

NUDGE TRANSPARENCY IS NOT REQUIRED FOR NUDGE RESISTIBILITY

GABRIEL DE MARCO

Oxford Uehiro Centre for Practical Ethics, University of Oxford

THOMAS DOUGLAS

Oxford Uehiro Centre for Practical Ethics and Jesus College, University of Oxford

In discussions of nudging, transparency is often taken to be important; it is often suggested that a significant moral consideration to take into account when nudging is whether the nudge is transparent. Another consideration taken to be relevant is whether the nudge is easy to resist. Sometimes, these two considerations are taken to be importantly related: if we have reason to make nudges easy to resist, then we have reason to make them transparent, insofar as a nudge's transparency is relevant to whether the nudge is easy to resist. In this paper, we critically scrutinize this view. First, we draw out the purported connections between transparency and resistibility, combining them into what we call the *Resistibility-to-Transparency Argument (RTA)*. Then, we do some conceptual groundwork elaborating on different forms of transparency mentioned in the debate. With this in place, we argue that in order to be plausible, the *RTA* must appeal to certain forms of transparency; those on which the transparency of a nudge depends on the nudgee's ability to become aware of the relevant facts about the nudge. We then argue against the common assumption that in order for a nudge to be easy to resist for an individual, that individual needs to have the capacity to easily become aware of the relevant nudge facts. We conclude that the connection between the easy resistibility of a nudge and its transparency is, at best, a weak one.

1. Intro

Consider the following cases:

Food Placement. In order to encourage healthy eating, cafeteria staff place healthy food options at eye-level, whereas unhealthy options are placed

Contact: Gabriel De Marco <gabriel.demarco@philosophy.ox.ac.uk>
Thomas Douglas <thomas.douglas@philosophy.ox.ac.uk>

lower down. Diners are more likely to pick healthy foods and less likely to pick unhealthy foods than they would have been had foods instead been distributed randomly.

Default Registration. In application forms for a driver's license, applicants are asked whether they wish to be included on the organ donation registry. In order to opt out, one needs to tick a box; otherwise, the applicant will be registered as an organ donor. The form was designed in this way in order to get more organ donors; applicants are more likely to be registered than they would have been had the default been *not* being included in the registry, and applicants needed to tick the box to opt *in*.

Interventions like these two are often called *nudges* (Thaler & Sunstein 2009; 2021).¹

Though many agree that it is, at least sometimes, permissible to implement nudges, there is a thriving debate about what features of nudges are relevant to their permissibility. One such feature is the transparency of the nudge. It is often suggested that transparency is relevant to—and perhaps even a necessary condition for—its permissibility.

Different explanations have been given for why transparency is relevant to nudge permissibility. For instance, some suggest that transparency can help with ensuring that nudgers can be held accountable, or with ensuring that nudging respects autonomy.² But one explanation alluded to by many is that the transparency of a nudge is relevant to its resistibility, which is another feature often thought to be relevant to the permissibility of nudging. Our aim in this paper is to evaluate this explanation, and, more generally, to evaluate the relationships between (1) the (easy) resistibility of a nudge, (2) the nudgee's capacity to (easily) become aware of the nudge, and (3) the nudge's transparency.

Here are some claims that some have made, or at least come close to making, regarding these relationships, and which can be arranged—as we have arranged them here—into an argument for nudge transparency.

P1. A nudge is permissible only if it is easy to resist.³

P2. A nudge is easy to resist only if the agent can easily become aware of it.

1. We do not offer a precise account of what counts as a nudge, and instead rely on paradigmatic examples that are usually taken to be nudges.

2. For discussion, see, for instance Anderson (2010: 374), Blumenthal-Barby (2021: 121), Clavier (2018: 370–71), Grüne-Yanoff (2012: 638), Ivanković and Engelen (2019), Rebonato (2014: 360), Schmidt and Engelen (2020), Soled et al. (2021: 64), Sunstein (2014: 584; 2015a: 428–29; 2015b: 144–59; 2016: 231–32), Thaler and Sunstein (2021: 327–28), Wilkinson (2013).

3. A brief clarification: we understand this as the ability to easily resist a nudge when one first encounters it.

- P_3 . An agent is capable of easily becoming aware of a nudge only if it is transparent.
- C. A nudge is permissible only if it is transparent.

Call this the *Resistibility-to-Transparency Argument*, or the *RTA* for short. P_1 – P_3 are strong in that they express necessary conditions. Typically, theorists commit to a weaker variation of P_1 , on which a nudge's not being easy to resist gives us a significant *pro tanto* reason against implementing the nudge.⁴

P_1 , or some variant of it, might be justified by virtue of further principles. For example, one might think that whether a nudge is easy to resist is relevant to the nudge's permissibility because it is relevant to whether the nudge preserves the nudgee's autonomy. The resistibility of a nudge might affect the nudgee's autonomy because, for instance, it affects whether the nudgee retains control over her actions, or because it affects whether the nudge counts as a controlling or substantially controlling influence.⁵

P_2 can also be weakened. For instance, one might think that although having the capacity to easily become aware of the nudge can help to make the nudge easy to resist, it is not necessary. However, most who hold some version of P_2 hold the strong version. For example, in her discussion of (non-)transparent nudges, Blumenthal-Barby maintains that "in order for something to count as a mere nudge, it must be easy to avoid (substantially noncontrolling), and it is hard to see how someone could easily avoid something they are unaware of" (Blumenthal-Barby 2021: 119).⁶ We will return to a weaker version of P_2 below.

P_3 will seem implausibly strong to many. It holds that a necessary condition of an agent's having the capacity to easily become aware of a nudge is that the nudge itself is transparent. Yet one might think that people can, at least in some cases, (easily) come to be aware of a nudge-style influence even if it is not transparent. For instance, if one reads the introduction to this paper right before going to a cafeteria in which all and only healthy foods are in prominent locations, one can easily become aware of the nudge even if the cafeteria does nothing to inform customers about it. However, as we discuss later (in Section 2), on some understandings of transparency, P_3 is quite plausible.

Once one adopts some variation of claims P_1 – P_3 , C—or perhaps a weaker version of it—will seem to follow straightforwardly. If we have, at least, a rea-

4. See, for example, Engelen (2019a; 2019b: 216), Kipper (2020), Lades and Delaney (2022), Majtenyi and Ruble (2019), Mills (2015), Moles (2015), Saghai (2013a; 2013b).

