http://social-epistemology.com
ISSN: 2471-9560

The Possibility of Epistemic Nudging: Reply to My Critics

Thomas Grundmann, University of Cologne, thomas.grundmann@uni-koeln.de

---

Grundmann, Thomas 2021. "The Possibility of Epistemic Nudging: Reply to My Critics." *Social Epistemology Review and Reply Collective* 10 (12): 28-35. https://wp.me/p1Bfg0-6m6.

SERRC
Social Epistemology
Review & Reply Collective
10 (12): 28-35. 2021.
https://wp.me/p1Bfg0-6m6

In "The Possibility of Epistemic Nudging" (2021), I address a phenomenon that is widely neglected in the current literature on nudges: intentional *doxastic nudging*, i.e. people's intentional influence over other people's beliefs, rather than over their choices. I argue that, at least in brute cases, nudging is not giving reasons, but rather bypasses reasoning altogether. More specifically, nudging utilizes psychological heuristics and the nudged person's biases in smart ways. The goal of my paper is to defend the claim that nudging, even when it bypasses reasoning, can result in justified beliefs and knowledge.

As I argue, it takes two things to accomplish this goal: suitable meta-epistemological views and appropriate circumstances. If a broadly reliabilist account of justified beliefs and knowledge is correct, and if the relevant belief-forming methods are externally individuated in the right way, then nudging to knowledge is possible. If, in addition, the nudger is knowledgeable, epistemically benevolent and systematically effective, then nudging to knowledge will become reality. In the paper, I use the thought experiment of POLITICAL LOYALTY to illustrate my point: Although biased party members hesitate to believe in the proven guilt of their charismatic leader, they finally become convinced of his guilt by negative frames that make him appear unsympathetic and suspicious. I propose that, intuitively, the party members achieve knowledge of their leader's guilt.

In their replies Neil Levy (2021) and Jonathan Matheson and Valerie Joly Chock (2021), put pressure on my argument from different angles. Levy thinks that a better case can be made for his view that nudging is giving testimonial reasons, and finds my objections to this view unconvincing. Matheson and Joly Chock, on the other hand, point out that acquiring knowledge through nudging (i.e. epistemic nudging) is compatible with evidentialism, even if nudging is not giving reasons. On their view, evidentialism provides an explanation of epistemic nudging that is superior to my own account, which, according to them, also suffers from a number of counterintuitive consequences. I am grateful to my critics for raising these concerns, because considering them deepens our perspective on the target phenomenon, and has made me think harder about the relevant epistemological issues. Nevertheless, I am convinced that my core claims can be defended against these criticisms.

Reply to Levy: Is Nudging Giving Reasons?

I will start with Levy's concerns. By way of background, let me quickly remind the reader of how I take nudging to work. Although it may sometimes involve deliberative cognitive processes (e.g. placing warnings on products), the bulk of nudges are targeted at S1-processes that are automatic and unreflective.[1] In the paper, I label these as *brute nudges*. I take it that the effects of brute nudges are typically mediated by the heuristics of the person who is nudged. Heuristics are associative cognitive mechanisms that reliably lead to true beliefs under specific circumstances, but systematically misfire outside of this context, where they turn into biases. Moreover, heuristics are not sensitive to further information, but lead robustly to their effects.

[1] For the following, see Kahneman 2011.

Even when heuristics reliably result in true beliefs, they typically respond to cues rather than reasons (or evidence). Heuristics lack the inferential or appropriate fittingness relation between input and output beliefs that is required for reason-responsiveness. Consider, for example, the availability heuristic. This delivers a belief about frequencies, based on how easily instances of a certain type are retrieved from memory. When things of one kind come to my mind more easily than things of another kind, I will take the frequency of the former to be higher than that of the latter. However, one cannot infer that something is more frequent than something else from the fact that it is more easily available. I propose that brute nudging typically operates on the basis of heuristics.

