Essays on Ethical Nudging

Master Thesis: Behavioral Economics

Prithvi Shashikant

23 August 2016

Erasmus University Rotterdam

Preface and Acknowledgements

This thesis would not have been possible without help from a staggering number of friends and well-wishers. I would like to extend my deepest thanks to Peter Wakker - my thesis supervisor. I have lost count of the number of hours we spent discussing our inevitably differing opinions on ethics, nudging, architecture, paternalism, and the kitchen sink. I have also lost count of the number of times he has agreed to let me change my defense date.

A huge shout-out to all my friends who spent time reading and constructively criticizing my thesis to make it the best it could be. Particularly Niclas Kandzia, with whom a discussion is both illuminating and never-ending; Pasi Rantanen, who made sure I did not overuse the word 'posit'; Linda Smolka, who was significantly more excited about my thesis topic than I was; Poorvi Iyer, who listened patiently while I desperately attempted to explain the nuances of benefit based obligation and ethical redistribution; Sten te Vogt, whose approval of the five commandments of Prithvian ethics means more to me than he realizes; and Debashish Roy, who recommended that I read *Justice*[1], without which this thesis would not be halfway where it is now.

Finally, I am incredibly grateful for the unwavering support from my family. My mother, who only suggested changes she knew I could deal with; and my father, who spent 2 hours reading my thesis and expressed interest in knowing more – the best compliment I could have received.

Niclas affectionately refers to this text as being more akin to a twelve-thousand-word opinion piece than to a master thesis. There may be some modicum of truth to that statement. I hope you enjoy the read.

[1] *Justice*: 'Justice: What's the Right Thing to do?' A book by Michael J Sandel.

Table of Contents

Chapter 1 – An Introduction to Nudging and Policy

What is Nudging?

Human decision making does not exist in a vacuum. We are susceptible to changing preferences based on the ways in which choices are presented to us. Nudging is an attempt to modify the environment in which we make choices (now popularly known as choice architecture) with the express intent to influence our behavior.

A nudge, as defined by Richard Thaler and Cass Sunstein, is "an aspect of choice architecture that alters people's behavior in a predictable way, without forbidding any options or significantly changing their economic incentives" (Thaler & Sunstein, Nudge: Improving decisions about health, wealth and happiness, 2009, p. 6). An important stipulation for an intervention to count as a nudge is that it must be cheap and easy to avoid. Defaulting the retirement savings rate at 7% with options to pick any other rate is a nudge. Forbidding any rate below 7% is not.

Nudging is more common now than ever, with governments, corporations, and even individuals realizing the enormous benefits to correctly implemented choice architecture (Thaler, Sunstein, & Balz, 2014). Concerns over the ethical constraints of nudging have come up just as quickly, with opinions ranging from calling Sunstein "the most dangerous man in America" (Baude, 2014), to scientifically valid objections about ethical nudging that we are "more likely to accept unethical nudging in the future if we become habituated to nudging now" (Selinger & Whyte, 2011).

Nudges have been shown to be effective in a vast array of settings. Companies like the Behavioral Insights Team in the UK have shown that when implemented correctly, nudges can influence behavior in a very significant way. This is why it is so important now to discuss ethics

in the context of nudging - to avoid a situation where we were so preoccupied with whether we could, that we never stopped to think if we should (Spielberg, 1993).

Before we attempt to understand the intricacies of ethical nudging, it is useful to discuss concepts that are salient in both ethical reasoning and nudge theory. We begin with two concepts – paternalism and libertarianism.

Paternalism

Paternalism is defined as the "interference of a state or an individual with another person, against their will, defended or motivated by a claim that the person interfered with will be better off or protected from harm" (Dworkin, 2016)**.** Paternalism traditionally arouses strong and polarized opinions. However, while the wording in the definition can bias people, there is nothing inherently wrong with paternalism as a philosophy – at least not as defined in this thesis.

Paternalism has frequently been criticized on the grounds that 'they will be better off' is not a good enough reason to interfere with people's lives. John Stuart Mill, a popular philosopher, wrote:

> ...the only purpose for which power can be rightfully exercised over any member of a civilized community, against his will, is to prevent harm to others. His own good, either physical or moral, is not a sufficient warrant. He cannot rightfully be compelled to do or forbear because it will be better for him to do so, because it will make him happier, because, in the opinion of others, to do so would be wise, or even right (Mill, 1966, pp. 1-147)

More widely accepted forms of 'soft paternalism' exist now, that attempt to influence behavior to improve welfare, without significantly impacting liberty and autonomy. I believe that it is not possible for ethical nudging to exist without the underlying objective of making people better off

– an important tenet of soft paternalism. Even a nudge that preserves liberty, or encourages freedom of choice is implemented with the idea that more choices or more liberty are good things for people to have.

This does not mean that all nudging follows paternalism – only that ethical nudging cannot exist independent of some form of explicit or implied paternalistic influence.

Libertarianism

Libertarianism is a political philosophy that upholds liberty as its principal objective. Libertarians seek to maximize autonomy and freedom of choice, and believe that the role of government is to protect individual rights (Boaz, 2016). The inherent belief here (as opposed to paternalism) is that people are capable of making their own decisions, and interventions have no business running people's lives.