5. See, for example, Kipper (2020), Majtenyi and Ruble (2019), Moles (2015), Saghai (2013a).

6. For more examples, see Barton and Grüne-Yanoff (2015), Capron (2015), Engelen (2019b), Gagnon and Le Nagard (2015), Lades and Delaney (2022), Ledderer et al. (2020), Mills (2015), Moles (2015), Nys and Engelen (2017), Saghai (2013a; 2013b).

son in favor of making nudges easy to resist, *and* a nudge's being easy to resist requires the capacity to easily become aware of the nudge, then it seems that, when we nudge, we have a reason to make it such that people can easily become aware of the nudge. The most plausible way to do this would seem to be via making the nudge transparent.

And some authors indeed endorse, or at least entertain, something close to this whole argument. Consider, for instance, how Barton and Grüne-Yanoff describe the state of some of the debate: "to be able to avoid a nudge, one needs in a first place to know that one is being nudged, which requires a sufficient degree of transparency of the nudge," and "[t]hus, to be avoidable (or resistible), nudges must be transparent enough" (2015: 347).⁷

To be clear, we do not mean to suggest that this is the only way to argue for the claim that we ought to, or have reason to, be transparent when nudging; there are other potential justifications for such a claim (though we remain agnostic on their success, for the purposes of this paper). However, an argument along the lines of *RTA* does seem to us to be a common way of coming to such a conclusion.

In this paper, we aim to explore the relationships between resistibility, awareness, and transparency—at least in the context of nudging—and the implications of these relationships for *RTA*. First, however, it will be necessary to do some conceptual groundwork. In the next section, we discuss the notion of transparency. We distinguish four different types of transparency and specify which types *P*₃ must refer to, if it is to be plausible. We then move on to consider *P*₂, a common assumption about the resistibility of nudges, and argue that it is false. Finally, we briefly consider the prospects of a significantly weaker version of the *RTA*, and offer some concluding thoughts about the implications of the arguments presented here.

2. Transparency

In saying that we ought to be, or at least, that we have strong reason to be, transparent when nudging, it is important to be clear both about (a) what facts need to be transparent, and (b) what it takes to be transparent about them. We think that the success of the *RTA* will partly depend on what answers one gives to these questions; and below, we consider some answers that theorists have given.

7. See also Mills (2015: 499ff.), and for a weaker version, Ivanković and Engelen (2019: 59).

2.1. What We Have Reason to Be Transparent About

If a nudge is to be transparent, what is it, exactly, that needs to be transparent? Consider again the nudge used in *Food Placement*. There are multiple facts that may, or may not be, transparent. For example:

- F*₁. The healthy food is at eye level.
- F*₂. The food's placement might influence one's deliberation and/or decision about what to choose.
- F*₃. Someone has ensured that *F*₁, with the knowledge that *F*₂, in order to influence nudges into picking the healthy food.

Or consider, instead, *Default Registration*. In this case, there are also multiple facts that may, or may not be, transparent:

- D*₁. Being registered for the organ donor list is the default.
- D*₂. The fact that this is the default will make one more likely to register for organ donation.
- D*₃. Someone has ensured that *D*₁, with the knowledge that *D*₂, in order to influence nudges into registering for the organ donor list.

We will have reason to refer to these different sorts of facts later, so we name them now. Call facts like *F*₁ and *D*₁ *environmental facts*. These are merely about the nature of the environment in which the agent is making a choice, and pick out a feature used by the nudger. Call facts like *F*₂ and *D*₂ *effect facts*. These facts concern the effects that features of the environment described in environmental facts can have on those exposed to them. Call facts like *F*₃ and *D*₃ *nudger facts*. These facts concern the actions, knowledge, and intentions of an agent (the nudger): the agent has ensured that the environmental facts hold, in knowledge of the effect facts, with the intention of influencing the person exposed to these environmental features.

We have exemplified these three types of facts in relation to particular nudges, but it is possible to imagine more general forms of them. Using the default nudge as an example, consider:

- G*₁. People often face choices in which one option is the default.
- G*₂. In such cases, people are often more likely to choose the default option than they would have been had it not been the default.
- G*₃. In many cases, someone has decided what should be the default, in the knowledge of *G*₂, in order to influence people into choosing that option.

If we have reason to be transparent about nudging, we likely have reason to be transparent about all three sorts of facts. Yet different views on transparency differ in the level of specificity of those facts. For instance, Hausman and Welch suggest that: “[p]eople should be informed of government efforts to exploit their decision-making foibles” (2010: 135). This would seem to require only transparency regarding fairly general environment, effect and nudger facts, perhaps along the lines of G_{1-3} .

According to Bovens, however, it needs to be the case that a watchful agent can “identify the intention of the choice architecture” (2009: 217) and that “it is possible to recognize each token interference” (2009: 216).⁸ This suggests facts much closer to the level of description found in F_{1-3} and D_{1-3} above. For instance, the fact that this particular food display was structured in order to influence the customer’s choices, or that the choice of default on this particular form might influence one’s choice about whether to become an organ donor.

2.2. What Does Transparency Require?

Different theorists, then, take different views about which facts need to be transparent in order for a nudge to be transparent. They also take different views regarding what it takes for those facts to be transparent. Here, we introduce what we take to be some of the main types, or versions, of transparency.

One type of transparency is a matter of the nudger’s being willing and/or able to publicly defend the nudge policy. For instance, Thaler and Sunstein refer to Rawls’s publicity principle, which they interpret as follows: “[i]n its simplest form, the publicity principle bans government from selecting a policy that it would not be able or willing to defend publicly to its own citizens” (2009: 247).⁹ And more recently, when discussing transparency and the publicity principle, they suggest that “[i]f government officials [nudge], they should be happy to reveal both their methods and their motives” (Thaler & Sunstein 2021: 328). This type of nudge-transparency implies that the nudger (at least where the nudger is the state) is able and/or willing to defend the policy of which the nudge is part, which would involve revealing general forms of all three sorts of facts mentioned above. We will call this type of transparency *hypothetical disclosure* transparency. Transparency of this sort does not require that the nudger *actually* do anything to communicate or defend the policy.

8. What makes an agent a *watchful* agent? Bovens does not offer an account, but does suggest that it is something that ordinary, untrained individuals can choose to become. For more detailed discussions of the concept of “watchfulness” at play, see Schmidt (2017: 410), Ivanković and Engelen (2019).

9. See also Thaler and Sunstein (2021: 327).

Hausman and Welch suggest that this type of transparency is too weak and argue that

One important way to protect against abuse and to respect autonomy is to make sure that the government actually inform people of efforts to shape their choices, not merely that it be able and willing to do so. People should be informed of government efforts to exploit their decision-making foibles, even if doing so undercuts their effectiveness. (2010: 135)

And in recent work, Sunstein offers a stronger form of the publicity-based transparency requirement, suggesting that nudges “should be visible, scrutinized, and monitored. To the extent feasible [nudge policies] should be subject to public scrutiny in advance, often through notice-and-comment rulemaking” (2015b: 148).