Levy disagrees with me. He thinks that "nudging is (typically, at any rate) giving reasons" (2021, 44). More specifically, he claims that "nudging is […] a way of offering testimony." And, when it succeeds, "[p]eople respond in ways that reflect this implicit testimony." In my paper, I present various putative counterexamples to Levy's thesis. My thought was that default effects, social referencing, framing and the affect heuristic are not reason-responsive. In his reply, Levy insists that each of these phenomena can be interpreted as being reason-responsive. He presents ingenious reconstructions of these effects as rational, albeit automatic and non-reflective, responses to implicit testimony. Before I address these effects one by one, it should be clear from the outset what is required to establish that nudging effects are reason-responsive. First, Levy must offer a potential explanation of the effect as reason-responsive; and second, he must argue that his explanatory hypothesis correctly describes the actually operative mechanism.

What about the *default effect*? Treating some option (or proposition) as default will increase the likelihood of this option being chosen (or of this proposition being believed) by the person who is nudged. In my paper, I claim that default effects can be explained by the effort that the agent must make in order to decide or believe contrary to the default. Accordingly, default effects are explained by the (cognitive) laziness of an agent who wants to avoid making any additional effort. In his response to me, Levy correctly points out that an alternative psychological explanation is available: By placing some option in the default position, the nudger conveys the information that this is the option that she recommends (Levy 2021, 45; see also McKenzie et al. 2006). So, the recipient receives evidence (i.e. awareness of the default) of a recommendation that may itself serve as higher-order evidence of some underlying first-order evidence that motivated the recommendation. Recognizing a default is thus (rather weak) third-order evidence for the default option being the right choice, or for the default proposition being true.

When we look at recent empirical studies of default effects, implicit recommendation figures as one explanatory factor among others, such as effort (or ease) and status quo bias, with the default as the reference point (Dinner et al. 2011). A recent meta-study suggests that implicit recommendation and status quo bias are the most dominant mechanisms (Jachimowicz et al. 2019, 174). But it seems clear to me that effort will play at least some role. Suppose that you are informed that the products in a supermarket are randomly distributed across different shelves. This defeats the assumption that the default (being placed at eye-level) implicitly expresses a recommendation. Nevertheless, you will surely continue to show a (slight) preference for products at eye-level. If this is correct, the simplest explanation will be that it is easier for you to select products at eye-level than picking up products from lower or higher shelves. At any rate, the psychological literature does not privilege implicit

SERRC
Social Epistemology
Review & Reply Collective
10 (12): 28-35. 2021.
https://wp.me/p1Bfg0-6m6

recommendation as the exclusive explanation of the default effect. But then one cannot conclude that this effect can *generally* be explained via giving and responding to testimonial reasons.

There is a further concern with the view that default effects are understandable as rational responses to implicit recommendations. Understood this way, the default can only constitute *weak* evidence for the default option. Default effects are, however, quite substantial, and influence our choices more strongly than even explicit recommendations. The case of organ donation vividly illustrates this point. Politicians and physicians have long called on people to donate their organs after death. Nevertheless, the rate of donors strongly varies with the default. This rate is much higher when the default is to opt out. This suggests that default effects cannot be reduced to the uptake of testimony.

Levy (2021, 45) believes that the explanation of default effects as rational responses to implicit recommendations is superior to the one favored by me, i.e. the explanation via laziness, because the former but not the latter facilitates a uniform explanation of nudging. Levy thinks that nudging can generally be explained as triggering rational responses by implicit recommendations. *Framing* seems to be a case in point. While sensitivity to frames can be fully captured by the fact that framing an option a certain way (e.g. as a gain rather than a loss) conveys the information that the speaker recommends this option, cognitive laziness cannot account for this phenomenon (Levy 2021, 46).