One of the most common criticisms of libertarianism stems from the fact that unregulated markets lead to increased social inequality. The presence of a governing body that, for lack of a better word, interferes with both free markets and people's lives is necessary for society to function within an optimal level of equality.

It is also unrealistic to believe that people are capable and willing to make decisions about all aspects of their lives, and doing so would lead to very real problems (including paralysis from too many choices and increased cognitive burden from making too many decisions). In this case, it is the responsibility of a governing body to set defaults, and enable welfare promoting decisions for citizens.

Government interference is staunchly anti-libertarian, and most nudging in public policy would qualify as interference of some sort. That said, freedom of choice – implemented correctly – is

very important in nudging. So much so, in fact, that restricting it is grounds for disqualifying an intervention from counting as a nudge.

I make no value judgements about libertarianism in this thesis. There is nothing inherently bad or wrong about libertarianism, just like there is nothing bad or wrong about paternalism.

Libertarian Paternalism

Libertarian Paternalism is a philosophy that suggests that it is possible for policy to respect freedom of choice and autonomy, while also affecting behavior and influencing choices in a way that makes people better off as judged by themselves (Thaler & Sunstein, 2009). The idea behind libertarian paternalism is that it will benefit people who behave irrationally by nudging them towards choices that make them better off (as judged by themselves), while maintaining the freedom to deviate to any other choice.

Libertarian Paternalism has traditionally been closely tied to nudging. They are, however, logically different terms. It is possible to modify choice architecture while keeping true to libertarian paternalistic guidelines (nudge people to reduce procrastination). It is also possible to implement policy that would not qualify as a nudge while keeping true to libertarian paternalistic guidelines (reduce procrastination by paying people money when they work). Similarly, many nudges follow libertarian paternalism, but nudges that violate libertarian paternalism also exist. Libertarian paternalism qualifies as a philosophy that can serve to guide policy, while implementing nudges is a type of policy in and of itself. I elaborate on nudging independently of libertarian paternalism later in this thesis.

Libertarianism rests on the principles of respecting people's autonomy and freedom of choice, with minimal intervention by public policy or government. This concept runs directly contradictory to paternalism – policy that restricts autonomy and choices with the inherent

belief that it is for the benefit of the people involved. As we discussed previously, paternalism is a salient premise in nudging. Even if we postulate that nudging does not run contrary to respecting autonomy (and we do), it is implicit that paternalism is present in the decision to nudge individuals toward a decision that is different than the one they would have picked had there been no nudge implemented. It seems difficult then, to imagine that both philosophies can co-exist.

Thaler and Sunstein claim that libertarian paternalism is not an oxymoron, and that it is perfectly possible to nudge people toward making choices that are better for themselves while maintaining freedom of choice. They say that nudging can be useful in situations where people tend to make sub optimal choices, and we frequently encounter situations where some form of choice architecture is inevitable (Sunstein & Thaler, 2003). In these situations, libertarian paternalism dictates nudging in the direction of welfare while maintaining freedom of choice.

Choice Architecture and Environment

There may be some disagreement about what constitutes as choice architecture. Sunstein includes an amusing anecdote in his paper about choice architecture that reads as follows:

> Consider in this light a tale from the novelist David Foster Wallace: "There are these two young fish swimming along and they happen to meet an older fish swimming the other way, who nods at them and says 'Morning, boys. How's the water?' And the two young fish swim on for a bit, and then eventually one of them looks over at the other and goes 'What the hell is water?'" This is a tale about choice architecture. Such architecture is inevitable, whether or not we see it. It is the equivalent of water. Weather is itself a form of choice architecture, because it influences what people decide. Human beings cannot live without some kind of weather. Nature nudges. (Sunstein C. , 2014)

It is possible to distinguish between two different types of choice architecture. One is created intentionally by sentient beings, whether or not there is an intention to nudge. The default temperature on a thermostat, the design of a cafeteria, and the type of handle on a door are all examples of this. The other is simply the environment in which we make our decisions – things that (choice) architects do not control, as described by Sunstein in the above citation. These can be weather, temperature, humidity, or any aspect of nature. I make this distinction because the concept that choice architecture can exist independent of an architect is a difficult one to grasp. For many people, the word 'architect' seems to spawn an immediate need for intentional design.

We are nudged by choice architecture both intentionally and unintentionally, and much of what nudges us was not created by man, much less by an identifiable architect. For example, researchers have found that the weather on the day people purchase a car can influence which factors they consider 'more important' (Busse, Pope, Pope, & Silva-Risso, 2012). They buy cars with sunroofs when it's warm, and cars with better heating when it's cold; even though cars are clearly a long term purchase and are designed to be used through various seasons. The environment in which they make their purchase cannot be controlled by man, yet clearly influences their decision making in a systematic way. The weather is nudging them.

