PANDION: The Osprey Journal of Research & Ideas

Volume 1 | Number 2 Article 5

2020

The Ethics and Applications of Nudges

Valerie D. Joly Chock
University of North Florida, n01051115@unf.edu

Faculty Mentor: Jon Matheson, PhD, Associate Professor
Department of Philosophy and Religious Studies
Philosophy Honors Thesis

The Ethics and Applications of Nudges

Cover Page Footnote

Thanks to my mentor Dr. Jon Matheson for providing helpful feedback, guidance, and support throughout the development of this paper.

The Ethics and Applications of Nudges

Valerie Joly Chock

Faculty Mentor: Jonathan Matheson, PhD
Department of Philosophy and Religious Studies
University of North Florida

Abstract
Nudging is the idea that people's decisions should be influenced in predictable, non-coercive ways by making changes to the way that options are presented to them. Central to the debate about nudging is the question of whether it is morally permissible to intentionally nudge others. Libertarian paternalists maintain that this can be the case. In this paper, I present the libertarian paternalistic criteria for the moral permissibility of intentional nudges. Having done this, I motivate two objections. The first one targets the moral permissibility of nudging in general. The second one targets the moral permissibility of only a subset of nudges. After evaluating both of these objections, I conclude that although they are unsuccessful, the evaluation of The Manipulation Objection shows that the libertarian paternalistic criteria for the moral permissibility of intentional nudges fails. I end by suggesting a possible revision to the criteria that avoids the problem, considering its limitations, and providing examples of its applications.

Introduction

We are not as rational as we think. While we like to think that our decisions are the result of rational reflection, the truth is that things we are unaware of actually influence our decisions all the time. Advances in cognitive and behavioral science reveal that the way that options are presented, what is referred to as the 'choice architecture', strongly influences our decisions. We tend to react to a particular option differently depending on how it is presented to us. Based on this data, Richard H. Thaler and Cass R. Sunstein came up with *nudging*, the idea that people's decisions and behaviors should be influenced in predictable, non-coercive ways by making small changes to the choice architecture. Central to the debate about nudging is the question of whether it is morally permissible to intentionally nudge other people. Libertarian paternalists maintain that this can be the case and offer a criteria for the moral permissibility of intentional interventions.

I start this project by presenting clear and coherent accounts of Nudging and Libertarian Paternalism that are faithful to how these concepts were originally introduced

by Thaler and Sunstein. My purpose in doing this is to dispel two misconceptions often found in the literature: (i) that nudges are intentional and avoidable, and (ii) that all instances of nudging are interventions grounded by Libertarian Paternalism. By clearing up the confusion surrounding these misconceptions, my project is able to move away from the question of whether it can be morally permissible to nudge, to more productive discussion about *how* it is permissible to do so. To answer this question, I present the Libertarian Paternalistic Criteria (LPC) for the moral permissibility of intentional nudges and defend it against two objections: Biased Choice Architects and The Manipulation Objection. The former objects to the moral permissibility of intentional nudging in general, whereas the latter objects only to a particular subset of nudges. While I show that both objections fail, I argue that the evaluation of The Manipulation Objection raises a different challenge for LPC. The Libertarian Paternalistic Criteria for the moral permissibility of intentional nudges fails because it ignores the moral distinction that exists between different kinds of nudges. Such distinction is not the one proposed by the objection, but one between what I call 'counteractive' and 'non-counteractive' nudges. I end by suggesting a possible revision to the criteria that avoids the problem, considering its limitations, and providing examples of its applications.

1. NUDGING

In their book *Nudge: Improving Decisions About Health, Wealth, and Happiness*, Richard Thaler and Cass Sunstein define a 'nudge' as "any aspect of the choice architecture that alters people's behavior in a predictable way without forbidding any options or significantly changing their economic incentives."[1] In other words, a nudge is a change in the way options are presented that steers people in a certain direction without forbidding any options, all while maintaining people's freedom of choice. Suggestions, warnings, defaults, and recommendations are all examples of common nudges. On the other hand, fines, mandates, threats, bans, and direct instruction are not, since they either forbid certain options or significantly reduce one's freedom of choice. A nudge is

1 Thaler and Sunstein (2009, 6).
Note that Thaler and Sunstein give the original definition of 'nudge' in terms of economic incentives. However, they accept that there are other kinds of incentives as well. Thus, a charitable interpretation must be inclusive of these other kinds. Note also that the definition given here treats 'nudge' as a noun. However, Thaler and Sunstein—as well as most authors who engage with the subject—also use the term as a verb to refer to the action of influencing people by means of nudges.

analogous to how a GPS works. The GPS suggests the best route, but it does not coerce drivers to take that route. The GPS allows drivers to go in a different direction if they so choose, but drivers often end up taking the suggested route. Likewise, a nudge influences people to make certain choices, but it does not coerce them to do so. It is possible for both the driver and the person being nudged to easily choose another available option, even though it is often likely that they will not. Nudging is, then, the practice of influencing people's decisions and behaviors in non-coercive ways by making small changes to the choice architecture.

A choice architect is a person responsible for presenting options. Salespeople, doctors, waiters, website designers, policy makers, and professors are only a few examples of choice architects who nudge. Their decisions about how options are presented to other people affect the decisions that other people make. This is supported by evidence from cognitive and behavioral science which suggests that we commonly make irrational decisions due to systematic errors in how we reason.[2] These errors typically result from heuristics, which are rules-of-thumb that people use to form judgements and make decisions. Heuristics, although sometimes accurate and useful, often lead to "severe and systematic errors,"[3] called 'cognitive biases', that result when people choose what the heuristics suggest even when the suggestion is a poor option. Thus, even when we make "good" decisions (outcome-wise), those decisions are still irrational (in some sense) when they are the result of biased decision-making.

The three main heuristics and cognitive biases identified by Tversky and Kahneman in their original research are the following:[4]

a) Representativeness

People tend to evaluate probabilities by the degree to which A is representative of B. That is, by how similar A is to B.[5] This heuristic takes place in stereotyping. For example, we think a 6'8" man is more likely to be a professional basketball player than

2 This research was pioneered by the 'heuristics and biases' work of psychologists Daniel Kahneman and Amos Tversky, which is rooted in dual-process theories. See Tversky and Kahneman (1974).
3 Tversky and Kahneman (1974, 1124).
4 Tversky and Kahneman (1974).
5 Tversky and Kahneman (1974, 1124).

a 5'6" man because most professional basketball players are taller than 5'6".[6] In this case, the heuristic results in an accurate judgment, but biases occur when similarity and frequency diverge. This is demonstrated in a thought experiment about a woman named Linda.[7] In this thought experiment, subjects are told that Linda is a young, single, outspoken, and bright woman who majored in philosophy. Subjects are then told that Linda is concerned with issues of discrimination and that she constantly participates in social justice demonstrations. After these descriptions are given to them, subjects are asked to choose whether they think Linda is: (i) a bank teller, or (ii) a bank teller active in the feminist movement. The majority of people tend to choose (ii) even though this is a logical mistake. It is not logically possible for a conjunction of events to be more likely than one of those events alone. In this case, it is more likely that Linda is a bank teller than a feminist bank teller because *all* feminist bank tellers are bank tellers. The cognitive bias that results from the representativeness heuristic is rooted in the fact that people rely on similarities and expectations to choose the description that seems to better match Linda according to the description they are given of her. Instinctively, Linda's description matches (ii) better than (i).[8]

b) Availability

People tend to assess the frequency or probability of an event according to the ease with which instances of such event are recalled.[9] The availability heuristic is closely related to biases of accessibility, salience, and recency. The media greatly influences what information is available to us. For example, people tend to believe that homicides are more common than suicides given that information about the former is more accessible (i.e., more frequently reported in the media) and thus, more easily recalled. However, more people die from suicide than they do from homicide.[10] Personal experiences and recent events are also widely available to us, and more easily recalled than events one

6 Thaler and Sunstein (2009, 26).
7 Thaler and Sunstein (2009, 26).
8 The representative heuristic also leads people to confuse random fluctuations with causal patterns. Examples of this are when people detect cases of repetition in the outcome of random processes and believe them to be meaningful patterns when they are actually just due to chance—e.g., sequence of coin tosses, the "hot hand" fallacy, etc.
9 Tversky and Kahneman (1974, 1127).
10 Thaler and Sunstein (2009, 25).