I am not convinced by Levy's argument, however. First, more uniform explanations are not always more likely to be true. For example, conspiracy theories offer extremely unified explanations of a whole range of events by a single agent. But this kind of explanation is typically too unified to be true. It might just be a fact that nudging effects result from various psychological mechanisms, just as automatic cognitive responses result from a plurality of different heuristics. Second, Levy's explanation of framing abstracts away from the fact that frames can also have their effects in non-conversational contexts where no testifier is involved. For example, when you consider a hypothetical case that describes an event in terms of a loss, the loss frame is operative in the same way as in conversational contexts. So Levy would need different explanations depending on whether the framing takes place inside or outside of conversational contexts. Finally, some nudging effects, such as triggering the affect heuristic, are strongly recalcitrant to Levy's explanation, as we will see shortly.

What about *social referencing*? In my paper, I argue that this related effect is a counterexample to Levy's reasoning account of nudging. What I had in mind was that people choose or believe what other people (e.g. of their in-group) choose or believe. In opposition to my assessment of this effect, Levy (2021, 46) points out that a consensus among my peers is an excellent reason for me to adopt their belief or decision, because it makes the belief's truth or the decision's rightness highly likely. I agree that my peers' consensus is, under appropriate conditions, a good reason to agree with them. The Condorcet Jury Theorem and the Miracle of Aggregation both point in this direction. However, is social referencing—or, to use another term, *social contagion*—in fact rationally responsive to this kind of reason? I doubt that it is, because social referencing robustly persists in the face of defeating information.

An eye-opening experiment was performed by Solomon Asch (1951). In this experiment, the test subject was asked which of three lines had the same length as a reference line. When confederates who collaborated with the experimenter gave an obviously false answer, the subject nevertheless agreed with them 37% of the time. It seems that this effect was not sensitive to the defeating visual information. In a different experiment, people are asked to put money into a box for every cup of coffee they get from the coffee machine. The number of paying people is significantly increased when a big picture of watching eyes is fixed behind the coffee machine. This works even though it is obvious for the subject that no real social interaction is involved. One might worry that the picture is effective simply because it conveys the information that the request to put money into the box is meant seriously. But the picture is more effective than even a serious request. In my view, both experiments suggest that it is social pressure rather than responsiveness to reasons that explains the effects.

The weakest point in Levy's account is his alternative explanation of the *affect heuristic*. This heuristic plays a crucial role in POLITICAL LOYALTY. In this case, the public relations manager makes the party members believe in their leader's guilt simply by presenting him with an unsympathetic and suspicious appearance. Levy claims that the affect heuristic relies on "testimony from […] oneself" (2021, 46). This is, of course, a merely metaphorical use of the term "testimony." I doubt that one can cash out the reasons given to the recipients in non-metaphorical terms. If the party members rely on their negative feelings about their leader when they believe that he is guilty, then these negative feelings may count as reliable cues, but they cannot be reasons for believing in his guilt, because the required inferential (or fittingness) relation is missing. If, however, they defer to their own belief about their leader, then this belief has already been formed. It cannot serve as a reason for itself. I do not see how Levy can interpret this case as an instance of reason-responsiveness.

To wrap up my discussion of Levy: He is right to propose that it is *possible* to explain many (but not all) nudging effects as triggering rational responses by giving reasons. The availability of this kind of explanation for most instances of nudging is amazing. However, the robustness and insensitivity of these processes to further defeating information suggests that the underlying psychological mechanisms are non-rational heuristics rather than reason-responsive processes. I must admit that this judgment is tentative. I fully agree with Levy that much more empirical research on the psychological mechanisms that underly nudging is called for if we are to decide this issue definitively.

First Reply to Matheson and Joly Chock: Do Epistemic Nudging and Evidentialism Go Together?

