefs; e.g. denying people's access to information about the side-effects of vaccinations (VACCINATION2) might cause some people to accept unwarranted conspiracy theories that are clearly false. Note that my claim is not that VACCINATION2 should be avoided just because it can have unintended consequences such as inviting unwarranted conspiracy theories. After all, any interventions, including VACCINATION, can have unintended consequences. Rather my claim is only that it is an empirically open question whether VACCINATION and VACCINATION2 are equally likely to produce true beliefs.

Another response to this worry is to deny the assumption that VACCINATION2 produces rational beliefs. Let us assume for the sake of argument that VACCINATION satisfies NMIC because the framing effect is inevitable. (Note that this is a fair assumption in this dialectical context. I happily concede that VACCINATION is not a serious contender for VACCINATION2 when it does not satisfy VC and NMIC.) If we make this assumption, then Ken's belief that *P* in the case of VACCINATION and Naomi's belief that *R* in the case of VACCINATION2 are under the influence of the framing effect. Even in the case of VACCINATION2, there is some information about vaccinations that is available to the public, and the information must be put in some frame. For example, Naomi has access to a variety of information about vaccination, such as general statistical information of side effects, which must be put in some frame.

Of course, we can think of a radical form of VACCINATION2 where Naomi does not have any access to information that must be put in some frame. But perhaps this radical intervention leaves very limited information available to the public. This seems to be a serious violation of Naomi's freedom of inquiry. In addition, this radical intervention faces a more serious worry of generating conspiracy theories rather than true beliefs.

3.4 Rational Nudges?

The second worry about my proposal is that I am conceding too much when I admit that Ken's nudged belief in the case of VACCINATION is irrational²². It is irrational, I assume, because the belief is under the influence of the framing effect, which is irrational. But the irrationality of the framing effect can be disputed.

Several authors (e.g. Gigerenzer 2015; Levy 2019a, 2021) argue that the biases and effects that are co-opted in nudges, including the framing effect and the status quo bias, are not irrational because they can be understood as a form of (rational) testimony. (This idea is sometimes associated with the ecological conception of

²² Grundmann's (2021) view of epistemic nudging is worth mentioning in this context (although his main issue is whether nudged belief can be knowledge or not, not whether it is rational or not). He argues that nudged belief can be knowledge assuming that, first, knowledge is safe belief and, second, the belief-forming method is individuated externally, relative to the epistemic agent's environment. As Grundmann admits, however, both of his assumptions are controversial.

rationality (Gigerenzer, 2015), which I set aside in this paper. But, as Levy (2021) points out, this idea can be examined independently of the ecological conception, which is why I discuss this idea here.) These biases and effects can be explained by what is known as information leakage (McKenzie, 2004; Sher & McKenzie, 2006, 2008). For example, when an organization introduces a particular default retirement plan (RETIREMENT), the default option “leaks” an implicit recommendation by that organization. McKenzie, Liersch, and Finkelstein (2006) found that the participants exhibited a tendency to infer that the policymakers recommend an option (e.g. enrolling on a retirement plan, becoming an organ donor, etc.) when the option is set as the default, and also that this tendency is causally relevant to the status quo effect exhibited by the participants. There is nothing irrational about being influenced by recommendations when people choose their retirement plan. Again, when information about vaccinations is put in a particular frame by an authority (VACCINATION), the frame “leaks” an implicit recommendation from that authority. In fact, Altay and Mercier (2020) take their study of vaccination framing, which we mentioned earlier, to be coherent with the information leakage hypothesis. There is nothing irrational about being influenced by recommendations when people need to judge the safety of vaccinations. Thus, a strong response to the irrationality problem is available, according to Gigerenzer and Levy.

However, I have some worries about such a strong response to the irrationality problem. First, if the relevant biases are *entirely* due to information leakage, then VACCINATION constitutes an intervention in the epistemic choice set rather than the epistemic choice architecture. In other words, VACCINATION is an epistemic incentive rather than an epistemic nudge, according to my classification. Gigerenzer and Levy might be correct about VACCINATION and the framing effect, but their argument is useless in the context of defending epistemic nudges over alternatives such as epistemic incentives; VACCINATION is not an epistemic nudge after all.