has not experienced personally or events one experienced a long time ago.[11] The effect of the bias that results from the availability heuristic is that easily recalled information has the potential to misinform people's judgements. And these judgements themselves could lead to irrational actions. One example here has to do with flood insurance purchases. Someone who has not experienced a flood in the immediate past is far less likely to purchase insurance (even if they live on floodplains) than someone who has experienced (or know someone who has experienced) one. This is the case even if the first person lives on a floodplain and the second lives in an area where there is no significant flood risk.[12]

c) Adjustment and Anchoring

In order to make decisions or determine an unknown, people tend to rely heavily on initial or familiar values as starting points, referred to as *anchors*, to make estimates by means of adjustment.[13] The bias that results from this heuristic occurs when the comparable reference is not a good similarity. Thaler and Sunstein give an example in which two groups of people are asked to guess the population of Milwaukee. One group is from Chicago and the other is from Green Bay. The first group does not know much about Milwaukee. They think it is the biggest city in Wisconsin, but they do not think it is as big as Chicago, which leads them to say that the population of Milwaukee is about one-third the size of that of Chicago (approx. one million). The group from Green Bay does not know the exact answer either, but they know that Green Bay has about one hundred thousand people. They also know that Milwaukee is larger than Green Bay, for which they guess three hundred thousand. In this case, people from Chicago make a guess too high whereas those from Green Bay guess too low. This shows that different people arbitrarily choose different anchors for the same unknown questions, and these anchors often produce very different answers.[14]

Kahneman and Tversky's research on heuristics and biases is rooted on psychological dual-process theories. Kahneman (2011) describes two thinking systems,

---

11 Thaler and Sunstein (2009, 25).

12 Thaler and Sunstein (2009, 25).

13 Tversky and Kahneman (1974, 1128).

14 Some other examples of cognitive biases include the tendency to favor pre-selected options (default effect), to avoid losses over making gains (loss aversion), to do things merely because other people do them (bandwagon effect), and many others.
For empirical evidence, and in-depth discussion on heuristics and cognitive biases, see Tversky and Kahneman (1974), Kahneman (2011), Ariely (2010), and Thaler and Sunstein (2009).

System 1 and System 2, which Thaler and Sunstein refer to as 'the Automatic System' and 'the Reflective System' respectively. System 2 thinking is deliberate, controlled, effortful, and conscious. System 1 thinking is rapid, uncontrolled, effortless, and subconscious.[15] Based on this distinction, it is possible to differentiate between two types of nudges: type-1 nudges and type-2 nudges. The latter engages individuals reflectively—they engage System 2 thinking—whereas the former does not—they engage only System 1 thinking.[16] The plate-size nudge is an example of a type-1 nudge.[17] This nudge consists of reducing the plates' size in cafeterias, which typically reduces consumers' food intake, and typically leads to a calorie intake reduction as well.[18] This usually happens without engaging the subject's reflective thinking because of perceptual biases that result in *mindless eating*.[19] The typical behavior is to first fill the plate with food and then eat the food typically without reflecting about it.[20]

An example of a type-2 nudge with the same goal—reducing calorie intake—is placing nutritional labels besides the food along with placards encouraging the consumption of low-calorie food. Here, the nudge works by engaging diners' reflective thinking. The placards engage subjects when they read the information on the label and proceed to reflect on their food choices (which typically results in diners consciously choosing lower-calorie foods or reducing the amount of food they put on their plate). These examples illustrate the difference between type-1 and type-2 nudges. That is, type-1 nudges engage people subconsciously—they "work in the dark"[21]—whereas type-2 nudges engage people reflectively.

---

15 Thaler and Sunstein write that "one way to think about this is that the Automatic System is your gut reaction and the Reflective System is your conscious thought" (2009, 21).

16 This distinction is similar to the ones made by Hansen and Jespersen (2013) and Sunstein (2015 & 2016). However, there are some important differences. Although Hansen and Jespersen's distinction, as well as Sunstein's, are also inspired by Kahneman's thinking systems, they add other components that are not relevant for the distinction I am concerned with.
Hansen and Jespersen add a transparency component that further categorizes nudges. Sunstein (who uses the language of 'System 1 nudges' and 'System 2 nudges'), on the other hand, adds an education component. None of these additional components are necessary to the distinction that is relevant for my argument. I am concerned merely with the thinking systems that are engaged in each type of nudge.

17 See Hansen and Jespersen (2013, 15).

18 For empirical evidence, see Wansink (2004).

19 The term 'mindless eating' was coined by Brian Wansink. It refers to consuming food without paying close attention to what and how much is being eaten. See Wansink et al. (2009).

20 Hansen and Jespersen claim that this is a type-1 nudge because "there is usually no conscious decision or choice made in this sequence of behavior in regard to how much to eat" (2013, 15).

21 Bovens (2008, 3).

One confusion in the literature revolves around the intentions of choice architects who nudge. Some authors argue that nudging is intentional by definition.[22] However, the definition of nudge given by Thaler and Sunstein does not include intention as a necessary component for nudging. According to Thaler and Sunstein, for nudging to take place, there need not be intentions involved. This is closely related to their claim that choice architecture and nudging are unavoidable. Thaler and Sunstein claim that "supermarkets, restaurants, and cafeterias are nudging consumers all the time whether they recognize this or not."[23] This implies that, besides intentions, awareness is not necessary either. It follows, then, that when choice architects nudge, they do not necessarily do so with this purpose in mind.[24] Nudging can happen without the choice architect's intentions or awareness. In both the type-1 nudge and type-2 nudge examples mentioned above, cafeterias nudge their customers via arranging the food or placing placards and nutritional labels. They are nudging their customers whether choice architects are aware of it and whether they want to or not.

1.1 Moral Permissibility

Although nudges have been used to positively influence behavior in various settings,[25] they have also been criticized. A great part of the literature on nudging focuses on its moral permissibility, with some views claiming that nudging as a whole is not morally permissible. These views present an initial worry to my project because, if nudging in general is not morally permissible, then analyzing and developing a criteria for the ethical use of nudges becomes pointless. This worry, however, can be overcome if the views against the moral permissibility of nudging in general are proven flawed.

From the outset, views that nudging as a whole is not morally permissible, face a

22 See Hansen and Jensen (2013) and Hansen (2016).

23 Thaler and Sunstein (2009, 3).

24 Hansen (2016) compares this to bragging. He claims that one need not be aware of the concept of 'bragging' to actually do this, nor they need the intend to 'brag' in individual instances in order to do so. In this sense, it is the same with nudging.

25 See Shafir (2013) for how nudges have been applied to public policy. See Thaler and Sunstein (2009) for specific examples on the increase of retirement savings (105–19), organ donations (177–84), healthy food consumption (262–3), and recycling (267–8). As well as on the reduction of environmental pollution (185–98), energy consumption (258–61), speeding (261–2), and urine spillage in public restrooms (268), among other things.

difficulty. That is, if we ought not to nudge, then not nudging must be possible.[26] However, it is impossible to not nudge. Thus, the initial response to these views should be obvious: nudging is inevitable, for which it is not reasonable to morally reject nudging as a whole. It is simply not reasonable to claim that one ought not do that which is unavoidable. This fact itself is enough to show that the views that find nudging in general morally problematic are flawed. Those who hold these views seem to be misguided by the idea that nudging is avoidable. But, as pointed out in the previous section, this is merely a misconception. Moving away from this misconception helps us realize that there is no point in critiquing the moral permissibility of nudging as a whole, however, it is possible to critique the moral permissibility of *intentional* nudges.

1.2 Bad Intentions

The Bad Intentions view claims that nudging in general is not morally permissible because choice architects can nudge "for bad."[27] This view is particularly concerned with situations in which choice architects pursue objectives that are not aligned with people's interests. Glaeser (2006) claims that nudging is not benign because it can make things worse if abused by "a less than perfect government."[28, 29] Part of Glaeser's worry is that nudges can be abused given that political leaders have a number of goals, and only some of these goals are aligned with individuals' well-being.[30] Thus, according to Glaeser, nudging in general is not permissible because choice architects may nudge for their own interests rather than those of the people they nudge. Nudging is simply a power too great to be wielded, so it should not be wielded at all.