Matheson and Joly Chock's reply is complementary to Levy's. In contrast to Levy, these co-authors seem to grant me that nudging "involves the use of non-epistemic factors to bring about a belief in the nudgee" (Matheson and Joly Chock 2021, 38). But they disagree with me about the consequences of this fact. In my paper, I argue that nudging cannot result in evidence-based knowledge (or evidence-based justified beliefs), since nudging is a non-epistemic causal intervention. Matheson and Joly Chock correctly point out that my conclusion follows only under the further assumption that *every* factor that is causally

SERRC
Social Epistemology
Review & Reply Collective
10 (12): 28-35. 2021.
https://wp.me/p1Bfg0-6m6

relevant for bringing about the belief contributes to its epistemic basis. But this further assumption is, according to them, false on any reasonable account of proper basing (39).

Matheson and Joly Chock consider two such accounts. One of them is the purely doxastic account of basing. According to it, a belief is based on a reason if the agent possesses this reason and has the meta-belief that it is a good reason for the belief. If this condition is satisfied, the doxastic account maintains, the agent will hold her belief *for* this reason. Since on this account the proper basis of a belief has nothing to do with its causal grounds, nudging does not prevent the belief from being properly based on evidence (38). I strongly disbelieve, however, that purely doxastic accounts can provide sufficient conditions for proper basing.[2] Believing for a certain reason requires one's belief to *respond* to this reason. And this requirement is stronger than calling for a rationalizing perspective on the target-belief. Moreover, as Turri (2011) argues, it seems possible for an agent to take two of his reasons to be good reasons for having a certain belief, while nevertheless forming this belief on the basis of only one of them. That a causal relation is necessary for proper basing is beyond question for me.[3]

The second account Matheson and Joly Chock consider is a moderately causal account of proper basing that requires a *sufficiently strong causal connection* to hold between a belief and its basis. This condition may be satisfied by reasons even if a completely non-epistemic nudge triggers the belief's occurrence. On this view, nudging causally *enables* the subject to respond to reasons. One may call this phenomenon *nudging to reason*. I agree that this is an interesting possibility that I did not consider in my paper. If nudges were to clear the nudgee's mind by neutralizing her prior biases, or if nudges were to establish reason-responsiveness in some other way, then nudging would be compatible with an evidential basis for the belief. On this view, doxastic nudging would interact with evidence as some additional causal factor.

Although this is an interesting possibility, I am not convinced that nudging (typically) operates in this way. Recall my prior claim that nudging utilizes cognitive heuristics. These heuristics (typically) take cues as input, and do not involve reasons or evidence at any later stage of their operation. So, they bypass reasoning altogether. If this is correct, then nudging does *not* enable or reestablish reason-responsiveness in the agent's doxastic behavior. You might think that POLITICAL LOYALTY is a striking counterexample to my claim. Doesn't the nudge bring the party members' beliefs in line with what the evidence suggests, i.e. that the political leader is guilty of murder? I agree. But this doesn't show that the nudge makes the party members *responsive* to the available evidence. It only brings about a belief that is also (propositionally) justified by the available evidence. The evidence of guilt that is neglected by the party members has a dual function in my case. It motivates the manager's intervention and it guarantees that no knowledge-defeater is available to the party members. However, this evidence does not play any causal role in the explanation of their terminal beliefs. So,

[2] I also do not believe that having this meta-belief is *necessary* for the basing relation. Otherwise, small children could not possess justified beliefs, because they typically lack the epistemic concepts needed for the required meta-belief.

[3] This is not to say that a causal relation is sufficient for proper basing. The possibility of deviant causal chains shows that more is needed. I find Turri's suggestion that the reason's causation of the belief must manifest one of the agent's cognitive traits or habits quite promising (see Turri 2011).

even if the evidence of the leader's guilt had not been available to the party members, the nudge would still have brought about their belief in his guilt. Affect heuristics simply work this way. They elicit the assessment of a person solely on the basis of specific feelings about that person.

Here, then, is the argument that I should have provided to support my claim that epistemic nudging and evidentialism conflict with each other:

(1) Nudging (typically) brings about beliefs, without evidence playing any suitably strong causal role.
(2) If evidence doesn't play a suitably strong causal role in bringing about a belief, it doesn't belong to its proper basis.
(3) Therefore, beliefs that result from nudging (typically) aren't based on evidence.