Second, Gigerenzer and Levy seem to be committed to the very strong assumption that the relevant biases are *entirely* due to information leakage. (If they are only partially due to information leakage, then the irrationality problem remains.) But such a strong assumption goes beyond what current empirical research supports. In fact, McKenzie and colleagues are careful not to be committed to such a strong view. About the framing effect, McKenzie says that “*at least some of the logically equivalent frames used by framing researchers*” (emphasis added; McKenzie 2004, 876) leak some information. About the status quo bias, McKenzie and colleagues say that “*default effects occur in part because policymakers’ attitudes can be revealed through their choice of default, and people perceive the default as indicating the recommended course of action*” (emphasis added; McKenzie et al. 2006, 414).

Third, it is certainly plausible that frames and defaults leak some information. But the causal relevance of the information leakage to the framing effect or the status quo bias is not obvious. In fact, Gigerenzer’s and Levy’s accounts could sound like a *post hoc* rationalization rather than a description of the actual causal process. For instance, Ken’s choice is (irrationally) influenced by the implicit

framing effect at a particular time, but he (or somebody else who is observing Ken's behavior) can rationalize his choice at a later time by saying that he was actually influenced by the recommendation in the frame. This is analogous to the famous study by Nisbett and Wilson (1977) in which participants' choice (of a pair of nylon stockings) was (irrationally) influenced by the implicit position effect, but they rationalized their choice at a later time by saying that they had actually been influenced by some relevant qualities (of the pair of nylon stockings). This worry is especially relevant in the context of self-report studies on the causal relevance of information leakage (e.g. Experiment 4 in McKenzie et al. 2006).

4 Conclusion

I characterized ELP as a weak form of epistemic paternalism (weaker than EAP) that recommends epistemic nudges rather than stronger epistemic paternalistic interventions. Epistemic nudges (primarily) modify the epistemic choice architecture rather than the epistemic choice set. The main aim of this paper was to defend ELP from the irrationality problem, which I took to be the most urgent problem for ELP. I admitted that epistemic nudges tend to produce irrational beliefs, but argued that epistemic nudging can be justifiable in some cases, in particular the cases in which both VC and NMIC are satisfied.

A limitation of this paper is that I have been neutral on the empirical issues as to whether some actual interventions such as VACCINATION satisfy VC and NMIC. Whether an intervention satisfies VC depends on empirical facts about its effectiveness. Again, whether an intervention satisfies NMIC depends on empirical facts about the inevitability of relevant psychological effects. Currently available empirical evidence (e.g. on the framing effect and its applications) is certainly consistent with the hypothesis that some epistemic nudges such as VACCINATION do satisfy VC and NMIC, but more empirical research is needed for a stronger justification of the hypothesis.

Acknowledgements I thank Rie Iizuka, Chie Kobayashi, Shane Ryan, Bart Engelen, and three anonymous referees for helpful feedback. Early versions of this paper were presented at Kansai University (online: February 2021), National Chung Cheng University (September 2019), National Taiwan University (September 2019), and Yonsei University (August 2019). I acknowledge the support of JSPS KAKENHI (grant number 21H00464).

Declarations

Competing interests The authors have no competing interests to declare that are relevant to the content of this article.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line