Both choice architecture and choice environment are capable of influencing decision making. Whether it is a supermarket designed intentionally to make everyone shop more, or the beautiful waterfall on the highway that consistently adds twenty minutes to everyone's trip because they stop to take pictures. Our primary focus will remain on intentional choice architecture when discussing ethicality in the context of nudges. It would seem, however, that the case for ethical nudging is likely to grow stronger with evidence about decision making in the context of naturally occurring choice architecture.

The specificities of what it would take for any active intervention to qualify as a nudge may be subtle and open to interpretation. However, through this thesis, only interventions that do not

restrict freedom of choice and do not significantly change economic incentives count as nudges. Any interventions that modify economic incentives (taxes/subsidies) or restrict choices (even with the intention of societal well-being) do not count as nudges. This is not to imply that interventions like these are any more or less effective, of course.

Chapter 2 – Ethical Frameworks

To understand what is and is not ethical, and what it is that makes the difference, it is important to outline frameworks by which we can gauge the ethicality of a nudge. I use two popular ethical frameworks in this thesis – Utilitarianism and Kantian ethics.

Utilitarianism

Utilitarianism, an ethical framework created by Jeremy Bentham, and improved by John Stuart Mill, states that "the highest principle of morality is to maximize happiness, the overall balance of pleasure over pain" (Sandel, 2010; see also Bentham, 1879). It is a popular concept in public policy, and is relatively simple to understand. Utilitarianism simply posits that the ethical thing to do would be the action that results in the most benefit, by maximizing pleasure and minimizing pain.

I think that utilitarianism has its flaws, primary among which is the fact that utility maximization may come at the cost of fundamental rights. Policies that involve harming a subset of the population to benefit the majority would be ethical based on utilitarian frameworks. Often, what is ethical or moral should go beyond weighing consequences and cost-benefit analyses. Some things are unethical just because they are wrong.

Whereas utilitarianism qualifies as, and was created with the intention of being an ethical standard in and of itself, it is more useful to us as a consideration in ethicality instead of an overarching framework. Utility maximization may not always be the ethical practice, but will remain an important factor in evaluating the ethics of policy.

Part of the utilitarian argument against nudging is based on the idea that individual actions are undertaken with the intention (and effect) of maximizing one's own utility, and nudges interfere with this. Unfortunately, to say that people's actions are utility maximizing simply because those are the actions they chose to perform is a nonstarter of an argument. There is ample evidence that people act against their own preferences (Thaler, 1988) and frequently do things that would (even self admittedly) not maximize their utility.

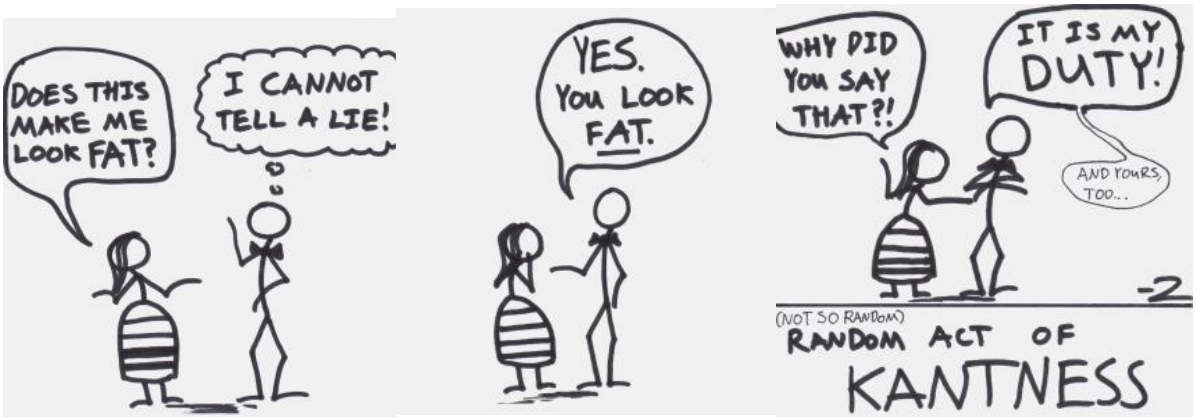
Another objection to nudging based on utilitarianism is that nudges simply impose non-monetary costs that affect decision making just like taxes and other monetary policy (Schnellenbach, 2012). In other words, nudges are capable of interfering with self-assessed utility by modifying incentives. A nudge to reduce addiction to video games by placing books at the front of a store and video games all the way at the back would be comparable to increasing the price of the video game to the point where buying the book is now the preferable option. Nudges are therefore capable of sacrificing short term utility for (possible) long term utility as seen by a welfare promoting state, and not the affected individual (Fischer & Lotz, 2014).

This is a relevant objection. Just because the definition of nudging outlines that economic incentives must not be changed, does not give us the benefit of imposing large cognitive, or non-economic costs through nudging. It is possible for a nudge to be unethical if it does so, but these kinds of violations of ethical nudging are less nuanced and easier to catch.

Kantian Ethics

Kantian ethics revolve around the principle that it is not the consequences of actions that determine how ethical they are, but that the actions themselves follow certain rules – or as Kant likes to call it, the categorical imperative (Kant, Wood, & Schneewind, 2002; see also Sandel, 2010). The categorical imperative has three constituent ethical principles. One, the maxim that one should act only in a way such that if that action was universalized, there would be no contradiction. Two, humans should be respected as rational beings, and used not as a means to an end but as an end in and of themselves. And three, that the autonomy of human beings should be respected.