Consider the two cafeteria nudge examples given above. In this context, Glaeser's argument says that nudging is not permissible because it is possible for the choice

---

26 This follows from Kant's "Ought Implies Can" principle, which states that the actions an individual ought to do must be possible for that individual to fulfill. This principle limits an individual's moral obligations relative to their abilities.

27 Thaler (2015).

28 Glaeser (2006, 135).

29 Glaeser focuses his argument on government and policy making. He takes the choice architect to be the government, but his claims and worries can be generalized and proposed in terms of other kinds of choice architects as well.

30 Glaeser (2006, 155).

architects to nudge in directions that are not aligned with the well-being of the diners. Glaeser's argument is one in which the minimum possibility for deviation from the welfare of the person being nudged is sufficient reason to reject nudging as a whole. But this is not the case. Accepting that nudging is not permissible because there is a *possibility* for the person being nudged to end up worse off rather than better off is like accepting that people should not speak because there is a possibility that they could say mean or offensive things to the listener rather than flattering comments. It is unreasonable to reject speaking under these conditions, even though it may be reasonable to reject the use of offensive terms in certain circumstances. In that same sense, it is unreasonable to reject nudging entirely, while it is reasonable to reject only particular instances of nudging.

A stronger case for the Bad Intentions view is one in which the nudge is ill-intentioned. For example, in cafeterias, choice architects could arrange the food with the goal of affecting diners' health in a negative way by nudging them towards the unhealthy food. This would be a case in which the choice architect *intentionally* does wrong. Here, it must be acknowledged that choice architecture can be altered in ways that allow the choice architect to introduce nudges for malicious reasons or with malicious goals. Cases where there are underlying malicious motivations on the part of the choice architect are the ones that make the Bad Intentions worry stronger. However, this is not sufficient to effectively reject nudging in general either. Malicious intent is a problem that should be objected to independently from nudging. If a person nudges with malicious intent, then that particular nudge is objectionable, but it is not the case that nudging as a whole should be rejected. It is simply not reasonable to reject nudging as a whole just because of the particular instances in which nudging is or could be ill-intentioned. Nudges can be independently impermissible, but from this, it does not follow that nudging in general is not permissible.

The Bad Intentions view fails to show that nudging as a whole is not permissible given that it is not effective in rejecting nudging in general but only the particular instances of nudging that are ill-intentioned. Further, these particular instances are objectionable not because they are nudges, but because they are ill-intentioned. Resisting the Bad Intentions view helps us move away from the question of whether it can be permissible to nudge, to more productive discussions about how it is permissible to do so. Thaler and Sunstein acknowledge that choice architects can

nudge for bad. For this reason, they argue in favor of "nudging for good."[31] That is, in favor of what they call 'Libertarian Paternalism,' a theory based on the use of nudges as tools for behavioral change aimed at improving people's welfare.

2. LIBERTARIAN PATERNALISM

Libertarian Paternalism concerns how options *should* be presented. It is motivated, in part, by the unavoidability of choice architecture. When options are presented, they simply *must* be presented in *some* way or other. Further, *any* way of presenting the options inevitably influences what people choose. According to libertarian paternalists, this, coupled with the fact that people often make irrational decisions, can make intentional nudges morally permissible. Since nudging will inevitably occur, choice architects should nudge well. However, an intentional intervention is not morally permissible *merely* in virtue of being a nudge. Libertarian paternalists propose a criteria for the moral permissibility of intentional interventions (hereafter 'LPC'), which gives a set of *sufficient* conditions for the moral permissibility of an intentional intervention. According to LPC, any intervention that is (i) a nudge, (ii) transparent, and (iii) aimed towards the welfare of those being nudged is morally permissible.[32] Here, (i) requires that the intervention must preserve freedom of choice—it cannot forbid or attach incentives or consequences to options[33]; (ii) requires that the influence must be easy to resist—people must be able to choose a different available option[34]; and (iii) requires that the intervention must be justifiably intended to make people better off—it must have people's best interest in mind. We can thus summarize LPC as the claim that intentional interventions are morally permissible if they preserve freedom of choice by means of transparent and welfare-aimed nudges.[35]

---

31 Thaler (2015) claims that whenever he is asked to autograph a copy of the book he co-authored with Sunstein, he signs is with the plea: "Nudge for Good."

32 This follows from Thaler and Sunstein (2009).

33 This means that it should alter the choice *context*, not the choice *content*. That is, modify *how* options are presented, not *what* options are presented.

34 Thaler and Sunstein write that the nudge must be "easy and cheap to avoid" (2009, 19).

35 Thaler and Sunstein recognize that "Libertarian Paternalism" seems to be a contradiction in terms. However, they "unsettle the conventional wisdom" by arguing that this term is a conceptual possibility. See Thaler and Sunstein (2003).

To see LPC applied to a case, consider the following example:

Cafeteria
Jess is a cafeteria manager who decides the way in which food is displayed. After reading some psychology studies, she learns that the food order at a cafeteria line substantially determines what ends on diners' plates—the majority of people tend to select the food that is placed first in line.[36] Aware of this, Jess now knows that whatever she decides regarding food placement will influence her diners' choices. So, she asks herself: "How *should* I arrange the food?"[37]

Jess could arrange the food: (a) So that diners are better off;[38] (b) at random; (c) to maximize her profits; (d) by banning all unhealthy foods. No matter which option she chooses, her decision will inevitably influence the diners' decisions. The fact that food must be arranged in *some way* is unescapable. Option (d) limits the available options, so it fails to meet (i) of LPC. Thus, in this case, option (d) is not a permissible intervention under LPC.[39] The remaining options (a–c), all meet condition (ii) of LPC because they do not make it significantly more difficult for diners to choose any of the other available food. However, not all (a–c) satisfy (iii). Option (b) fails to satisfy (iii) because to arrange the food at random is to consciously ignore the diners' best interests. Option (c) fails to satisfy (iii) because it either reflects the architect's selfish purposes or fails to even consider the diners' welfare. Therefore, from Jess' set of options, (a) is the only option justified by LPC—and, thus, (a) is morally permissible.

Cafeteria also allows us to see that there is a distinction between instances of nudging and interventions grounded by LPC. The former is a change in the choice architecture that preserves freedom of choice in a way that influences people's decisions

36 Over 75% of people take the first food at a buffet, and the first three foods encountered comprise 66% of everything they take. For empirical evidence, see Wansink and Hanks (2013).

37 Note that in *Cafeteria*, Jess has always been nudging diners (unintentionally and unconsciously). Her learning of the ways customers' decision-making is affected by food placement does not make her *start* nudging but only *become aware* that she nudges.

38 In this context, what would make the customers better off can be defined in terms of health. For example, less sugary or less fatty food would make diners better off.

39 It is the position of libertarian paternalists that, all things being equal, choice-preserving architecture ("nudges") is morally preferable to choice-limiting architecture ("shoves"). However, it is not the position of libertarian paternalists that limiting options is *never* morally permissible. They recognize that, in some cases, limiting choices is morally justified. For more on this, see Sunstein (2014b).

independently of whether the choice at the center of such influence makes the individual better off or worse off. The latter is a welfare-aimed instance of nudging. Although interventions grounded by LPC are instances of nudging, not all instances of nudging are interventions grounded by LPC. This is the case because instances of nudging can influence people to make choices that either increase, decrease, or maintain their welfare, whereas interventions grounded by LPC will always influence people to make choices that *increase* their welfare. In *Cafeteria*, options (a–c) are instances of nudging, but only (a) is an intervention grounded by LPC.

3. OBJECTIONS

3.1 Biased Choice Architects

LPC's condition (iii) has it that all ill-intentioned nudges are morally impermissible. Thus, the Bad Intentions view considered in the Nudging section is not a worry for LPC. There is, however, one objection to LPC that targets the moral permissibility of all nudges. The worry here is that (iii) focuses on the choice architect's intentions, not on the actual consequences of the nudge. Thus, even if LPC is met, there is still a possibility for nudging to go wrong despite the choice architect's best intentions—i.e., for choice architects to steer people in directions that would make them worse off instead of better off, even if the nudge is aimed at what the choice architect is justified in believing would increase the welfare of those being nudged. This is the case especially when the choice architects have biases that affect the way they present options.