Given this argument, I disagree with Matheson and Joly Chock when they conclude: "Epistemic nudging is only a problem for evidentialism when it is saddled with implausible accounts of proper basing" (2021, 39).

Second Reply to Matheson and Joly Chock: Is My Own Account in Deep Trouble?

In addition, Matheson and Joly Chock worry that my reliabilist account of knowledge and justified beliefs faces a number of standard objections, and they also raise a novel problem for it. Let me start by addressing the standard objections. First, the externally individuated nudging process must, on my account, reliably generate justified beliefs and safely generate knowledge. Since reliability and safety apply to process types rather than tokens, the relevant type must be identified first. And here the generality problem kicks in: The actually operative process belongs to many different types that may vary in their reliability (or safety); and it is indeterminate which of these types is the epistemically relevant one. Of course, I can't offer a convincing solution to the generality problem here. Notice, however, that various suggestions have been made in the literature for solving this problem (e.g. Alston 1995; Beebee 2004; Comesaña 2006; Lyons 2019).

Second, I propose that safe nudging processes *suffice* to generate knowledge and that reliable nudging processes *suffice* to produce justified beliefs. No further evidence accessible to the agent is required. This may be implausible in itself. But it also seems to motivate the wrong verdict about BonJour's Norman case (Matheson and Joly Chock 2021, 39). In this seminal case, Norman unknowingly has the reliable and non-evidential ability of clairvoyance and forms the true belief that the president is in New York City on the basis of this capacity. From Norman's point of view, he forms this belief out of the blue, without relying on any evidence. Intuitively, Norman's belief about the president's whereabouts is not justified, although it is reliably formed. I don't take these objections to be decisive, however. To begin with, instances of knowledge or justified beliefs that are not based on evidence are more commonsensical than one might think. For example, privileged self-knowledge or introspectively justified beliefs are not based on evidence of one's lower-order mental states (Shoemaker 1996). The introspected mental state cannot serve as evidence for the higher-order belief (because the latter cannot be inferred from the former), nor is there any further piece of evidence mediating the relation between the mental state and the introspective belief

SERRC
Social Epistemology
Review & Reply Collective
10 (12): 28-35. 2021.
https://wp.me/p1Bfg0-6m6

about it. It is not even clear whether testimonial knowledge is based on evidence rather than reliable cues.[4] Next, since Norman does not know about his reliable clairvoyance, its manifestation looks to him like an unreliably formed hunch. He thus acquires an undercutting defeater that removes his prima facie justification.[5] This explanation of why Norman's belief is unjustified is compatible with the claim that reliable belief-formation is sufficient for justified beliefs (if defeaters are absent).

More interesting is Matheson and Joly Chock's novel objection to my account (Matheson and Joly Chock 2021, 40-41). According to my account, doxastic nudging is an externally individuated belief-forming process that generates justified beliefs only if the nudger is knowledgeable (concerning the induced beliefs), epistemically benevolent (to the nudgee), and steers the nudgees' beliefs systematically and effectively. If the nudging process had easily produced false beliefs in some other nudgee, then it would not result in knowledge even in those nudgees who are nudged to true beliefs. Matheson and Joly Chock take this dependence of individual knowledge on group effects to be implausible, and at the same time unavoidable for me. They believe that nudging can succeed on one person by yielding the desired knowledge, while amplifying another person's false belief. In contrast to Matheson and Joly Chock, I doubt that it is always implausible for individual knowledge to depend on a method's effects on other people.

Suppose you read the time from a clock that unbeknownst to you has stopped. Since you read the clock only when it represents the time correctly (say, twice a day), you always believe in the correct time. But other people read the clock when it misrepresents the time. It seems perfectly plausible to say that here the false beliefs of other people are relevant to your lack of knowledge. Moreover, even if it were plausible that doxastic effects on other people are irrelevant to individual knowledge, my account could easily adopt this feature. One might just relativize the externally individuated belief-forming nudging process to individual nudgees. And there may be good reason for doing so. After all, certain biases may be operative in some people but absent in the person at hand. If this is true, the relevant, partly externally individuated processes should be differently typed. I don't see why this would be incompatible with my own account.