Kant would argue that it is unethical to lie, not because the consequences of lying may reduce utility, but because it is ethically wrong to lie in the first place. The maxim of 'under certain conditions it is acceptable to lie' would universalize into a contradiction since people would then never be willing to accept anything as truth. It also treats humans as a means to an end (the end being the reason for lying to another person, whether it is to protect their feelings or to serve your own purposes).



Immanuel Kant – Ruining date night since 1785 (Brown, 2013)

The ethical objection to nudging based on Kantian ethics is that nudges affect the autonomy of the affected individuals by manipulating their intrinsic preferences. That is, the third maxim above is relevant. However, the concept of intrinsic preference is questionable. For example, people reverse preferences very consistently (Tversky, Slovic, & Kahneman, 1990). It is reasonable to assume that nudging impacts autonomy – since influencing behavior is one of the primary motives of nudging. Recall, however, that nudging already occurs through unintentional choice architecture and choice environment.

> As Reiss (2013: 299) puts it, 'humans with bounded rationality and willpower are subject to myriad influences anyway, and most of them do not aim to improve consumer well-being' – which implies that the notion of 'authentic' preferences does not make conceptual sense (Fischer and Lotz 2014: 11) (Schubert, 2015).

Sunstein (2014) also pointed out that if autonomy is our concern, nudging may sometimes be required to preserve it. His logic is that autonomy requires informed decision making, and eliminating biases and improving the flow of information through nudges are great ways to promote autonomy.

I believe nudging is exempt from the requirement of respecting human autonomy. Given that nudging occurs regardless of intention, is sometimes required to maintain autonomy, and people are free to choose an alternative they are not being nudged toward; it is difficult to argue that nudging is anti-autonomy.

That being said, nudging is capable of affecting autonomy in situations where people do not understand enough about the context of their decisions to effectively exercise their freedom of choice. In these situations it is possible to say that nudging does not perfectly respect autonomy. One of our objectives in promoting ethical nudging is to arrive at nudges where this compromise is an acceptable one to make.

An important distinction between utilitarianism and Kantian ethics is that utilitarianism is concerned with the consequences of actions, whereas Kantianism is concerned with the actions themselves. I believe that a combination of both is required to understand how to implement nudges ethically.

A note on Public Policy vs. Private Nudging

Many of the ethical principles that apply to nudging in public policy apply to private firms that nudge as well. The difference between the two is that private corporations have no obligation to the public other than to follow regulations - their obligations are to their stakeholders. Nudges implemented by private companies have no need to be ethical to the level that governments do. It is possible that they derive a benefit from being ethical – i.e. the public notices and rewards them with higher revenue. In a sense, the benefits that private corporations realize from being ethical are likely to be utilitarian in nature. Whether this takes away from the ethicality of their actions is a different discussion altogether.

Chapter 3 - Prithvian Ethics

As this thesis is an effort to contribute to the literature on ethical nudging, it would be amiss for me to be a bystander in the matter of what is considered ethical. I propose guidelines to promote ethical behavior specifically in the context of nudging – hereby referred to as Prithvian Ethics.

Some tenets of Prithvian Ethics are:

1) Ethical and unethical nudges exist outside the realm of libertarian paternalistic guidelines
2) For ethical nudging, intentions matter; whether you can divine them or not

3) Transparency in nudging – implemented correctly – can help promote ethics
4) Costs on rational agents through nudging can be ethically justified
5) Ethical nudging involves a situation dependent judgement call between forcing active choice and setting defaults

It is important to remember that the line between ethical and unethical is subjective. This makes it difficult to set out direct rules that if followed, would constitute an ethical nudge. What I can do, however, is to set out guidelines that show what contributes positively and negatively to nudging ethically.

Nudging without Libertarian Paternalism

Libertarian Paternalism suggests that nudges should "influence choices to make people better off as judged by themselves" (Thaler & Sunstein, Nudge: Improving decisions about health, wealth and happiness, 2009, p. 5), while maintaining freedom of choice. This condition seems like a reasonable way to judge whether a nudge was effective in the right sense. However, without casting aspersions on people's ability to make good decisions for themselves, one can very easily point out the multiple biases that go into how people perceive past events. We are prone to the self-attribution bias, which makes us believe that events that turned out well were because we made it so, and things that went bad were because of unavoidable circumstance or decisions made by someone else (Myers, 2014). For example, a nudge toward reducing obesity rates may work extremely well, but an individual asked to point out the factors that led to his weight loss in hindsight may attribute the success to his iron will and discipline and not the effectiveness of the nudge.

Self-Attribution in Action (Calhoun, 2012)

We are also affected by the hindsight bias, which makes it difficult for us to remember accurately what we believed in the past after events have already unfolded (Roese & Vohs, 2012). For example, a nudge to promote inclusiveness in society may be successful, but an individual asked about inclusiveness after the fact may believe he has always been a promoter of equal rights. Daniel Kahneman also points out that people's memory of an event can be heavily biased depending on a variety of factors including an overweighting of the last thing they remember having experienced (Kahneman D. , 2011).