Gigerenzer (2015) argues that "nudging people into what is best for them requires choice architects who actually know what is best for others."[40] He claims that this requirement involves a contradiction. On the one hand, the choice architects are subject to the very same biases as the people whom they are trying to nudge. On the other hand, choice architects are supposed to be able to discern what constitutes welfare and what would make people better off in order to nudge them in that direction.[41] Gigerenzer offers empirical evidence for his argument on the basis of systemic biases, dubbed as the 'SIC Syndrome' in the context of health care. This includes biases related to self-defense, innumeracy, and conflicts of interest, which, when in place, make the

40 Gigerenzer (2015, 367).
41 Gigerenzer (2015, 367).

choice architect "steer the public into directions that are not in their best interest."[42] One of the examples Gigerenzer offers is about health care providers who practice defensive medicine, recommending a second-best option over an option that is best for the patient in order to protect themselves from potential lawsuits.[43] According to Gigerenzer, this shows that nudging is not permissible even when aimed at making people better off, because if the choice architects are as biased as those they are trying to nudge, then the outcome could be one that makes people worse off rather than better off, independently of the choice architect's intentions.[44]

In response to this objection, it is important to first note that, although Biased Choice Architects is presented as an objection to LPC, at its core, it is really an objection against the moral permissibility of nudging as a whole. Thus, this objection is flawed in the same way that the Bad Intentions view presented in the Nudging section is. Both of these objections mistakenly assume that nudging can be avoided when this is not the case. Further, what is presented as problematic in Biased Choice Architects—i.e. that choice architects can err even when their intentions are in the right place—is not really a problem for LPC but a point in its favor.

Recall that condition (i) of LPC is that the intervention must be a nudge. It requires the intervention to preserve freedom of choice, which means it must alter the choice *context* and leave the choice *content* intact. In other words, it must modify *how* options are presented, not *what* options are presented. This is a liberty-preserving kind of choice architecture. The alternative to this is a liberty-limiting kind of choice architecture, which includes interventions that limit the choice *content* by either forbidding some options or by adding consequences to certain options (e.g. punishments or fines).[45] Although it is true that choice architects are subject to biases that may affect the directions towards which they nudge people, it is crucial to understand that this is one of the reasons why LPC concerns itself only with such interventions. Nudges are not infallible, they allow people to choose options that are different from the ones the choice architect nudges them towards. Thus, in cases like the health care example provided by Gigerenzer, the patients being nudged are able to choose what they prefer or what is best for them even if this is not the same as what

42 Gigerenzer (2015, 276).
43 Gigerenzer (2015, 276).
44 For more examples and empirical evidence on biased choice architects, see Gigerenzer (2015).
45 These interventions are known as *shoves*. See Sunstein (2014b).

the health care provider recommends. This would not be possible if the intervention were to exist in a liberty-limiting choice architecture.[46]

Take, for example, a *Cafeteria* scenario in which Jess is subject to biases that lead her to believe that what is best for diners are sugary and fatty (unhealthy) foods, as opposed to healthy foods. In this case, if she decides to arrange the food with the welfare of her customers' in mind, she would do so by placing the unhealthy options first because this is what, according to her, is in the customers' best interests. Here it is clear that even though Jess is well-intentioned, her biases lead her to nudge in a way that do not result in making the diners better off—i.e., she errs. However, given that she is merely altering the choice context and not the choice content, her customers are able to choose healthy food over the unhealthy food Jess is nudging them towards. Cases like this make arguing in favor of liberty-preserving choice architecture (nudging) more reasonable than arguing in favor of liberty-limiting choice architecture because nudges allow people to easily resist what is not in their best interests.

3.2 The Manipulation Objection[47]

Perhaps the most common objection to nudging targets its moral permissibility by appeal to manipulation.[48] This objection claims that although nudging in general can be morally permissible, there is a particular subset of nudges that is morally impermissible in virtue of being manipulative. This objection does not target all nudges, but type-1 nudges exclusively. The charge from manipulation is that type-1 nudges are manipulative in a way that type-2 nudges are not, and that this makes a relevant moral difference.[49] The objection claims that a nudge is manipulative if it influences people without sufficiently engaging their reflective and deliberative

---

46 In a liberty-limiting choice architecture, the health care provider's recommendation would not be merely a recommendation. It would be the only available option for the patient, or at least the only one they could choose without significant negative consequences.

47 This section (3.2) was previously published in Florida Philosophical Review, 19 (Spring 2020). It was part of the winning undergraduate paper at the 63rd Annual Meeting of the Florida Philosophical Association. See Joly Chock (2020).

48 See Bovens (2008), Nagatsu (2015), Hansen and Jespersen (2013), and Hausman and Welch (2010).

49 People in general tend to share this perception. For empirical evidence on people's evaluations of nudges, see Sunstein (2015 & 2016) and Hagman et al (2015).

capacities.[50, 51] The charge from manipulation can be presented as follows:

The Manipulation Objection (TMO)

1) All interventions that engage people *only* subconsciously are manipulative.
2) Type-1 nudges engage people only subconsciously.
3) Type-1 nudges are manipulative. (1–2)
4) All manipulative nudges are morally impermissible.
5) Therefore, type-1 nudges are morally impermissible. (3–4)[52]

Subliminal messages[53] are interventions that engage people only subconsciously.[54] Without getting into empirical details, suppose that subliminal messages are capable of influencing people's choices.[55] Suppose that, as part of a public health campaign, subliminal messages encouraging calorie intake reduction are introduced in TV programs. As a result, many people subconsciously reduce their calorie intake.

---

50 This is an overarching understanding of 'manipulation' drawing from various sources within the literature. Sunstein (2015b, 443–444) considers accounts by Barnhill (2014), Wilkinson (2013), Faden and Beauchamp (1986), and Raz (1988).

51 Nagatsu claims that "nudges are ethically problematic to the extent that they change individual behavior in such a way that is not responsive to the agent's reasoning process" (2015, 487).

52 Note that TMO makes no claims about the moral permissibility of type-2 nudges. This is motivated by the desire to keep some nudges while excluding others. The desire to leave type-2 nudges unscathed is itself motivated by the assumption that type-2 nudges are not manipulative (or at least not manipulative in the relevant sense), and thus not morally problematic, because they engage people consciously.

53 The word 'subliminal' means 'below threshold'. More specifically, below the threshold of consciousness. I am using 'subliminal messages' to refer to either auditory or visual messages presented below the average limits of human perception. In other words, to a signal or message that is typically unperceived consciously, yet perceived subconsciously.

54 Whether or not subliminal messages are nudges is contested. Some critics argue that subliminal messages are not nudges. However, some others, including Thaler and Sunstein, find it difficult to differentiate the two because all it takes for an intervention to be a nudge is for it to preserve freedom of choice. That is, to not limit options or attach incentives or consequences to any of them. Since subliminal messages do not do any of this, it is reasonable to think that they are nudges. However, what is commonly resisted is whether they are morally permissible. Thaler and Sunstein argue that even if they are nudges, subliminal messages are not instances of Libertarian Paternalism. My argument in this paper assumes that subliminal messages are nudges. More specifically, type-1 nudges given that they engage us only subconsciously.

55 Subliminal messages are commonly tied to negative connotations. Some people believe that subliminal messages can never improve the welfare of those who are subject to such messages. If this is true, then subliminal messages are not a challenge for LPC because they fail to meet (iii). For the purpose of this paper, my assumption is that subliminal messages can be used to increase the welfare of those who are subject to them.

However, they may be unaware that they are doing so, largely because they are unaware of the subliminal message itself. Even though the outcome may be desirable and beneficial, the intuition here is that such a use of subliminal messages is manipulative. This intuition gives reasons to accept premise 1.