On this type of transparency, the nudger needs to ensure that the nudge is subject to public scrutiny—as suggested by Sunstein—or that people are informed of the nudge—as suggested by Hausman and Welch. Being transparent in this way requires *actual*, not merely hypothetical, disclosure and/or informing. A further detail to be developed concerns whom, in particular, the policy needs to be disclosed to, or who needs to be informed of it. We return to this shortly.

Another type of nudge transparency frequently encountered in the literature depends on the nudgee’s epistemic state. For instance, Soled and co-authors suggest that “[n]ontransparent nudges operate by bypassing a person’s awareness. Often, but not always, defaults and certain framing effects powerfully influence choice without the person’s being at all aware that they are influenced by these phenomena” (2021: 64).¹⁰ In a similar vein, Kiener justifies his view that a nudge described in a clinical vignette is not transparent by noting that the nudgee “is not at all aware that the medical team arranged the situation to exploit her confirmation bias and make her appreciate the medical evidence” (2021: 4206). Transparency of this sort requires that the nudgee is aware of a relevant fact (or set of facts). Call this *actual awareness* transparency. A nudge-style influence can be transparent in this way even if the nudger does not disclose effect or nudger facts to the nudgee, and even if the nudger is neither willing nor able to defend the nudge publicly. It is enough that the nudgee un.masks the nudge.¹¹

10. For a similar description, see Blumenthal-Barby (2021: 118–19).

11. Of interest may be a case introduced by Blumenthal-Barby, which involves the nudgee’s being aware of the nudge—and so the nudge is transparent on the actual awareness account—while the nudger is neither trying to, nor willing to, disclose the fact that they are nudging the nudgee—and so is not transparent in the ways required by either hypothetical or actual disclosure transparency (Blumenthal-Barby 2021: 121).

A somewhat weaker type of transparency would require that the nudgee can easily become aware of the relevant fact (or set of facts) (Hansen & Jespersen 2013: 17–18), or that a watchful agent can (Bovens 2009: 217; Schmidt 2017: 409–10); call such transparency *hypothetical awareness* transparency.¹²

To summarize, we have distinguished, roughly, the following four types of transparency.

Hypothetical Disclosure Transparency: A nudge is transparent when the nudger is able and/or willing to defend the policy of which the nudge is part, which would involve revealing the relevant facts.

Actual Disclosure Transparency: A nudge is transparent when the nudger discloses/informs the relevant facts to the public (or perhaps a subset of the public).

Hypothetical Awareness Transparency: A nudge is transparent when the (perhaps watchful) nudgee can easily become aware of the relevant facts.

Actual awareness Transparency: A nudge is transparent when the nudgee is actually aware of the relevant facts.

With these distinctions in place, we are now in a position to make some initial points about the relationship between the transparency of a nudge and the nudgee’s awareness of the nudge, or their capacity to easily become aware of it.¹³

2.3. Transparency and Awareness

Recall:

*P*₃. An agent is capable of easily becoming aware of a nudge only if it is transparent.

We can now see that the plausibility of this claim will depend on what type of transparency it references. If it refers to hypothetical awareness transpar-

12. As we read them, Ivanković and Engelen (2019) use “transparent by design” as a version of transparency that is either an actual, or a hypothetical, awareness version.

13. For a different way of categorizing types of transparency, see Ivanković and Engelen (2019: 50–53).

ency, P_3 turns out to either be a conceptual truth, or close to it. If understood as referring to this type of transparency, P_3 effectively comes to: “an agent is capable of easily becoming aware of a nudge only if the (perhaps watchful) agent can easily become aware of the relevant nudge facts.” By contrast, if it refers to actual awareness transparency, P_3 is clearly too strong: an agent can have the capacity to easily become aware of a nudge without being, in fact, aware of it.

What about disclosure-based forms of transparency? These forms of transparency cannot support the strong claim, made by P_3 , that transparency is *necessary* for nudgees to have the capacity to easily become aware of a nudge. There are clearly cases in which individuals can easily unmask a nudge without being explicitly informed about it, for example, because the nudges happen to be well-informed about nudge practices, and vigilant as to their presence.¹⁴ And if actual disclosure is not necessary for unmasking a nudge, neither is hypothetical disclosure.

Nevertheless, forms of actual disclosure transparency that require disclosure to nudgees may support a weaker version of P_3 restricted to cover only some nudgees. For any nudge, there will presumably be some nudgees—for example, those not already well-informed about this type of nudge—who can easily become aware of the nudge if certain nudge facts are disclosed to them, and would not have this capacity otherwise.

However, this does not apply to hypothetical disclosure transparency, since this sort of transparency has little or no bearing on a nudgee’s awareness of the nudge: a nudger’s merely being willing and/or able to defend the nudge policy would not confer the capacity to easily become aware of the relevant facts on nudgees. Similar thoughts apply to versions of actual disclosure transparency in which the disclosure is made to persons other than the nudgees. Forms of actual disclosure transparency in which the nudge must be disclosed *to the nudgees* are more plausibly relevant to the nudgees’ awareness.

In light of these thoughts, we suggest that proponents of the *RTA*, or weaker versions of it, would do best to invoke either hypothetical awareness transparency, or a form of actual disclosure transparency in which the disclosure must be made to nudgees.

With this in place, we now turn to P_2 , which concerns the relationship between the (easy) resistibility of a nudge, and the capacity to easily become aware of the nudge.

14. In the extreme case, suppose that Cass Sunstein walks into a cafeteria in *Food Placement*. The nudge need not be disclosed to him for it to be the case that he can easily become aware of it.

3. Awareness and Resistibility

Recall:

P2. A nudge is easy to resist only if the agent can easily become aware of it.

The plausibility of this claim—which we will henceforth refer to as the ‘awareness condition on easy resistibility’—will clearly depend on what it takes for a nudge to be easy to resist. Yashar Saghai has provided the most developed and influential account of this.¹⁵ Saghai suggests that a nudge intended to get nudges to ϕ is easily resistible if the nudge is able to effortlessly oppose the nudger’s pressure to get her to ϕ , were she to not want to ϕ (2013a: 489). The notion of ability required is:

Ability for Easy Resistibility. B is able to easily resist A’s influence [if and only if]:¹⁶

1. B has the capacity to become aware of A’s pressure to get her to ϕ (attention-bringing capacities);
2. B has the capacity to inhibit her triggered propensity to ϕ (inhibitory capacities);
3. B is not subject to an influence, or put in circumstances that would significantly undermine the relatively effortless exercise of attention-bringing and inhibitory capacities. (Saghai 2013a: 489)

For our purposes, the most relevant conditions are the first and third. The first mentions a capacity to become aware of the pressure to get one to ϕ ; and the third claims that the agent is not subject to some sort of obstacle that would prevent the effortless exercise of this capacity. Combined, they suggest that in order for a nudge to be easy to resist for an agent, she must have the capacity to, relatively effortlessly, become aware of the nudger’s pressure to get her to ϕ . This can be seen as simply a fuller specification of the awareness condition on easy resistibility; one on which ease of resistance is understood in terms of effort, and the relevant type of awareness is taken to be awareness of the *nudger’s pressure*.