It is always good for your nudge to be positive in hindsight, and nudging so people are better off as judged by themselves is a good way to avoid crossing the line from libertarian paternalism into full blown paternalism (which is not to imply paternalism is bad or wrong). However, it is important to consider that people's recollection of past events, and their ability to construct counterfactual[2] scenarios of their own lives, is imperfect at best.

Proponents of libertarian paternalism would point out that if a nudge does not make people better off as judged by themselves, it would violate the inherent principle of nudging, which is to

[2] Counterfactual: Implies 'what would have happened', a concept popular in experimental economics. The idea is to generate an idea of what would have happened if the nudge was not implemented (counterfactual), and compare that to what happens when the nudge is actually implemented – to see the difference. Needless to say, this is easier said than done.

not restrict freedom of choice. An intervention that does not make people better off as judged by themselves would carry with it the implicit assumption that a rational person would not choose to undergo the intervention at all. Therefore an intervention that violates libertarian paternalism also violates the definition of nudging which insists on not restricting choice.

The point of contention here is likely to be that the phrase 'making people better off as judged by themselves' can be interpreted as being relevant as soon as the nudge is implemented, or retroactively after the effects of the nudge have materialized. For libertarian paternalism to work independently of nudging, we have to assume that it is possible for people to be nudged toward an option that does not make them better off as judged by themselves – at the time when they are nudged. This would remove the objection to restricting freedom of choice when nudging independently of libertarian paternalism (that in the present, a person would not be nudged unless they believe the nudge would make them better off as judged by themselves, since they hold the option to deviate from the nudge). I believe that this is a perfectly reasonable assumption to make. People are both capable of not realizing they are being nudged, and also may have no judgement about whether being nudged is making them better off.

Viewed retroactively, libertarian paternalism can be violated without contradicting the definition of a nudge. The welfare promoting portion of libertarian paternalism is concerned with making people better off as judged by themselves, and not by a governing body or third party. But as we just pointed out, people are not adept at judging the past and comparing it rationally to the present. It is the freedom of choice as a critical component of the process that is significantly more important than the desirable outcome of self-assessed welfare. People also hold the free, and preferably costless option to switch in favor of maximizing their own utility if it runs contradictory to the direction in which they are being nudged. As long as the implemented nudge sticks to the definition of unrestricted options and no significant change in

economic incentive, an intervention can be called a nudge without having to satisfy both parts of libertarian paternalism. Only the freedom of choice (libertarian) part is necessary.

Libertarian paternalism is not married to nudging. The attempt to nudge people in line with libertarian paternalism is admirable, but not the only way to nudge. For example, influencing choice architecture so that the default option is changed to the one that preserves the most liberty seems a perfectly acceptable way to nudge. Policy makers are aware of the heuristics and biases that make people choose sub optimally, and nudging toward de-biasing people and then letting them make choices for themselves would also seem to be an ethical way to nudge. Even by Thaler and Sunstein's definition, this means of nudging would not be libertarian paternalistic, since de-biasing people may not be done with the express intention of making them better off as judged by themselves, only preserving their liberty and autonomy in making more rational decisions. In this scenario, the nudged individual is better off as judged by the policy maker. They may or may not be better off as judged by themselves. As previous discussed, when they make this judgement and how strongly they feel about it will affect ethicality.

Nudges that are created only for the benefit of the corporations or individuals doing the nudging can be unethical, but would still be nudges. Thaler (2015) echoes this in a New York Times opinion piece, pointing out that he is troubled by choice architecture in the private sector and urging people to 'nudge for good.'

Both ethical and unethical nudges can and do exist outside of libertarian paternalism. The exploration of the questions around ethicality of nudging assumes significance due to these instances.

Intentions, and why they matter

Why are intentions so important in ethics? The obvious answer is that intentions can be judged to reveal how ethical a nudge is. Society may be more willing to forgive an intervention that was created with good intentions but was not effective, than one that ended up working just fine, but was created by a government that people did not trust with intentions that people did not believe in.

A nudge implemented by the government in Sweden encouraged citizens to actively choose their social security portfolios. The policy was ineffective at best. The staggering amount of choices made available in the interest of 'being fair' led to two-thirds of the population making terrible investment decisions (Thaler & Sunstein, 2009). I would argue that this was a badly implemented nudge, but not an unethical one. The difference is that the intentions of the nudge were clear – the government wanted to encourage people to choose well for retirement, and also ensured there was a good default plan.

There is another reason why intentions are vital in ethical nudging. In *Justice*, Sandel (2010) points out that it is possible for moral claims of reciprocity to hold without explicit acts of consent. The example he provides reads as follows.