Premise 2 is true by definition, and premise 3 follows from 1 and 2. Premise 4 can be motivated by appeal to autonomy. Hausman and Welch (2010) define 'autonomy' as "the control an individual has over his or her own evaluations and choices."[56] They argue that when nudges do not involve "rational persuasion", the subject's autonomy is diminished.[57] Here the degree of reflection that takes place in an individual's decision-making is proportional to the amount of autonomy they exercise. Since manipulative nudges engage individuals only on subconscious levels, the degree to which they exercise their autonomy, if at all, is very low. Thus, given that for a nudge to be permissible, it is necessary for the individual to exercise their autonomy, and to exercise their autonomy, it is necessary for them to engage in System 2 thinking, the central worry is that when nudges engage an individual merely subconsciously, the individual is not able to exercise their autonomy at a degree that would render the nudge morally permissible. This presumably leads to the choice architect potentially having more control over the individuals' choices than the individuals themselves. It is such a diminishing of autonomy what makes type-1 nudges morally impermissible according to TMO.[58]

LPC[59] makes no distinction regarding which System (1 or 2) is engaged in nudging. Thus, under LPC, type-1 nudges and type-2 nudges are equally permissible. If TMO succeeds and all type-1 nudges are morally impermissible, then LPC's sufficiency claim is false because LPC fails to account for the moral difference between type-1 nudges and type-2 nudges.

3.2.1 Implications

TMO makes a moral distinction between type-1 nudges and type-2 nudges. This distinction is motivated by the desire to leave type-2 nudges unscathed, which is

56 Hausman and Welch (2010, 128).

57 Hausman and Welch (2010, 127).

58 In contrast, Hansen and Jespersen claim that type-2 nudges facilitate and increase freedom of choice and "empowerment" by means of reflective engagement (2013, 24).

59 Recall that LPC holds that for an intentional intervention to be morally permissible, it has to be (i) a nudge, (ii) transparent, and (iii) aimed toward the welfare of those being nudged.

problematic. Such a project rests on the assumption that type-2 nudges exist independently from type-1 nudges, but this is simply not the case. While type-1 nudges and type-2 nudges are exclusive—a nudge is either of the type-1 or type-2 variety—they are importantly related. It is impossible to cleanly separate the two types of nudges because while System 1 thinking can—and often times does—operate on its own, System 2 thinking always depends on System 1. For instance, in the nutritional labels type-2 nudge example, reading and reflection only come into the picture after the subject notices and reacts to the placards. However, the noticing and reacting to the placard does not engage System 2 thinking. Without this initial reaction (System 1), there would be no reflection (System 2). There is simply no way to skip straight to engaging the subject's System 2 thinking.

Given that the distinction between these types of nudges is not as clear as it is assumed by proponents of TMO, if the objection is successful, it not only undermines the moral permissibility of type-1 nudges but the moral permissibility of type-2 nudges as well. In other words, if type-1 nudges are morally impermissible and type-2 nudges cannot exist without type-1 nudges, then there would be no morally permissible nudges at all. By arguing that type-1 nudges are impermissible, proponents of TMO inadvertently make the case that *all* nudges are impermissible. This is a radical result. If TMO succeeds, many everyday interactions would be morally impermissible. To appreciate the radical nature of this conclusion, think about the prevalence of nudges in our everyday lives. People dress well for job interviews which nudges employers to take them seriously, students raise their hands which nudges professors to call on them, people make recommendations to their friends which nudges them to make particular choices, etc. All these 'everyday nudges' would be impermissible if we accept TMO's conclusion, given how inextricably connected type 1 and type 2 nudges are. To accept TMO's conclusion is to accept that we do something morally wrong much of the time. This is unreasonable. TMO proves too much. Therefore, its conclusion should be rejected.[60]

---

60 This evaluation of TMO follows the 'Moorean Shift', which is a way of objecting to an argument by appealing to commonsense. That is, by showing that rejecting its conclusion is more reasonable than accepting the conjunction of its premises. The move was originally proposed by G. E. Moore in response to skepticism. See Moore (1939).

This way of rejecting the objection relies on the absurdity that results from its conclusion. This rejoinder claims that something goes wrong in the argument for TMO, but it fails to identify what exactly that is. Thus, although considering TMO's implications is sufficient to reject the objection, it is not a satisfying solution. Something else should be added to complement this rejoinder. Something that can help us point out what goes wrong in the TMO argument.

3.2.2 Evaluation

Rejecting that all type-1 nudges are morally impermissible is not the same as accepting that all type-1 nudges are morally permissible. There is an important difference between type-1 nudges like the plate-size nudge and interventions like subliminal messages that creates a moral distinction between them. In this section, I address Thaler and Sunstein's attempt to make such a distinction, followed by what I think is a better alternative to do so.

3.2.2.1 Monitoring

Thaler and Sunstein condemn subliminal messages because it is impossible to monitor them. They claim that "manipulation of this kind is objectionable precisely because it is invisible and thus impossible to monitor."[61, 62] They justify this by appeal to LPC's transparency condition. With this understanding of transparency, they implicitly introduce unreasonable expectations about people's capacity to monitor nudges. On their account, then, for a nudge to be transparent, it must be practically monitorable; and for it to be practically monitorable, the nudge must be practically detectable. This brings about a problem. In the plate-size example, the nudge is not practically monitorable.[63] Thus, if subliminal messages are impermissible because they are not practically monitorable, then nudges like the plate-size nudge and many 'everyday nudges' are impermissible too. Here, we can see that the transparency

61 Thaler and Sunstein (2009, 246).

62 Note that a person with the right technological equipment could detect messages that would be undetectable otherwise. Thaler and Sunstein's claim, then, does not mean that subliminal messages are in *fact* "impossible" to monitor but perhaps only that they are *practically* undetectable.

63 This is the case even more so if the person being nudged is unaware of the concept of nudging and the biases that influence their thinking and decision-making.

condition is extremely restrictive.[64] This is sufficient reason to worry about Thaler and Sustein's rejoinder to TMO—i.e. to not accept how practically monitorable a nudge is as a good divider between permissible and impermissible type-1 nudges. Thaler and Sunstein's solution is a possible one but it overly restricts the set of morally permissible nudges. Thus, their rejoinder to TMO is not satisfying.[65]

3.2.2.2 Counteracting

A more promising response makes a distinction between the *ways* in which interventions engage our biases. Subliminal messages are morally problematic because they exploit our biases in a way that other type-1 nudges do not—they decrease our sensibility to reasons. Biases prevent our rationality status from being optimal. Nudges are intended to *counteract* already operant biases that prevent us from making choices aligned with our welfare. Subliminal messages do not *counteract* biases. Rather, they *activate* them. Thus, in a sense, while some nudges ("counteractive") either maintain or elevate our rationality status by counteracting our biases and not affecting our sensibility to reasons, other nudges ("non-counteractive") lower it even more under its already suboptimal level by activating additional biases on top of the already operant ones and decreasing our sensibility to reasons.[66]

Suppose Jess wants to influence her diners to consume healthy desserts (e.g. fruit) over less healthy options (e.g. cupcakes/donuts) because doing so will improve their welfare. She knows diners are biased to fill their plates with what is presented first. But now she learns about subliminal messages. Thus, she considers two possible options to

---

64 Hansen and Jespersen make a similar point regarding the transparency condition being too restrictive. However, they are concerned mainly with nudging in the context of public policy and the private sector. To make the point, they analyze the connection between the transparency condition and disclosure in relation to John Rawls' *publicity principle* (2013, 15–17).

65 Note, too, that there is another way in which Thaler and Sunstein's distinction restricts the set of morally permissible nudges even more. It rules out a lot of type-2 nudges for the reasons I mention above having to do with type-2 nudges' reliance on type-1 nudges. This is the case because a great deal of type-1 nudges are not practically monitorable. Thus, if non-practically monitorable type-1 nudges are impermissible, then all the type-2 nudges that rely on them are impermissible as well.