15. One brief clarification: Saghai does not merely offer this as an account of a nudge’s being easy to resist; he suggests that for an intervention to *be* a nudge, as opposed to something more serious, it needs to be easy to resist.

16. We have replaced a potentially ambiguous “when” with “if and only if.” In replies to comments on his original paper, Saghai makes it explicit that “when” should be understood as a bi-conditional: “B can easily resist A’s influence *if and only if* B is not subject to influences, or put in circumstances that would either significantly undermine the relatively effortless exercise of the capacities that enable B to oppose A’s pressure to get her to ϕ , or weaken those very capacities” (2013b: 499–500, emphasis ours).

As noted earlier, the awareness condition on easy resistibility has also been endorsed, albeit more briefly, by others.¹⁷ However, we disagree with it. We think, and in the remainder of this section argue, that, in order for a nudge to be easy to resist for an agent, she need not be aware of the nudger's pressure, or indeed of any effect or nudger facts, nor need she have the capacity to easily become aware of either of these sorts of facts. If this is right, then the connection between a nudge's being transparent and a nudge's being easy to resist will be, at best, weaker than the one found in the *RTA*.

3.1. *The Case of Jacob*

We begin with a case. Suppose that Giovanni, the CEO of a company, decides that he wants his workers to pay for the more expensive, unhealthy snacks in the company cafeteria. He thereby implements a variety of nudge-style influences, including ordering the cafeteria workers to put unhealthy food at eye level, without letting any of his office employees know about the program of interventions. Now consider Jacob, who has for years avoided unhealthy foods, in large part because they work against his goal of having a finely-tuned athletic body. When he encounters the unhealthy food at eye-level, he simply searches for the healthier option, and picks that.

Jacob, it seems to us, has resisted a nudge.¹⁸ Further, we suggest, it was easy for him to resist it. Finally, his easily resisting the nudge was something he was able to do; the nudge was easily resistible for him.

Yet Jacob was not aware that the food placement might have any effect on his deliberation, decision, or behavior whatsoever.¹⁹ Nor was he aware that the unhealthy food was placed at eye-level with that purpose in mind; that is, he was not aware of the relevant effect and nudger facts. Nor need it be the case that Jacob had the capacity to easily become aware of the relevant facts. This nudge seems to be easily resistible for Jacob to resist for reasons unconnected with his awareness—actual or hypothetical—of the relevant facts.

What can we conclude from the case of Jacob? Minimally, we can conclude that an agent can resist a nudge without being aware of the relevant effect and nudger facts. Further, we suggest that a nudge can be resistible,

17. See, for instance, Barton and Grüne-Yanoff (2015: 247), Cratsley (2015), Engelen (2019a: 52; 2019b: 216), Engelen and Nys (2020: 148), Heilmann (2014), Mills (2015: 501), Moles (2015: 649), Nys and Engelen (2017: 204), Schmidt 2017: 410, n.11).

18. Although Giovanni's intervention uses the same techniques as those in *Food Placement*, some will disagree with the claim that it is a nudge. Such readers are welcome to interpret "nudge" in the following as "nudge-style influence" instead.

19. Other than, perhaps, some mundane effects, like making it such that he has to reach down, instead of straight in front of him, to pick the food he prefers.

and easily resistible, for an agent even if they are not, will not become, and do not have the capacity to easily become, aware of the relevant facts. From this, it follows that the awareness condition on easy resistibility is false. Some may, however, object.

3.2. *Objection 1 – Jacob Did Not Resist*

One might object to our argument by suggesting that Jacob did not, in fact, resist the nudge. Why might one hold this position? Perhaps because one thinks that, in order to resist something, there needs to be something to resist. In the case of a nudge, it needs to have had some sort of influence on one's mental states—say, an increased motivation to pick the unhealthy snack—in order for there to be anything to resist. Alternatively, one might think that resistance requires some sort of effort; one cannot resist something without making an effort to do something. Yet it is not clear that the nudge had any such influence on Jacob. Nor did Jacob make any sort of effort; he merely acted contrary to the nudge. Thus, one might argue, this is not an instance of resisting a nudge; in fact, it is not an instance of resisting *anything*.

Consider, then, a different case. Suppose that Liliana, who is on a diet, also works at Giovanni's company. While in line at the cafeteria, she sees the unhealthy food, and is a bit tempted by it. We can assume that the placement had an influence on the extent to which she is tempted. Recognizing the temptation, she performs a relatively easy self-control exercise. Liliana reminds herself of her plan to eat healthily, and why she has it. She thinks about how a diet intended to change her eating habits is going to be difficult, and that once she starts making exceptions, it's just going to be easier to make exceptions later on; her choice now is connected to her ability to make healthy choices later on. After this, she decides to leave the unhealthy food behind, moves to the cash register, and pays for her meal.

In this case, it is true that the nudge had an influence on Liliana's motivational states, and that she had to make some effort to refrain from grabbing the unhealthy food. We can say much the same about Liliana as we did about Jacob. It is intuitive that she resisted the nudge, and that, in resisting the nudge, she did something she was able to do. Yet, as was the case with Jacob, she was not aware of, nor need it have been easy for her to become aware of, the relevant facts. Was it easy for her to do what she did? This depends on how we are to understand ease and difficulty, but we think that it is plausible that, in this case, it was easy for her to do what she did, and so it was easy for her to resist the nudge.

But suppose that one denies this, suppose that one thinks that it was not easy for Liliana to resist the nudge in this case, though it might also not have

been hard. What can we say to this objector? We might begin by pointing out that the problem with the case of Jacob was supposed to be that it doesn't count as an instance of resistance, because there was no effect on his mental states for him to resist, or because there was no effort. Granting this point for the sake of argument, we then offered the case of Liliana. This case is similar to Jacob's in various respects, yet we stipulated that the nudge had an effect on her mental states, and we modified it to include effortful resistance. If Jacob's case does not count as an instance of resistance, in virtue of the fact that there was no effort, and Liliana's does not count as a case of *easy* resistance, in virtue of the fact that there was effort, what must a case be like in order to have easy resistance? Without an answer to this question, we find it difficult to take the objection seriously.