> Many years ago, when I was a graduate student, I drove across the country with some friends. We stopped at a rest stop in Hammond, Indiana, and went into a convenience store. When we returned to our car, it wouldn't start. None of us knew much about car repair. As we wondered what to do, a van pulled up beside us. On the side was a sign that said, "Sam's Mobile Repair Van." Out of the van came a man, presumably Sam.
>
> He approached us and asked if he could help. "Here's how I work," he explained. "I charge fifty dollars an hour. If I fix your car in five minutes, you will owe me fifty dollars. If I work on your car for an hour and can't fix it, you will still owe me fifty dollars."

> "What are the odds you'll be able to fix the car?" I asked. He didn't answer me directly, but started poking around under the steering column. I was unsure what to do. I looked to my friends to see what they thought. After a short time, the man emerged from under the steering column and said, "Well, there's nothing wrong with the ignition system, but you still have forty-five minutes left. Do you want me to look under the hood?"
>
> "Wait a minute," I said. "I haven't hired you. We haven't made any agreement." The man became very angry and said, "Do you mean to say that if I had fixed your car just now while I was looking under the steering column you wouldn't have paid me?"
>
> I said, "That's a different question" (Sandel, 2010).

In the example, Sandel is pointing out that 'we have not made an agreement' and 'I would still have paid you if you fixed my car in the first fifteen minutes' are statements that can co-exist. If the repair man fixed his car in the first few minutes, Sandel would have owed him money because the repair man had performed a service (i.e – fixing the car). This does not mean the repair man had been hired, since there was no agreement or consent. Consequently, Sandel would not owe the man any money if he had not fixed the car. Obligation can exist without an act of consent or a contract. This is known as a contract based on mutual benefit (i.e you fixed the car, I pay you money even though I did not consent to an agreement with you).

The concept of mutual benefit based contracts can be extended to nudging. The act of consent in nudging is problematic, because it may be very difficult for a person to express consent (or lack thereof) to being nudged. Imagine eating at a cafeteria and then saying "wait a minute, the size of the plates meant I ate less than I usually would have. I did not consent to that!" However, if the nudge is being implemented with the intention to help people, and the nudged provide benefit to the nudger (through taxes in public policy, or simply by adding value by being exposed to the nudge), then the contract is ethical by mutual benefit instead of consent.

That is why, sometimes, the important question to ask is not 'did I or did I not consent to being nudged.' It is to ask 'was it fair that I was nudged?' And understanding the intention that went into implementing the nudge – whether it was done with the benefit of the nudged population in mind, is critical in answering that question.

This runs directly in parallel to a concept we covered earlier, about the paternalistic implications of ethical nudging. The assumption is that if a nudge is not created with the intention of making someone better off, it cannot be ethical. In that case, the mutual benefit portion of the contract has not been satisfied, and so the implementation of the contract with no consent cannot occur. However, a nudge that is implemented with the right intentions can be exempt from the consent requirement, since it follows the principle of mutual benefit.

It may seem trivial to say policy should be created with the right intentions. Defining 'right' is difficult, and divining the intentions behind nudges can be difficult sometimes. That being said, ethical nudging must have input on the intentions behind policy – for two reasons. One, logical inferences can be made about the ethicality of a nudge based on what the 'nudger' is trying to achieve. And two, if the intention is one of mutual benefit, we have a positive effect on ethicality regardless of consent.

Transparency done right

Transparency in nudging implies openness, communication, and accountability in implementing a nudge. When evaluating the usefulness of transparency in ethical nudging, there are two main considerations. One - does being transparent really make the nudge more ethical? And two – does the transparency come at the cost of effectiveness?

Ceterus paribus, transparency almost always contributes positively to ethicality. Our concern is really with the fact that transparency can come at the cost of effective nudging. Thaler and

Sunstein are very clear about nudges, specifically in public policy, having to be as transparent as possible (Thaler & Sunstein, 2009). Others, however, argue that transparency in nudging can significantly impact the effectiveness of the nudge. It is notable that the impact transparency has on effectiveness may be overstated. Loewenstein, Bryce, Hagmann, & Rajpal (2014), found that revealing the existence of a default in the context of making end-of-life decisions did not significantly impact the effectiveness of the nudge.

Explicitly stating that a nudge has been implemented is not as efficient as letting it work in the background, for three reasons. One, some people may realize what is going on and de-bias themselves. Two, particularly vindictive people may consciously choose to work against the nudge as a reactive act of defiance against the choice architect. And three, being transparent can involve revealing irrelevant and confusing information to the nudged population.

While it may be true that some people realizing a nudge is in play will actively work to de-bias themselves, I argue that this does not have a negative effect on the ethicality of the nudge. In fact, it is quite the opposite. The loss of potential benefits from people being nudged in the absence of transparency are more than made up for by the fact that people who consciously work to de-bias themselves are going through an important, autonomy promoting process in decision making. Individuals who think actively about the nudge at work and decide to de-bias undergo some process of understanding the nudge and deciding for themselves what their best course of action is. This is better than blindly following a default, or even making a decision by themselves before the nudge was in play – since the presence of the nudge will contribute to the decision made after de-biasing. Imagine a person who realizes the default retirement contribution is designed to encourage him to save more. He is not comfortable with this arrangement, and so spends some time thinking about how much he wants to save, and then changes his savings rate to the one that he believes works best for him.