66 Recall that the sense of rationality that is relevant here is one closely tied to biases. The more biased the decision-making process is, the lower the rationality status of the individual and the choice they make, independently of whether the choice is "good" in terms of outcomes or consequences. So, even when one's decision is "good" (outcome-wise), it is irrational (in some way) if it is the result of biased decision-making.

nudge her customers: (i) placing healthier desserts first in line, or (ii) placing subliminal messages or playing a subliminal song throughout the cafeteria. The former is a counteractive nudge. This nudge relies on changing the environment to prevent people's already operant bias (tendency to choose whatever is placed first in line) from leading them towards poor choices (unhealthy over healthy food). This nudge utilizes the same already operant bias to influence people to make better choices not by eliminating the bias but by counteracting it with a change in the environment (arranging the food so that healthy foods are placed first). Further, counteractive nudges do not decrease people's sensitivity to reasons because, although people might not recognize how the choice architecture affects their choices about what they eat (or even recognize that this is doing so at all), they can at least recognize the choice architecture itself. They recognize that food is placed in some way and that they have different options to choose from, even if that all happens at System 1 level without reflection. The idea here is that, if people started to think about and list all the possible reasons they could have for choosing the food they put on their plate, they could at some point list the placement of the food as a reason.[67] What matters here is not that they *actually* list the food placement as a reason for choosing the food they put on their plate but that there is a *potential* for them to do so.

Subliminal messages do not rely on utilizing already operant biases or changing the environment but rather on activating *additional* biases (tendency to be susceptible to subconscious auditory/visual message) to the subject's thinking. Non-counteractive nudges decrease people's sensitivity to reasons because people are unable to recognize the subliminal message within the choice architecture. Thus, the potential to list the subliminal message as a reason for choosing their food is non-existent. It is not only that they cannot *actually* list the subliminal message as a reason but that there is no *potential* for them to do so.

Given this distinction, it is possible to offer an account of morally permissible nudges based on the way nudges work. A nudge is *permissible* when it does not decrease our sensibility to reasons[68] and *impermissible* when it does.[69] This distinction

67 Note that this is the case even though the placement of the food is affecting their choices at a System 1 level.

68 Although it is possible for a nudge to increase our sensibility to reasons, it need not do so in order to be morally permissible. All that matters here is that the nudge does no decrease our sensibility to reasons.

69 This is the case even if both the counteractive and the non-counteractive nudges influence the person being nudged towards the same desirable choice, as in the example given above.

is compatible with the intuition that motivates TMO. That is, that the relationship between autonomy and reflection plays a role in the moral permissibility of nudges. The difference is that, for proponents and supporters of TMO, *actual* reflection about and determination of the reasons for our choices is necessary for exercising one's autonomy, and thus for a nudge to be morally permissible. According to my view, the *potential* for such reflection and determination is sufficient for exercising one's autonomy, and thus for a nudge to be morally permissible.

Whether this rejoinder defeats premise 1 (All interventions that engage people only subconsciously are manipulative) or premise 4 (All manipulative nudges are morally impermissible) of TMO depends upon how we understand manipulation. Central to the manipulation literature, is the question of whether or not manipulation is inherently impermissible. This is a difficult question with no clear answer.[70] But independently of what answer one adopts, my rejoinder does enough to prove that TMO is unsuccessful by showing that there is no moral problem with counteractive nudges even when they are type 1 nudges. My rejoinder can be a response to TMO for those who take manipulation as inherently impermissible and for those who do not. On the one hand, my rejoinder allows for counteractive nudges to be manipulative but permissible.[71] Thus, if manipulation is not inherently impermissible, then premise 1 is true and my rejoinder defeats premise 4 by showing that not all manipulative nudges are morally impermissible. On the other hand, my rejoinder can also allow for counteractive nudges to be permissible by not being manipulative.[72] Thus, if manipulation is inherently impermissible, then premise 4 is true and my rejoinder defeats premise 1 by showing that counteractive nudges are not manipulative, for which not all interventions that engage people only subconsciously are manipulative.

---

70 The extreme view that manipulation is inherently impermissible resembles Kant's hardline view that lying is always wrong and can be justified in a similar way. For this view, see Kant's ethical writings in Gregor (1996). For the view that manipulation is merely *prima facie* wrong, and thus, not inherently impermissible, see Baron (2014).

71 The possible position to take here is that counteractive nudges are manipulative because they engage individuals merely subconsciously, but they are morally permissible because they do not decrease one's sensibility to reasons. Thus, do not prevent one from exercising one's autonomy.

72 The possible position here requires a revision in the definition of 'manipulative nudge' from merely engaging the individual subconsciously to not allowing the individual to exercise their autonomy. Thus, counteractive nudges are not manipulative because they allow the individual to exercise their autonomy.

4. LPC*

Disproving TMO proves that not *all* type-1 nudges are morally impermissible. However, this is not the same as proving that all type-1 nudges *are* morally permissible. According to my response *some* (non-counteractive) type-1 nudges are impermissible. Thus, even though it is sufficient to reject TMO, my rejoinder does not eliminate the problem for LPC because, if correct, it has the consequence that meeting LPC is insufficient for an intentional intervention to be morally permissible. One way of solving the problem is to add the condition that the intervention must be (iv) counteractive. The revised sufficiency claim with the additional condition is then:

> LPC*: Any intervention that is (i) a nudge, (ii) transparent, (iii) aimed towards the welfare of those being nudged, and (iv) counteractive is morally permissible

This revision takes care of the challenge from manipulation while maintaining that most libertarian paternalistic interventions are morally permissible given that this revision minimally restricts the set of nudges that are morally permissible under the existing criteria.

Independently of whether LPC* is successful, my conclusion holds—i.e. TMO fails as an objection against the moral permissibility of *all* type-1 nudges. Perhaps my rejoinder to TMO is not more convincing/satisfactory than the other ones considered in this section. Nonetheless, it seems to get at something more reasonable and practically applicable than the rest. The rejoinder to TMO from considering the implications fail to explain what goes wrong with the objection. Thaler and Sunstein's rejoinder places an unreasonable expectation over people who are nudged and significantly restricts the set of permissible nudges. In contrast, my rejoinder explains what goes wrong with TMO while also suggesting a possible solution that does not place an unreasonable expectation over people who are nudged and just minimally restricts the set of permissible nudges. All while complementing the intuition that TMO's implication—i.e. that all nudges, including 'everyday nudges', are impermissible—is unreasonable.

5. APPLICATIONS

Not only are nudges everywhere, they are also highly effective. Thus, any good criteria for the ethical use of intentional nudges must be applicable in practice. One example of an intervention grounded by LPC* that has been proven to significantly impact people's choices about becoming organ donors is a simple change of default

options in organ donation forms. A study shows that when forms have an opt-in default (requires explicit consent—people have to check a box if they want to be donors) less than 20% of people typically become donors. When forms have an opt-out default (presumes consent—people have to check a box if they *do not* want to be donors) over 98% of people typically become donors.[73] This disparity is due to the default effect—tendency to stick with what is pre-selected, regardless of what that is. Such default is designed into the donation form.

Interventions grounded by LPC* have been used to effectively increase people's retirement savings and recycling, as well as to reduce pollution and speeding, among other things.[74] However, one worry with LPC* is that it is perhaps too ideal to the point where it may not be feasible in practice. This worry has to do with condition (iii), which excludes some nudges that intuitively seem morally permissible. That is, nudges that do not make the person being nudged better off but do not make them worse off either. We can think of the kind of nudges that work in this way as *neutral* nudges.[75] One example of a neutral nudge is an intervention involving an intermediate choice architect. Suppose I go to the supermarket with my parents. My dad loves sugar cookies but he is diabetic, so it is in his best interests to limit his consumption of such cookies. I know that if I try to nudge him into buying fruits instead of cookies, the nudge will likely be unsuccessful. That is, he will likely resist it. However, if my mom tries to nudge him, there is a high probability that the nudge will be successful.[76] So, instead of nudging my dad directly, I decide to nudge my mom into nudging my dad. Here my mom's nudge to my dad is an intervention grounded by LPC* because it is aimed at improving his welfare. My nudge to my mom, however, is not an intervention grounded by LPC* because that intervention is

---

73 Johnson and Goldstein (2004).

74 Thaler and Sunstein (2009).

75 It is important to recognize that to neutrally nudge *per se* is not actually possible. When options are presented, they *must* be presented in *some* way. That is significant, since *any* way of presenting options inevitably influences people's decisions. Thus, choice architects are *always* influencing the decisions of those to whom they present options. No nudge is ever perfectly neutral. Thus, to "nudge neutrally" is better understood as to "nudge as neutrally as possible".