3.3. Objection 2 – Jacob and Liliana Resist Something, But Not the Nudge – Version I

One might push back on our claims in a different way. One might agree that what Jacob and Liliana do is (relatively) easy for them, and one might think that both are cases of resistance. However, one might argue, Liliana and Jacob do not resist the *nudge*, nor do they resist *Giovanni's pressure*. Rather, they resist the effects of the nudge, or the effects of Giovanni's pressure: the (increased) motivation to eat the unhealthy snack.

Suppose that this is right, one might then ask what more we need to have an instance of someone resisting a nudge. We consider two ways of answering this question. First, perhaps the objector thinks that, in order to resist a nudge, one must not only prevent it from having any effect on one's behavior, as one would if one acted contrary to the intention behind the nudge, one must also reverse the effect that the nudge has on one's motivational states. Although we find this implausible, one might offer some support for this objection by considering the intentions that lie behind a nudge. Nudges are intended to modify at least some individuals' behavior, but they are also intended to do this via an influence on an agent's mental states. It might be thought that, to resist a nudge, one must block or reverse *all* of its intended effects. Thus, even if Liliana chooses contrary to the nudge, she will have failed to resist it if she does not reverse (or block) its motivational effects, for example, by reducing the strength of the motivation that was increased by the nudge.²⁰ In the current case, she does not do this, she merely acts contrary to the nudge-increased motivation.

20. This might be in line with Blumenthal-Barby's suggestion that "ease of resistance is in part about the ability to inhibit whatever urge or propensity is triggered by the nudge" (2021: 120).

Suppose, then, that Liliana used a slightly different self-control technique. Rather than reminding herself of her plan to eat healthily, and the reasoning that led her to making the plan, she instead uses a different trick. She tries to imagine the unhealthy food as something disgusting—for instance, she tries to think of the chocolate pudding as mud. The strategy immediately works, and her motivation to pick the unhealthy food is reduced. Unlike the previous version of the case, in which she strengthens her motivation to act contrary to the desire for unhealthy food, in this version, she exerts effort in reducing the motivation that was increased by the nudge. Yet, in order to modify the case in this way, we need not change features about her awareness, nor her capacity to become aware, of the relevant facts; it still need not be the case that she was able to easily become aware of these facts. This version of the objection, then, does not save the awareness condition on easy resistibility.

3.4. Objection 3 – Jacob and Liliana Resist Something, But Not the Nudge – Version II²¹

There is, however, a different way of developing this objection. On this way of developing the objection, resisting the nudge, rather than its effects, requires that, when acting contrary to the nudge, one *intends* to act contrary to the nudge; that is, the content of one's intention must make reference to the fact that one's action is contrary to the nudge. But this requirement is not satisfied in the cases of Jacob and Liliana. Although these agents might intend *to pick the healthy option, to pick the salad, or to not pick the chocolate pudding*, they do not intend *to act contrary to the nudge*.²² Thus, these agents do not resist the nudge, and these cases fail as counterexamples to *P2*. Further, on this view, *P2* is straightforwardly true. Since resistance requires that the content of one's intention make reference to the fact that the action is contrary to the nudge, and in order for the intention to have this content, the agent must be aware that they are being nudged, resistance requires awareness that one is being nudged. Thus, in order for a nudge to be easily resistible, the nudgee must have the capacity to easily become aware of relevant nudge-facts.²³

We agree that a way of resisting a nudge is to do so with the intention to act contrary to the nudge, and that resisting in this way requires awareness of the

21. We are grateful to two reviewers for offering this style of objection.

22. Or, along the lines of one reviewer comment, they do not intend to “refuse to behave as the nudger wishes.”

23. Another reviewer offers a slightly different, and weaker, version of the objection, on which Jacob and Liliana resist the nudge *qua* feature of the choice environment, but they do not resist it *qua* nudge. Our response should apply to this version as well.

nudge. We also agree that, if this were the *only* way to resist a nudge, then our examples fail, and we have straightforward reasons for thinking that *P2* is true. However, we do not think that this is the only way to resist a nudge. Partly, we think this because we find it plausible that Jacob and Liliana do resist the nudge. But further, if this were the only way to resist a nudge, we would have to conclude not only that Jacob and Liliana do not resist the nudge, but also that the nudge is not *easily resistible* for them, insofar as they do not have the capacity to easily become aware of the relevant nudge facts. This, we think, is hard to believe.

Moreover, if resisting a nudge requires intending to act contrary to the nudge, this will cast doubt on the moral significance of resistibility. To see this, suppose that there are two reasonable ways of understanding “resisting a nudge.” On one understanding, resistance requires that the agent intend to act contrary to the nudge. Call this *I-Resistance*. An agent does not *I-Resist* a nudge if they do not have this intention, and Jacob and Liliana do not *I-Resist* the nudge. On a second, broader, way of understanding resistance—call it *NI-Resistance*—resistance does not have this necessary condition, and Jacob and Liliana *NI-Resist* the nudge. Now recall *P1*:

P1. A nudge is permissible only if it is easy to resist.

Recall that *P1* can come in weaker versions. On one such version, if a nudge is not easy to resist, then we have a strong, *pro tanto*, reason against implementing the nudge. Is *P1*—or one of its somewhat weaker variations—plausible when interpreted in terms of *I-Resistance*? What about with respect to *NI-Resistance*? Though we think that *P1*—or, rather, some weaker variant of it—has some plausibility with respect to *NI-Resistance*, we do not think it is plausible with respect to *I-Resistance*.

Consider again *Food Placement*, from the introduction. This was a case of a nudge similar to Giovanni’s nudge, except that the food placement is intended to get more people to pick the healthy food option over the unhealthy one, rather than the reverse. Understanding *P1*—or its somewhat weaker variations—in terms of *I-Resistance* would suggest that such an influence is very problematic if nudgees do not have the capacity to easily become aware of it. On the strong version of *P1*, it would be impermissible; on the somewhat weaker version we suggested, we have a strong *pro tanto* reason against it. And this is so even if all the following are true: the nudge has only a very small effect on a nudgee’s motivation, the nudgee picks the unhealthy food she would have picked in the absence of the nudge, and she finds it very easy to act contrary to the nudge in this way; and her acting contrary to the nudge in this way is a reflection of her values or desires, and the result of normal deliberation which is not significantly influenced by the nudge. If *P1* is understood in terms of

I-Resistance, the nudge will be impermissible—or, on weaker versions of *P1*, we will have strong reason not to implement this nudge—because the agent does not have the capacity to easily become aware of the relevant facts, and thus to form an intention to resist it. But it is hard to believe that this nudge is impermissible or that there is a strong reason against it. Thus, if easy resistibility is to be relevant to the permissibility of a nudge, we think it would be much more plausible to understand easy resistibility in terms of *NI-Resistance* than in terms of *I-Resistance*.