The people who choose to work against the choice architect out of pettiness or vindictiveness lose out on both the potential benefits of the nudge, and the autonomy-promoting thought process involved in de-biasing. These people will be negatively affected by transparency.

The third concern – that transparency can reveal irrelevant or confusing information about the nudge – wins out over the other two in revealing the possible negative repercussions of transparency. I propose a specific means to implement transparency in nudging, which serves the purpose of improving ethicality with minimal effects on nudging efficiency.

A good way to promote transparency without sacrificing effectiveness is to inform people about the intention behind the nudge without revealing how the nudge works. Bruns, Kantorowicz-Reznichenko, Klement, Luistro Jonsson, & Rahali (2016) found that when people are informed about the intention behind a nudge designed to increase charity contributions, the effectiveness of the nudge actually increases. In their experiment, people were given 10 euros and asked how much money they would like to donate to charity, with a default option preselected. One of the transparency treatments revealed that the default was selected to encourage higher contributions to charity. The other transparency treatment simply revealed to people that the default may have an effect on their decisions. In other words, treatment 1 revealed the nudge and the intention behind it, while treatment 2 only revealed the nudge. The results showed that treatment 1 worked better than preselecting the same default with no information revealed (no transparency), and treatment 2 (perhaps predictably) worked worse (Bruns et al., 2016).

Transparency through revealed intentions without sacrificing effectiveness is likely to work best in situations where the nudged population can identify with the goal of the nudge. It is easy to improve effectiveness of a nudge by revealing that it is designed to promote pro social behavior. I provide multiple examples of scenarios where this kind of transparency can be implemented.

Studies have shown that different arrangements of food in a cafeteria can change eating behavior (Dayan & Bar-Hillel, 2011; Rozin, et al., 2011).

(Blogs, 2010)

The food can be arranged to achieve many different ends. These include welfare options that make most people better off (arrange for healthy eating), random arrangement, neutral choice (try to arrange based on what people would have picked themselves), and profit maximization (encourage people to buy the most expensive food).

Most people would agree that option 1 is a good combination of viable and reasonable. There is still a feeling of manipulating choice that may make some people queasy, but clearly the food has to be arranged in some manner. If we settle on option 1 as the best given the circumstances, we can implement intention-based transparency to increase the ethicality of the nudge. People could then be informed: "The food in this cafeteria has been arranged to encourage healthier eating."

Often, being transparent is an important part of the nudge itself. For example, "Smoking is injurious to health," displayed on a cigarette pack with a picture of a diseased lung, carries the implicit statement of "This pack has been designed to curb smoking."

In some buildings, nudges are implemented to encourage people to take the stairs instead of the elevator, either by informing them of the benefits of saving the environment, or by subtly indicating that taking the stairs is healthier. To improve transparency (and consequently ethicality), a sign could read, “This building has been designed to promote a healthier lifestyle”

The vein is similar to the “Smile, you’re on camera!” statements in buildings and elevators. They tell you they are watching, and you know it is for security reasons to ensure your safety – but there is no need to point out exactly where the cameras are. The cameras can be fake, incapable of preserving recorded footage, or being monitored by uninterested, half-asleep guards. The effect on behavior will be the same, and there is of course, no need to reveal that this is the case.

(Sticker 'Smile! You're on Camera', 2016)

When attempting to increase transparency to make nudging more ethical, it is important to remember that how the architect chooses to implement transparency will have an impact on the end objective. Statements like “You are being nudged” may be counterproductive, since not all people know what a nudge is, and they may be extremely concerned at their behavior being modified subconsciously or unconsciously.

A combination of a statement that confirms that a nudge exists, and an explicit statement of the intention behind it are sufficient for transparency. It is not necessary to point out exactly what the nudge is – doing this may significantly impact the effectiveness.

As discussed earlier in this section, sometimes the information revealed may do little to benefit the nudged population. It may even confuse or irritate them – accomplishing little, while risking a lot. This is why we focus on implementing transparency in pro-social contexts, and in a very specific manner. Revealing questionable intentions, even if they may be good for the nudged population, would cost far too much in effectiveness of the nudge. Additionally, revealing how the nudge works can confuse people and make them suspicious about sub-conscious behavior modification.

When implemented correctly, and in the right contexts, transparency can help ethicality both by itself, as well as by revealing good intentions. Intentions which, as we covered in the previous tenet, are crucial to ethical nudging.

Costs on Rational Agents and Redistribution of Resources

The hope when implementing a good nudge is that it can help irrational decision makers without unduly harming rational decision makers. It is easy to consider nudges that can have large (direct or indirect) costs on rational people. We know that people sometimes make hasty and irrational decisions when they are in 'hot' states, like excitement or sexual arousal (Kahneman D, 2003; Ariely & Loewenstein, 2006). A nudge that could help is to implement a cooling off period before the decision is executed. However, depending on the length of the cooling off period, the nudge could impose a serious cost on rational decision makers.