Conclusion

Let me summarize the results of my virtual debate with my critics: Brute nudging (typically) operates on cognitive heuristics and is thus unresponsive to reasons. It also (typically) does not enable or reestablish reason-responsiveness. So, if it brings about knowledge or justified beliefs in the nudgee, it must do so in some non-evidentialist manner. Finally, it has turned out that acquiring non-evidential knowledge or non-evidentially justified belief might not be

[4] What I have in mind here is not Moran's argument (2005, 5-6), according to which signs that are known to be deliberatively produced cannot serve as evidence. This argument has been convincingly criticized by Keren (2012). For me it is more relevant that anti-reductionists generally have a problem with claiming that the testifier's bare assertion suffices as evidence for its truth.
[5] For this criticism, see Grundmann 2004.

as exceptional as it appears to the evidentialist. So, epistemic nudging may be possible, even if the prospects of rational or evidentialist-friendly interpretations of nudging are rather dim.

References

Alston, William. 1995. "How to Think about Reliability." *Philosophical Topics* 23 (1): 1-29.

Asch, Solomon. 1951. "Effects of Group Pressure on the Modification and Distortion of Judgment." In *Groups, Leadership and Men* edited by Harold Guetzkow, 177-190 Pittsburgh, PA: Carnegie Press.

Beebe, James. 2004. "The Generality Problem, Statistical Relevance and the Tri-Level Hypothesis." *Nous* 38 (1): 177-195.

Comesaña, Juan. 2006. "A Well-Founded Solution to the Generality Problem." *Philosophical Studies* 129 (1): 27-47.

Dinner, Isaac et al. 2011. "Partitioning Default Effects: Why People Choose Not to Choose." *Journal of Experimental Psychology Applied* 17 (4): 332-341.

Grundmann, Thomas. 2004. "Counterexamples to Epistemic Externalism Revisited." In *The Externalist Challenge* edited by Richard Schantz, 65-77 (Berlin and New York: De Gruyter.

Grundmann, Thomas. 2021. "The Possibility of Epistemic Nudging." *Social Epistemology* 1-11. https://doi.org./10.1080/02691728.2021.19451560.

Jachimowicz, Jon et al. 2019. "When and Why Defaults Influence Decisions: A Meta-Analysis of Default Effects." *Behavioural Public Policy* 3 (2): 159-186.

Kahneman, Daniel. 2011. *Thinking, Fast and Slow*. London: Penguin.

Keren, Arnon. 2012. "On the Alleged Perversity of the Evidential View of Testimony." *Analysis* 72 (4): 700-707.

Levy, Neil. 2021. "Nudging is giving Testimony: A Response to Grundmann." *Social Epistemology Review and Reply Collective* 10 (8): 43-47.

Lyons, Jack. 2019. "Algorithms and Parameters: Solving the Generality Problem for Reliabilism." *Philosophical Review* 128 (4): 463-509.

Matheson, Jonathan and Valerie Joly Chock. 2021. "The Possibility of Epistemic Nudging: Reply to Grundmann." *Social Epistemology Review and Reply Collective* 10 (8): 36-42.

McKenzie, Craig et al. 2006. "Recommendations Implicit in Policy Defaults." *Psychological Science* 17 (5): 414-420.

Moran, Richard. 2005. "Getting Told and Being Believed." *Philosophers' Imprints* 5: 1-29.

Shoemaker, Sidney. 1996. *The First-Person Perspective and Other Essays*. Cornell University, New York: Cambridge University Press.

Turri, John. 2011. "Believing for a Reason." *Erkenntnis* 74 (3): 383-397.