So, although we deny that the *only* way to resist a nudge is to do so with the intention of acting contrary to it—and thus deny that *I-Resistance* offers an adequate account of “resisting a nudge”—we also think that for the purposes of determining whether a nudge is permissible, resistibility, when understood in terms of *I-Resistance* is not relevant.

3.5. A Sufficient Condition for Nudge Resistance

The discussion above mostly concerns disagreement between us and our objectors about what it means for an agent to resist a nudge. This discussion points towards the following sufficient condition on resisting a nudge:

Sufficient for Resistance. If an agent is exposed to a nudge intended to get agents to *A*, and the agent intentionally acts or omits in a way contrary to *A*, then the agent resists the nudge.²⁴

Recall *Default Registration*. While filling out his registration for a driver’s license, Joe faces a question about being added to the organ donor list, with registration as the default. Suppose that, after thinking about it for a few seconds, he ticks the box, and opts out of being registered for organ donation. In this case, there is a nudge intended to get people to enroll as organ donors: it is the default option. Further, Joe decides otherwise, and intentionally opts out of the organ donation program, acting contrary to what the nudge is intended to get him to do. *Sufficient for Resistance* would tell us that Joe resisted the nudge, which coheres with our intuitions about this case.

We wish to highlight some features of *Sufficient for Resistance*. First, notice that SR does not make a claim about the *ease* of resistance; it simply makes a

24. If one thinks, like one version of our first objector, that in order to resist a nudge, it must have an effect on one’s motivation, one can adopt a similar condition which replaces “an agent is exposed to . . . to *A*” with “If a nudge, intended to get agents to *A*, increases an agent’s motivation to *A*.” If one thinks that resistance requires *exerting effort*, one can replace “and the agent intentionally acts or omits” with “and the agent *exerts effort in* intentionally acting or omitting.”

claim about one way to resist a nudge, regardless of whether it is easy or hard to resist.

Second, notice that *Sufficient for Resistance* is merely a sufficient condition on an agent resisting a nudge. We do not intend to provide a full account of what it is for an agent to resist a nudge; it is consistent with *Sufficient for Resistance* that there are other ways of resisting a nudge, and that there are cases in which an agent resists a nudge yet does not meet this condition.

Third, *Sufficient for Resistance* does not imply that *merely failing* to act in the way the nudge is intended to get one to act is sufficient for resisting the nudge. For example, suppose that an agent, while looking at the healthy food that is at eye-level, suffers a stroke and dies. In this case, it is true that he did not do as the nudger intended; he did not pick the healthy food. Yet, as far as *Sufficient for Resistance* is concerned, this will not be sufficient to count as an instance of resistance.

Finally, a clarification of what is meant by “the agent intentionally acts or omits in a way contrary to A.” The agent’s intentionally omitting in a way contrary to A, we think, just means that the agent intentionally does not A. However, the agent’s intentionally acting in a way contrary to A is a bit more complicated. Agents can perform multiple actions simultaneously; they can, literally and metaphorically, walk and chew gum at the same time. If a nudge is intended to get agents to A, and a nudgee B-s, where B-ing is a different action than A-ing, then this may be consistent with the agent’s A-ing as well.²⁵ Thus, we do not simply mean something like “the agent intentionally B-s, and B-ing is not a kind of A-ing.” However, we also do not mean something as strong as “the agent intentionally B-s at *t*, and the agent’s B-ing at *t* is inconsistent with her A-ing at *t*.” The agent who faces healthy food at eye-level and picks the unhealthy snack instead might resist the nudge. But her picking the unhealthy snack does not entail that she does not *also* pick the healthy food; she can pick both. If the agent intentionally acts in a way contrary to A, we suggest, then she both (i) intentionally B-ed, where B-ing is an action and is not a kind of A-ing, *and* (ii) she does not A; with the scope of “intentionally” restricted to (i).²⁶

Sufficient for Resistance can account for the claim that in the case of Jacob and in both variations of the case of Liliana, the nudgee resists the nudge.

25. It may be worth noting that in some cases, one does only have, effectively, inconsistent alternatives. For instance, Joe’s alternatives were to either check the box or not check the box; he cannot realistically do both.

26. One might think that this is much ado about nothing; if an agent intentionally B-s while knowing that they are not A-ing, and perhaps, while having known that in deciding to B, this will have the consequence that she does not A, then she intentionally omits to A. If this is true, then this would all be much simpler. However, we are not convinced that this is true.

4. A Weaker P₂?

What is the upshot of the discussion in the previous section? Here, we offer some conclusions. Throughout, we limit ourselves to considering the resistibility, awareness, transparency and permissibility of individual nudges targeted at individual nudgees. Implementing nudge policies for a whole population raises additional issues that we cannot consider here.

Recall *RTA*:

- P*₁. A nudge is permissible only if it is easy to resist.
- P*₂. A nudge is easy to resist only if the agent can easily become aware of it.
- P*₃. An agent is capable of easily becoming aware of a nudge only if it is transparent.
- C*. A nudge is permissible only if it is transparent.

In Section 2, we argued that *P*₃ is plausible with respect to hypothetical awareness transparency—on which the nudgee has the capacity to easily become aware of the relevant facts—and that a weaker version might be plausible with respect to some forms of actual disclosure transparency—on which nudgers disclose their nudging. However, it is not plausible with respect to actual awareness transparency—on which the nudgee is aware of the relevant facts—nor with hypothetical disclosure transparency—on which the nudger is merely willing and/or able to defend the nudge policy. In Section 3, we argued that *P*₂ is false. This suggests that, even transparency of the sort that makes *P*₃ plausible is not required for a nudge to be easy to resist. Thus, *RTA* fails.

Yet one might think that a weaker relationship still holds. For example, one might think that, if a nudge exhibits hypothetical awareness transparency or actual disclosure transparency then it will *tend* to be easier to resist, and thus, insofar as one has reasons to make nudges easy to resist, one will often have reason to nudge transparently.²⁷

This is a weaker view than the ones we have discussed, and involves a different sort of claim. This weaker view is, we think, more difficult to assess, in part because it depends on as-yet-unsettled empirical matters. However, we do not think that this weaker view is obviously true; it is not clearly the case that the capacity to easily become aware of nudge facts tends to make a nudge easier to resist. Here, we offer some brief comments on why we think this.

²⁷. For an expression of this weaker version, see, for example, Ivanković and Engelen (2019: 45).

First, consider that, at least according to some, the efficacy of a nudge either determines, or is evidence of, its resistibility: the more effective the nudge, the harder it is to resist, or the more reason we have to believe that it is hard to resist.²⁸ If one adopts this line, then it is worth pointing out, as others have,²⁹ that current, albeit limited, research has often found that neither nudge transparency nor nudges' awareness of effect and nudger facts, affects a nudge's efficacy.³⁰ Although we think these findings are not conclusive and can, at best, be generalized only to some types of nudges, they are the only and best empirical evidence we have available. Thus, as things stand, we have some reason to hesitate before accepting the claim that transparency regarding and/or awareness of the nudge can make the nudge easier to resist.