76 The nudge being "successful" means that the desired outcome is achieve. In this case, that is to nudge my dad into buying fruit instead of cookies. There are many potential ways in which he could be nudged, as well as several potential reasons for why my mom's attempt to nudge my dad would likely be more effective than my attempt. The specifics of the case are not relevant and do not affect my argument, so I will not elaborate on the details here.

aimed at my dad's welfare, not at my mom's welfare. My mom is neither better off nor worse off as a result of that nudge.

Nudges that make the choice architect better off while not significantly affecting the welfare of the person being nudged are also neutral nudges. For example, let's say a friend is craving a burger and asks me if I want to go out for dinner with her. There are several restaurants we could go to which have burgers in their menus. However, not all of them offer good vegetarian options. Since I am a vegetarian, I will nudge my friend into choosing one of the restaurants that offer the best vegetarian meals. Assuming the quality of the burgers is almost the same in all available restaurant options, this nudge does not significantly affect the welfare of my friend because she will eat a burger no matter where we go. Her choice of meal remains available, so my nudge does not make her better off or worse off. The nudge does make *me* better off because I will get a better meal than I would have otherwise.

Some mundane nudges that do not have a significant impact on the welfare of either the choice architect or the person being nudged are also neutral nudges. Suppose you need to fill out a form. There are two identical pens in front of you and I nudge you to choose the pen to the left. Your options are identical, and choosing the pen to the left does not improve your welfare any more than choosing the pen to the right does. Intuitively, this and the two other neutral nudge examples mentioned above seem morally permissible. According to LPC*, however, they are not. This is because they are not aimed towards the welfare of the person being nudged. The supermarket example is aimed at the welfare of a third person, the dinner example is aimed at the choice architect's welfare, and the pens example is not aimed at welfare at all. If condition (iii) of LPC* is preventing these examples from being morally permissible, then it seems that a different set of sufficient conditions for the moral permissibility of intentional interventions is needed to account for neutral nudges. What seems to be important for such a set is that the intervention does not make the person being nudged worse off, not necessarily that it makes them better off. That is, the set of conditions under which neutral nudges are morally permissible is one in which

nudges do not produce harm to the person or group of people being nudged.[77]

Any standard that aims to categorize nudges from an ethical standpoint while also aiming to be practically applicable must take into account neutral nudges. LPC* excludes neutral nudges, but that does not mean that its sufficiency claim is false.[78] It just means that the bar the criteria sets is too high to make LPC* an applicable ethical standard in practice. A more reasonable standard is to nudge neutrally. Thus, it is possible to propose a set of sufficient conditions for the moral permissibility of intentional interventions that is similar to LPC* but allows for neutral nudges to be permissible.

> LPC**: Any intervention that (i) is a nudge, (ii) transparent, (iii) produces no harm to those being nudged, and (iv) is counteractive is morally permissible.

Modifying condition (iii) from aiming towards welfare to producing no harm allows for neutral nudges that are intuitively perceived as morally permissible to be counted as such. LPC**, then, expands the extent to which nudges can be ethically implemented by offering a criteria for the moral permissibility of intentional interventions that is more applicable in practice than the one offered by LPC*.

5.1 Permissible vs. Preferable

Although, LPC* leads to the categorization of neutral nudges as impermissible, this set of sufficient conditions for the moral permissibility of intentional interventions is still preferable over LPC**. Nudging as neutrally as possible is the lowest bar that must be met for an intentional intervention to be permissible. However, choice architects should aim to increase the welfare of those they nudge whenever this is

---

77 Note that the view I propose here is not dependent on any particular view of 'harm.' The same is the case with LPC*—it does not depend on any particular view of 'welfare'. These two terms are complex and difficult to define because a range of considerations must be taken into account for such a task, e.g. immediate vs distant harm/welfare, and the relation of harm/welfare to various things such as health, economic and social status, among other things. Furthermore, it may be the case that there is no *one* definition for each of these terms. It is likely that in different circumstances, different considerations outweigh the rest. Trying to define these terms, however, is outside the scope of my project. Further, the view I present here stands without endorsing any particular view of 'harm' or 'welfare.'

78 The fact that LPC* is a set of sufficient conditions for the moral permissibility of intentional interventions does not mean that it is the *only* set of sufficient conditions. It is possible for there to be other sets of sufficient conditions for the moral permissibility of intentional interventions that is different from LPC*.

possible. Recognizing that this is preferable should motivate choice architects to aim for more than just the minimum ethical standard. In other words, it is morally good enough to nudge neutrally, but it is morally better to nudge for good. To nudge neutrally (i.e., to nudge according to LPC**) is ethical, but to nudge for good (i.e., to nudge according to LPC*) is a more preferable alternative.

5.2 Examples

Throughout this paper, when talking about LPC* I focused exclusively on the welfare of the person being nudged. It is important to note that, although all that LPC* requires is that the person being nudged ends up better off, interventions grounded by LPC* do not benefit *only* the person being nudged. In many cases, the choice architects themselves or third parties not directly affected by the intervention benefit from it as well. This section contains some cases that exemplify this.

5.2.1 Towel and Linen Reuse in Hotels

The use of signs that encourage hotel guests to reuse their towels and linens, is one example of a small, cost-effective nudge that brings about benefits for everyone involved. These signs typically display messages similar to the following: "Almost 75% of guests help the environment by using their towels more than once. You can join your fellow guests and help save the environment by reusing towels during your stay." For a number of possible reasons ranging from the basic power of social norms that lead people to follow the crowd, peer pressure, or perhaps even the perceived kinship among hotel guests, a lot of people end up reusing their towels.[79] This nudge meets LPC* because it is easy to resist (people can simply refuse to reuse their towels with no consequence), it increases the welfare of the person being nudged by increasing the welfare of society as a whole (it increases water-conservation while reducing greenhouse gas emissions that contribute to climate change), and it is counteractive (it is possible for the person being nudged to attribute their choice to reuse towels to the message). Nudging hotel guests to reuse their towels and linens is also beneficial for the choice architects (hotels) because it reduces their costs related to electricity, water, and soap usage, as well as the labor of restocking and washing towels. Less washing also translates into a reduction in replacement costs given that

79 See Goldstein, Cialdini, and Griskevicius (2008).

the wear and tear on towels and linens is reduced, which extends their lifespan. The American Hotel and Lodging Association estimates that this simple nudge reduces the number of loads of laundry washed—as well as the related costs—by 17%.[80] Calculations based on a 72% occupancy rate and 22% guest participation rate in a typical 300-room hotel estimate a reduction of water usage by 51,840 gallons and detergent usage by 346 gallons per year. The estimated translation of such results into energy, water, and labor savings are approximately $10,407 and $13,876 annually for towels and linens respectively.[81]

5.2.2 Soap Dispenser Alarm

Safeguard, a soap company, invented the "Germ Alarm" to increase hand washing among public toilet users.[82] The Germ Alarm is a soap dispenser warning alarm that alerts people to wash their hands after using the toilet. By using pressure sensors, every time someone gets out of a stall, the alarm is triggered and can only be deactivated by pressing the wall-mounted soap dispenser. Since its introduction during Global Handwashing Day in October 2013, these Germ Alarms have been installed inside public toilets in fast food restaurants, offices, and schools across the Philippines. This nudge meets LPC* because it is easy to resist (people can simply ignore the alarm and leave the restroom with no added consequences), it improves the welfare of those being nudged by reducing the likelihood of them getting or spreading a disease, and it is counteractive (it is possible for the people being nudged to attribute their choice of washing their hands to the sounding of the alarm). This nudge also benefits the choice architects (the places where the alarms are located) by decreasing the number of disease-causing germs that are spread in their environment. The soap company itself is also benefitted in that more people washing their hands means more soap consumption, which translates to more sales and more profit for the company.