to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Ahlstrom-Vij, K. (2013). *Epistemic paternalism: a defence*. Springer.
- Ahlstrom-Vij, K. (2018). Epistemic paternalism. In K. Grill, & J. Hanna (Eds.), *The Routledge Handbook of the philosophy of Paternalism* (pp. 261–273). Routledge.
- Altay, S., & Mercier, H. (2020). Framing messages for vaccination supporters. *Journal of Experimental Psychology: Applied*, 6(4), 567–578.
- Barton, A., & Grüne-Yanoff, T. (2015). From libertarian paternalism to nudging—and beyond. *Review of Philosophy and Psychology*, 6(3), 341–359.
- Bermúdez, J. L. (2022). Rational framing effects: A multidisciplinary case. *The Behavioral and Brain Sciences*, <https://doi.org/10.1017/S0140525X2200005X>
- Bortolotti, L. (2020). *The Epistemic innocence of Irrational Beliefs*. Oxford University Press.
- Bovens, L. (2009). The ethics of nudge. In T. Grüne-Yanoff, & S. O. Hansson (Eds.), *Preference change: approaches from Philosophy, Economics and psychology* (pp. 207–219). Springer.
- Bullock, E. C. (2018). Knowing and not-knowing for your own good: the limits of epistemic paternalism. *Journal of Applied Philosophy*, 35(2), 433–447.
- Druckman, J. N. (2001). Evaluating framing effects. *Journal of Economic Psychology*, 22(1), 91–101.
- Dworkin, G. (2020). Paternalism. *Stanford Encyclopaedia of Philosophy*. URL: <https://plato.stanford.edu/entries/paternalism/>
- Engelen, B. (2019). Nudging and rationality: what is there to worry? *Rationality and Society*, 31(2), 204–232.
- Engelen, B., & Nys, T. (2020). Nudging and autonomy: analyzing and alleviating the worries. *Review of Philosophy and Psychology*, 11(1), 137–156.
- Gallagher, K. M., & Updegraff, J. A. (2012). Health message framing effects on attitudes, intentions, and behavior: a meta-analytic review. *Annals of Behavioral Medicine*, 43(1), 101–116.
- Gigerenzer, G. (2015). On the supposed evidence for libertarian paternalism. *Review of Philosophy and Psychology*, 6(3), 361–383.
- Goldman, A. I. (1991). Epistemic paternalism: communication control in law and society. *The Journal of Philosophy*, 88(3), 113–131.
- Grill, K. (2014). Expanding the nudge: designing choice contexts and choice contents. *Rationality Markets and Morals*, 5, 139–162.
- Grüne-Yanoff, T., & Hertwig, R. (2016). Nudge versus boost: how coherent are policy and theory? *Minds and Machines*, 26(1), 149–183.
- Grundmann, T. (2021). The Possibility of epistemic nudging. *Social Epistemology*, <https://doi.org/10.1080/02691728.2021.1945160>
- Hagman, W., Andersson, D., Västfjäll, D., & Tinghög, G. (2015). Public views on policies involving nudges. *Review of Philosophy and Psychology*, 6(3), 439–453.
- Hansen, P. G., & Jespersen, A. M. (2013). Nudge and the manipulation of choice: a framework for the responsible use of the nudge approach to behaviour change in public policy. *European Journal of Risk Regulation*, 4(1), 3–28.
- Hausman, D. M., & Welch, B. (2010). Debate: to nudge or not to nudge. *Journal of Political Philosophy*, 18(1), 123–136.
- Banerjee, S., & John, P. (2021). Nudge plus: incorporating reflection into behavioral public policy. *Behavioural Public Policy*, 5717(332), 1–16.
- John, P., Smith, G., & Stoker, G. (2009). Nudge nudge, think think: two strategies for changing civic behaviour. *The Political Quarterly*, 80(3), 361–370.
- Jolls, C., Sunstein, C. R., & Thaler, R. (1998). A behavioral approach to law and economics. *Stanford Law Review*, 1471–1550.
- Kühberger, A. (1998). The influence of framing on risky decisions: a meta-analysis. *Organizational Behavior and Human Decision Processes*, 75(1), 23–55.
- Levy, N. (2017). Nudges in a post-truth world. *Journal of Medical Ethics*, 43(8), 495–500.