Second, it is quite possible to imagine realistic cases in which the relevant kinds of transparency in fact work against resistibility. Suppose, for instance, that, in order to increase pension savings by its employees, an employer enrolls employees in a pension scheme by default; employees have to actively opt-out to avoid inclusion in the scheme. And suppose that the employer informs employees that both 1) this policy tends to increase pension savings, and 2) it adopted the policy with this purpose in mind. If this transparency makes employees more likely to explicitly deliberate about whether to opt out, or to deliberate more thoroughly, *and* it turns out that the pension scheme lines up well with their interests, then transparency about the nudge might make it more difficult for nudges to act contrary to it. Or suppose that, rather than stimulating employees to deliberate, transparency simply provides one with further evidence on which to deliberate. The nudgee might think that those who chose this plan as the default are in a better position to make a judgment about which plan is better for them, and that the fact that it is the default is evidence of that judgment.³¹ That is, they might see the employer's disclosure as a credible recommendation. This, too, might make it harder to opt out. In both of these sorts of cases, it seems, transparency makes it harder to act contrary to the nudge, and thus to resist it, at least in the way suggested by *Sufficient for Resistance*.

As with the appeal to empirical studies, we think that these cases give us at best a weak reason to *reject* the claim that nudge transparency (of the right

28. We are not convinced that this relationship holds. At the very least, we think the relationship between a nudge's efficacy and whether it is easy to resist is more nuanced. For examples of theorists suggesting such a relationship, see Engelen and Nys (2020: 148), Houk (2019: 414), Moles (2015: 666), Rebonato (2012: 203), Schmidt and Engelen (2020: 5).

29. See, for example, De Ridder et al. (2020: 6), Ivanković and Engelen (2019: 47), Lades and Delaney (2022: 80, n. 7), Schmidt (2017: 409), Thaler and Sunstein (2021: 324).

30. See, for example, Bruns et al. (2018), Kroese et al. (2016), Loewenstein et al. (2015), Paunov et al. (2019a; 2019b; 2020), Steffel et al. (2016).

31. For related discussion, see Schmidt (2017: 409).

sort) is conducive to easy resistibility. However, we do think they give us some reason to hesitate before *accepting* the claim that it does. In any case, this claim is significantly different than the one that we mainly focused on, and we suspect that whether transparency (of the right sort) helps to make a particular nudge easier to resist will need to be determined on a case-by-case basis, with respect to particular nudge interventions.

5. Conclusion

Our main conclusion in this paper is that, even if one thinks that, in order to be permissible, a nudge needs to be easy to resist, this does not give us much, if any, reason for thinking that it also requires transparency. This is because, as we have argued, a common assumption regarding the easy resistibility of a nudge is false: one does not need the capacity to easily become aware of a nudge in order to be able to easily resist it.

We wish to stress, again, that we do not mean to suggest that there is *no* reason for being transparent when nudging; we have been interested here only in *resistibility-based* reasons for transparency.

We conclude with two brief points.

First, since autonomy of action is plausibly related to agents' abilities to act in certain ways, or on the basis of certain motives, we suspect that some of our discussion will extend to whether agents' autonomy is compromised when they are nudged in a non-transparent fashion. That is, we suspect our arguments will also create a challenge for the view that non-transparent nudges invariably diminish the nudgee's autonomy.

Second, we think that discussions of nudging may often help themselves to assumptions concerning human agency that are not realistic. Although we think that many nudges and nudge-style interventions are easy to resist for nudgees, we do not think that nudgees will often have the capacity to easily become aware of effect or nudger facts. The use of defaults, framing, the placement of food, etc. are common fare in discussions of nudging, but it is not clear to us how much of the population is generally aware of these sorts of interventions, or their potential effects. Without this knowledge, we suspect, it will not be easy to become aware of effect and/or nudger facts. Further, even in those cases where a nudgee can easily become aware of effect facts, it may not be so easy to determine whether there is a corresponding nudger fact, insofar as it may not be easy to determine whether there is, in fact, a nudger behind the feature of the environment. If we are right on these points, then it follows that many nudgees do not have the capacity to easily become aware of the relevant facts, even though they may be able to easily resist nudges.

To put it bluntly, we suspect that many nudges are like Jacob and Liliana—both in terms of their capacity to easily become aware of the relevant facts, and of their ability to easily resist the nudge. Adopting the awareness condition on easy resistibility would thus suggest that for many nudges, and for many nudges, the nudge is not easy to resist. Requiring easy resistibility for permissibility would thus mistakenly create a serious challenge to the permissibility of many nudges.

Acknowledgments

We would like to thank members of a writing group on arational influences, who commented on a previous version of this paper, an audience at the Society for Applied Philosophy Conference 2022, and an audience at the 16th World Congress of Bioethics. For discussion, we would like to thank Viktor Ivanković, and for very helpful written comments, we would like to thank Maximilian Kiener and Peter Schaber, and two anonymous reviewers for this journal.

Funding

We would like to thank the European Research Council for their funding [Consolidator Award 819757]. Thomas Douglas would like to thank, for their funding, the Uehiro Foundation on Ethics and Education.

References

- Anderson, J. (2010). *Nudge: Improving Decisions about Health, Wealth, and Happiness*, Richard H. Thaler and Cass R. Sunstein. Yale University Press, 2008. x + 293 pages. [Paperback edition, Penguin, 2009, 320 pages.]. *Economics & Philosophy*, 26(3), 369–76.
- Barton, A. and T. Grüne-Yanoff (2015). From Libertarian Paternalism to Nudging—and Beyond. *Review of Philosophy and Psychology*, 6(3), 341–59.
- Blumenthal-Barby, J. S. (2021). *Good Ethics and Bad Choices: The Relevance of Behavioral Economics for Medical Ethics*. MIT Press.
- Bovens, L. (2009). The Ethics of Nudge. In *Preference Change* (207–19). Springer.
- Bruns, H., E. Kantorowicz-Reznichenko, K. Klement, M. Luistro Jonsson, and B. Rahali (2018). Can Nudges Be Transparent and Yet Effective? *Journal of Economic Psychology*, 65, 41–59.
- Capron, A. M. (2015). Are There Public Health Domains in “Domain-Specific” Health Nudging? *The American Journal of Bioethics*, 15(10), 47–49.
- Clavien, C. (2018). Ethics of Nudges: A General Framework with a Focus on Shared Preference Justifications. *Journal of Moral Education*, 47(3), 366–82.