---

80 See Howard (2014).
81 See American Hotel and Lodging Association (n.d.).
82 See PSFK (2014).

5.2.3 Urinals Around the World

One of Thaler's favorite illustrations of an intervention grounded by LPC* dates back to the early 1990s at the Schiphol airport in Amsterdam. When the floors in the men's restrooms were getting too sticky, the cleaning manager implemented a nudge to try to reduce "spillage" around urinals. This nudge took the form of an etched photorealistic image of a fly placed into each urinal just above the drain. The idea was to give men something to aim at. The result was an 80% reduction in spillage and a cleanup reduction of 20% on average, which translated to an 8% reduction in bathroom cleaning costs at the airport.[83] Since then, urinal flies have been showing up in restrooms all over the world, including Terminal 4 of New York's JFK airport, Moscow, Munich, Singapore, Seattle, Detroit airports, Purdue University, the University of Colorado, Broward Community College, and throughout Holland.[84] The nudge has also been implemented with images other than a fly. A soccer-themed "target" with a small plastic goal in the bowl's center was implemented in Bonn, Germany.[85] This nudge meets LPC* because it is easy for the people being nudged to not aim at the image, it improves the welfare of those being nudged by preserving a clean environment, and it is counteractive (it is easy for men to attribute their action to the sticker in the urinal). Placing a picture of a fly in urinals is a simple, inexpensive way to reduce spillage that also benefits choice architects by significantly reducing their cleaning costs.

6. CONCLUSION

Central to the debate about nudging is the question of whether it is morally permissible to intentionally nudge others. I started my project by moving away from this question to the more productive question about *how* it is morally permissible to intentionally nudge people. In response to this, I presented the Libertarian Paternalistic Criteria (LPC), which proposes a set of sufficient conditions for the moral permissibility of intentional nudges. Having done this, I defended such criteria from the Biased Choice Architects objection, which claims that nudging in general is morally impermissible given the potential of choice architects to err. I also defended

---

83 See Ingraham (2017).
84 Thaler and Sunstein (2009).
85 Thaler and Sunstein (2009).

LPC from The Manipulation Objection, which claims that a particular type of nudges is morally problematic because they are not relevantly different from standard cases of manipulation. After evaluating these objections, I concluded that although both are unsuccessful, the evaluation of The Manipulation Objection shows that LPC's sufficiency claim is false. Thus, I proposed LPC* as a possible revision of the criteria that avoids the problem. I end the paper by considering the applicability of LPC* as a standard in practice. After concluding that LPC* excludes nudges that are intuitively perceived as morally permissible, I propose LPC**, a different set of sufficient conditions for the moral permissibility of nudges that renders such nudges morally permissible.[86]

86 Thanks to my mentor, Dr. Jon Matheson, for providing helpful feedback, guidance, and support throughout the development of this paper.

Bibliography

Ariely, D. (2010). *Predictably Irrational: The hidden forces that shape our decisions.* New York, NY: Harper Perennial.

American Hotel and Lodging Association. (n.d.). Green Guidelines: Towel & Linen Reuse Programs. Retrieved from https://www.ahla.com/resources/green-guidelines-towel-linen-reuse-programs-0

Barnhill, A. (2014). What is Manipulation? In C. Coons & M. Webster (Eds.), *Manipulation: Theory and Practice* (pp. 51-72). Oxford, England: Oxford University Press.

Baron, M. (2014). The Mens Rea and Moral Status of Manipulation. In C. Coons & M. Webster (Eds.), *Manipulation: Theory and Practice* (pp. 98–120). Oxford, England: Oxford University Press.

Bovens, L. (2008). The Ethics of Nudge. In T. Grüne-Yanoff and S.O. Hansson (Eds.), *Preference Change: Approaches from Philosophy, Economics, and Psychology* (pp. 207–219). New York, NY: Springer.

Faden, R. & Beauchamp, T. (1986). *A History and Theory of Informed Consent.* Oxford, England: Oxford University Press.

Gigerenzer, G. (2015). On the Supposed Evidence for Libertarian Paternalism. *Review of Philosophy and Psychology, 6*(3), 361–383.

Glaeser, E. L. (2006). Paternalism and Psychology. *The University of Chicago Law Review, 73*(1), 133–156.

Goldstein, N. J., Cialdini, R. B., & Griskevicius, V. (2008). A Room with a Viewpoint: Using Norms to Motivate Environmental Conservation in Hotels. *Journal of Consumer Research, 35*(3), 472–482.

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. (2016). The Definition of Nudge and Libertarian Paternalism: Does the Hand Fit the Glove? *European Journal of Risk Regulation, 7*(1), 155–174.

Hansen, P. & Jespersen, A. (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. *The Journal of Political Philosophy, 18*(1), 123–136.

Howard, B. C. (2014). Hotels Save Energy with a Push to Save Water. Retrieved from news.nationalgeographic.com/news/energy/2014/02/140224-hotels-save-energy-with-push-to-save-water/

Ingraham, C. (2017). What's a urinal fly, and what does it have to do with winning a Nobel Prize? Retrieved from https://www.washingtonpost.com/news/wonk/wp/2017/10/09/whats-a-urinal-fly-and-what-does-it-have-to-with-winning-a-nobel-prize/?noredirect=on

Johnson, E. J. & Goldstein, D. G. (2004). Defaults and Donation Decisions. *Transplantation, 78*(12), 1713–1716.

Joly Chock, V. (2020). The Moral Permissibility of Nudges. *Florida Philosophical Review, 19*(1), 33-47.

Kahneman, D. (2011). *Thinking Fast and Slow.* New York, NY: Farrar, Straus and Giroux.

Kant, I. (1996). *Practical Philosophy* (Ed.) M. J. Gregor. Cambridge (England): Cambridge University Press.

Moore, G. E. (1939). Proof of an External World. *Proceedings of the British Academy, 25*, 273–300. (Reprinted in Baldwin, pp. 147–70, 1993.).

Nagatsu, M. (2015). Social Nudges: Their Mechanisms and Justification. *Review of Philosophy and Psychology, 6*(3), 481–494.

PSFK. (2014). Soap Dispenser Alarm Shames People Into Washing Their Hands. Retrieved from https://www.psfk.com/2014/04/alarm-soap-dispenser.html#!E7xwd

Rawls, J. (1971). *A Theory of Justice.* Cambridge (England): Harvard University Press.

Raz, J. (1988). *The Morality of Freedom.* New York, NY: Oxford University Press.

Shafir, E. (2013). *The behavioral foundations of public policy.* Princeton, NJ: Princeton University Press.

Sunstein, C. R. (2014). Nudging: A Very Short Guide. *Journal of Consumer Policy, 37*(4), 583–588.

Sunstein, C. R. (2014b). Nudges vs. Shoves. *Harvard Law Review Forum, 127*(6), 210–217.

Sunstein, C. R. (2015). Do People Like Nudges? *Administrative Law Review, 68*(2), 177–232.

Sunstein, C. R. (2015b). The Ethics of Nudging. *Yale Journal on Regulation, 32*, 414–450.

Sunstein, C. R. (2016). People Prefer System 2 Nudges (Kind Of). *Duke Law Journal, 66*, 121–168.

Thaler, R. H. (2015). The Power of Nudges, for Good and Bad. Retrieved from https://www.nytimes.com/2015/11/01/upshot/the-power-of-nudges-for-good-and-bad.html

Thaler, R. H. & Sunstein, C. R. (2003). Libertarian Paternalism is not an Oxymoron. *The University of Chicago Law Review, 70*(4), 1159–1202.

Thaler, R. H. & Sunstein, C. R. (2009). Nudge: Improving decisions about health, wealth, and happiness. Westminster, London: Penguin Books.

Tversky, A. & Kahneman, D. (1974). Judgement Under Uncertainty: Heuristics and Biases. *Science, 185*, 1124–1131.

Wansink, B. & Hanks, A. S. (2013). Slim by Design: Serving Healthy Foods First in Buffet Lines Improves Overall Meal Selection. *PLoS ONE, 8*(10).

Wansink, B., Just, D., & Payne, C. (2009). Mindless Eating and Healthy Heuristics for the Irrational. *American Economic Review: Papers & Proceedings, 99*(2), 165–169.

Wansink, B. (2004). Environmental Factors that Increase the Food Intake and Consumption Volume of Unknowing Consumers. *Annual Review of Nutrition, 24,* 455-79.

Wilkinson, T. M. (2013). Nudging and Manipulation. *Political Studies, 61*(2), 341–355.