- Levy, N. (2019a). Nudge, nudge, wink, wink: nudging is giving reasons. *Ergo*, 6(10), 281–302.
- Levy, N. (2019b). No-platforming and higher-order evidence, or anti-anti-no-platforming. *Journal of the American Philosophical Association*, 5(4).
- Levy, N. (2021). *Bad beliefs: why they happen to good people*. Oxford University Press.
- McKenna, R. (2020). Persuasion and epistemic paternalism. In G. Axtell, & A. Bernal (Eds.), *Epistemic paternalism: conceptions, Justifications and Implications* (pp. 91–106). Rowman & Littlefield.
- McKenzie, C. R. (2004). Framing effects in inference tasks—and why they are normatively defensible. *Memory & Cognition*, 32(6), 874–885.
- McKenzie, C. R., Liersch, M. J., & Finkelstein, S. R. (2006). Recommendations implicit in policy defaults. *Psychological Science*, 17(5), 414–420.
- Meehan, D. (2020). Epistemic vices and epistemic nudging: a solution? In G. Axtell, & A. Bernal (Eds.), *Epistemic paternalism: conceptions, Justifications and Implications* (pp. 249–261). Rowman & Littlefield.
- Mertens, S., Herberz, M., Hahnel, U. J., & 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.
- Nisbett, R. E., & Wilson, T. D. (1977). Telling more than we can know: verbal reports on mental processes. *Psychological Review*, 84(3), 231–259.
- Pritchard, D. (2013). Epistemic paternalism and epistemic value. *Philosophical Inquiries*, 1(2), 9–37.
- Rini, R. (2017). Fake news and partisan epistemology. *Kennedy Institute of Ethics Journal*, 27(2), E–43.
- Rothman, A. J., & Salovey, P. (1997). Shaping perceptions to motivate healthy behavior: the role of message framing. *Psychological Bulletin*, 121(1), 3.
- Ryan, S. (2016). Paternalism: an analysis. *Utilitas*, 28(2), 123.
- Ryan, S. (2018). Libertarian paternalism is hard paternalism. *Analysis*, 78(1), 65–73.
- Samuels, R., & Stich, S. (2004). Rationality and psychology. In A. R. Mele, & P. Rawling (Eds.), *The Oxford Handbook of rationality* (pp. 279–300). Oxford University Press.
- Samuels, R., Stich, S., & Bishop, M. (2002). Ending the rationality wars: how to make disputes about human rationality disappear. In R. Elie (Ed.), *Common sense, reasoning and rationality* (pp. 236–268). Oxford University Press.
- Schmidt, A. T. (2019). Getting real on rationality—behavioral science, nudging, and public policy. *Ethics*, 129(4), 511–543.
- Schmidt, A. T., & Engelen, B. (2020). The ethics of nudging: an overview. *Philosophy Compass*, 15(4), e12658.
- Sher, S., & McKenzie, C. R. (2006). Information leakage from logically equivalent frames. *Cognition*, 101(3), 467–494.
- Sher, S., & McKenzie, C. R. (2008). Framing effects and rationality. In N. Chater, & M. Oaksford (Eds.), *The probabilistic mind: prospects for bayesian cognitive science* (pp. 79–96). Oxford University Press.
- Stanovich, K. E. (2004). *The Robot's Rebellion: finding meaning in the age of Darwin*. University of Chicago Press.
- Stein, E. (1996). *Without good reason: the rationality debate in Philosophy and Cognitive Science*. Oxford University Press.
- Sunstein, C. R. (2018). Misconceptions about nudges. *Journal of Behavioral Economics for Policy*, 2(1), 61–67.
- Sunstein, C. R. (2020). *Behavioral Science and Public Policy*. Cambridge University Press.
- Sunstein, C. R., & Thaler, R. H. (2003). Libertarian paternalism is not an oxymoron. *The University of Chicago Law Review*, 70(4), 1159–1202.
- Thaler, R. H., & Sunstein, C. R. (2003). Libertarian paternalism. *American Economic Review*, 93(2), 175–179.
- Thaler, R. H., & Sunstein, C. R. (2008). *Nudge: improving decisions about Health, Wealth, and happiness*. Yale University Press.
- Thaler, R. H., & Sunstein, C. R. (2021). *Nudge: The Final Edition*. Penguin Books.
- Todd, P. M., & Gigerenzer, G. (2007). Environments that make us smart: ecological rationality. *Current Directions in Psychological Science*, 16(3), 167–171.
- Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and Biases. *Science*, 185(4157), 1124–1131.
- Tversky, A., & Kahneman, D. (1981). The framing of decisions and the psychology of choice. *Science*, 211(4481), 453–458.

Vallier, K. (2016). On the inevitability of nudging. *Georgetown Journal of Law & Public Policy*, 14, 817–828.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.