- Cratsley, K. (2015). Nudges and Coercion: Conceptual, Empirical, and Normative Considerations. *Monash Bioethics Review*, 33(2), 210–18.
- De Ridder, D., J. Feitsma, M. V. D. Hoven, F. Kroese, T. Schillemans, M. Verweij, T. Venema, A. Vugts, and E. De Vet. (2020). Simple Nudges That Are Not So Easy. *Behavioural Public Policy*. Advance online publication. <https://doi.org/10.1017/bpp.2020.36>
- Engelen, B. (2019a). Ethical Criteria for Health-Promoting Nudges: A Case-by-Case Analysis. *The American Journal of Bioethics*, 19(5), 48–59.
- Engelen, B. (2019b). Nudging and Rationality: What Is There to Worry? *Rationality and Society*, 31(2), 204–32.
- Engelen, B., and T. Nys (2020). Nudging and Autonomy: Analyzing and Alleviating the Worries. *Review of Philosophy and Psychology*, 11(1), 137–56.
- Gatignon, H., and E. Le Nagard (2015). *Manipulating Consumers is Not Marketing: A Commentary on Cass R. Sunstein's "Fifty Shades of Manipulation"* (SSRN Scholarly Paper No. ID 2686547). Social Science Research Network.
- Grüne-Yanoff, T. (2012). Old Wine in New Casks: Libertarian Paternalism Still Violates Liberal Principles. *Social Choice and Welfare*, 38(4), 635–45.
- Hansen, P. G. and A. M. Jespersen (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. and B. Welch (2010). Debate: To Nudge or Not to Nudge*. *Journal of Political Philosophy*, 18(1), 123–36.
- Heilmann, C. (2014). Success Conditions for Nudges: A Methodological Critique of Libertarian Paternalism. *European Journal for Philosophy of Science*, 4(1), 75–94.
- Houk, T. (2019). On Nudging's Supposed Threat to Rational Decision-Making. *The Journal of Medicine and Philosophy: A Forum for Bioethics and Philosophy of Medicine*, 44(4), 403–22.
- Ivanković, V. and B. Engelen (2019). Nudging, Transparency, and Watchfulness. *Social Theory and Practice*, 45(1), 43–73.
- Kiener, M. (2021). When Do Nudges Undermine Voluntary Consent? *Philosophical Studies*, 178(12), 4201–26.
- Kipper, J. (2020). Irresistible Nudges, Inevitable Nudges, and the Freedom to Choose. *Moral Philosophy and Politics*. Advance online publication. <https://doi.org/10.1515/mopp-2020-0013>
- Kroese, F. M., D. R. Marchiori, and D. T. D. de Ridder (2016). Nudging Healthy Food Choices: A Field Experiment at the Train Station. *Journal of Public Health*, 38(2), e133–e137.
- Lades, L. K., and L. Delaney (2022). Nudge FORGOOD. *Behavioural Public Policy*, 6(1), 75–94.
- Ledderer, L., M. Kjær, and E. K. Madsen (2020). Nudging in Public Health Lifestyle Interventions: A Systematic Literature Review and Metasynthesis. *Health Education and Behavior*, 47(5), 749–64.
- Loewenstein, G., C. Bryce, D. Hagmann, and S. Rajpal (2015). Warning: You Are About to Be Nudged. *Behavioral Science & Policy*, 1(1), 35–42.
- Majtenyi, K. and M. Ruble (2019). Nudges and Budes. *The American Journal of Bioethics*, 19(5), 72–74.
- Mills, C. (2015). The Heteronomy of Choice Architecture. *Review of Philosophy and Psychology*, 6(3), 495–509.

- Moles, A. (2015). Nudging for Liberals. *Social Theory and Practice*, 41(4), 644–67.
- Nys, T. R. and B. Engelen (2017). Judging Nudging: Answering the Manipulation Objection. *Political Studies*, 65(1), 199–214.
- Paunov, Y., M. Wänke, and T. Vogel (2019a). Ethical Defaults: Which Transparency Components Can Increase the Effectiveness of Default Nudges? *Social Influence*, 14(3–4), 104–16.
- Paunov, Y., M. Wänke, and T. Vogel (2019b). Transparency Effects on Policy Compliance: Disclosing How Defaults Work Can Enhance Their Effectiveness. *Behavioural Public Policy*, 3(2), 187–208.
- Paunov, Y., M. Wänke, and T. Vogel (2020). Combining Defaults and Transparency Information to Increase Policy Compliance. *Social Psychology*, 51(5), 354–59.
- Rebonato, Riccardo (2012). *Taking Liberties: A Critical Examination of Libertarian Paternalism*. Palgrave Macmillan.
- Rebonato, Riccardo (2014). A Critical Assessment of Libertarian Paternalism. *Journal of Consumer Policy*, 37(3), 357–96.
- Saghai, Y. (2013a). Salvaging the Concept of Nudge. *Journal of Medical Ethics*, 39(8), 487–93.
- Saghai, Y. (2013b). The Concept of Nudge and Its Moral Significance: A Reply to Ashcroft, Bovens, Dworkin, Welch and Wertheimer. *Journal of Medical Ethics*, 39(8), 499–501.
- Schmidt, A. T. (2017). The Power to Nudge. *American Political Science Review*, 111(2), 404–17.
- Schmidt, A. T. and B. Engelen (2020). The Ethics of Nudging: An Overview. *Philosophy Compass*, 15(4), e12658.
- Soled, D., N. W. Dickert, and J. Blumenthal-Barby (2021). When Does Nudging Represent Fraudulent Disclosure? *The American Journal of Bioethics*, 21(5), 63–66.
- Steffel, M., E. F. Williams, and R. Pogacar (2016). Ethically Deployed Defaults: Transparency and Consumer Protection through Disclosure and Preference Articulation. *Journal of Marketing Research*, 53(5), 865–80.
- Sunstein, C. R. (2014). Nudging: A Very Short Guide. *Journal of Consumer Policy*, 37(4), 583–88.
- Sunstein, C. R. (2015a). The Ethics of Nudging. *Yale Journal on Regulation*, 32(2), 413–50.
- Sunstein, C. R. (2015b). *Why Nudge?: The Politics of Libertarian Paternalism* (reprint ed.). Yale University Press.
- Sunstein, C. R. (2016). Fifty Shades of Manipulation. *Journal of Marketing Behavior*, 1(3–4), 213–44.
- Thaler, R. H. and C. R. Sunstein (2009). *Nudge: Improving Decisions About Health, Wealth and Happiness* (1st ed.). Penguin.
- Thaler, R. H. and C. R. Sunstein (2021). *Nudge: The Final Edition*. Allen Lane.
- Wilkinson, T. M. (2013). Nudging and Manipulation. *Political Studies*, 61(2), 341–55.