Stock markets across the world use "circuit breakers" to suspend trading activity if stock prices crash too much too soon. The by-laws define the actual parameters for implementing the circuit

breaker, and stock exchanges are associations whose bye-laws are created by and binding on all members. The intent of the circuit breaker is to provide a period of pausing and reflection when panic overtakes a market crash. However, rational players who are ready and willing to act are incapacitated and subjected to avoidable loss by the implementation of the circuit breaker, even if for a few minutes.

It is important that nudges do not punish people for being rational. To improve the ethicality of a nudge, one can refer to asymmetrically paternalistic policies as a benchmark. Asymmetric paternalism is a form of paternalism that addresses any costs (direct or indirect) borne by rational individuals from regulation aimed primarily at irrational individuals. “A regulation is asymmetrically paternalistic if it creates large benefits for those who make errors, while imposing little to no harm on those who are rational” (Camerer, Issacharoff, Loewenstein, O'donoghue, & Rabin, 2003).

Good examples of implementation of asymmetrically paternalistic nudges include nudges that modify choice architecture to draw attention to important information. The implication is that rational agents already possess this information, and irrational agents would benefit from noticing it. But some costs are harder to see than others.

Mitchell (2005) points out that nudges are frequently accompanied by redistribution of resources from rational to irrational persons, which is a cost on rational agents that is not always immediately apparent. He provides an example to illustrate this, based on a nudge designed to increase participation in employee pension plans.

> Employers confronted with increased participation must either redistribute funds among plan participants by reducing individual match amounts or infuse the plan with additional funds that may result in degradations of other employee benefits (alternatively, public employers could seek to pass the cost on to taxpayers, which may still have very marginal redistributive effects from the

rational to the irrational, or private employers may seek to pass the increased plan costs on to customers) (Mitchell, 2005).

These redistributive effects are important considerations when evaluating the ethical aspects of a nudge. From a utilitarian perspective, the redistribution may be small enough to ignore, or simply minimize and move on. Our ethical dilemma here centers around the fact that punishing rational decision makers is an inherently wrong practice, and justifying it takes more than simply attempting to minimize the costs. There is something to be said for not imposing large costs on rational people in the name of societal welfare. Especially when the rational people involved did not consent to being 'used' to improve irrational people's well-being.

I argue that redistribution of resources from the rational to the irrational can be justified depending on the context and intention behind the nudge. Aristotle's theory of justice (Sandel, 2010) claims that to identify what is ethical, we must understand the purpose, or *telos* of the social practice in question. Questions about what is ethical are, in part, questions about what virtues should be rewarded and honored (Sandel, 2010). It is not possible to argue that redistribution or costs on rational agents are ethical, without first understanding the purpose of the context in which the nudge is being implemented.

Imagine there are a finite number of flutes in the world, being distributed to the population. Aristotle would argue that the purpose of a flute is to be played, and that the ethical thing to do would be to give the flutes to the best flute players (Sandel, 2010). Nudging less competent players to pick up the flute would be an unethical form of redistribution, since it violates the purpose or *telos* of the flute. As would distributing the flutes to the richest players, or giving them out at random.

Consider, however, a nudge implemented to encourage the less privileged to take advantage of government benefits (Bhanot & Violante, 2016). The nudge simplifies the process that poor people have to go through to get the benefits they are entitled to, therefore increasing uptake of

government benefits. This nudge can be ethically justified as long as one presumes that one of the objectives of government resources (and in fact the creation of this benefit system) is to ensure a certain standard of living for the least privileged, even if it comes at the cost of redistribution of wealth (encouraging more people to take up benefits can either increase taxes, or reduce the resources available to the under privileged who did not need to be nudged to take advantage of the benefits available).

To say that costs of nudging on rational agents (whether direct or through redistribution) must be minimized, while true, is only a part of the discussion. Nudging in line with the purpose of the good or service involved helps offset the negative effects of 'using' rational decision makers; because any costs that are being imposed on them now exist only because they serve to advance the purpose of the service being used. Therefore, the ethicality of two similar nudges can depend a great deal on the context in which they are implemented. Promoting healthy eating through nudging is ethically justifiable in a cafeteria whose purpose is to encourage healthy eating. The same nudge, implemented in a cafeteria whose purpose is to provide food as cheaply as possible would not be ethical, from a redistribution perspective.

Multiple nudges can be implemented in any context, but not all of them align with the *telos* of the practice. Importantly, implementing a nudge that does not align with the *telos* does not change the purpose of the service. While this does not make that particular nudge inherently unethical, it may have a negative effect on ethicality if it imposes direct or indirect costs on rational agents. This would, of course, not be the case if the nudge aligned with the *telos*.

There is a distinction between the ethics of implementing a nudge in the context of the existence of the good, and the ethical nature of the good in question. The *telos* argument does not claim that the purpose of the good is inherently ethical or unethical (though of course it can be). It only deals with whether redistribution is ethical, given the purpose of the good in question. If society agrees that the purpose of public parks is to promote exercise, nudging people to go to

the park and run is ethical – not just because it improves societal welfare, but because it satisfies the purpose of the existence of the park. This is even though the nudge imposes a cost on rational people who were running in the park already, because the more crowded the park is, the more unpleasant it becomes to run there.

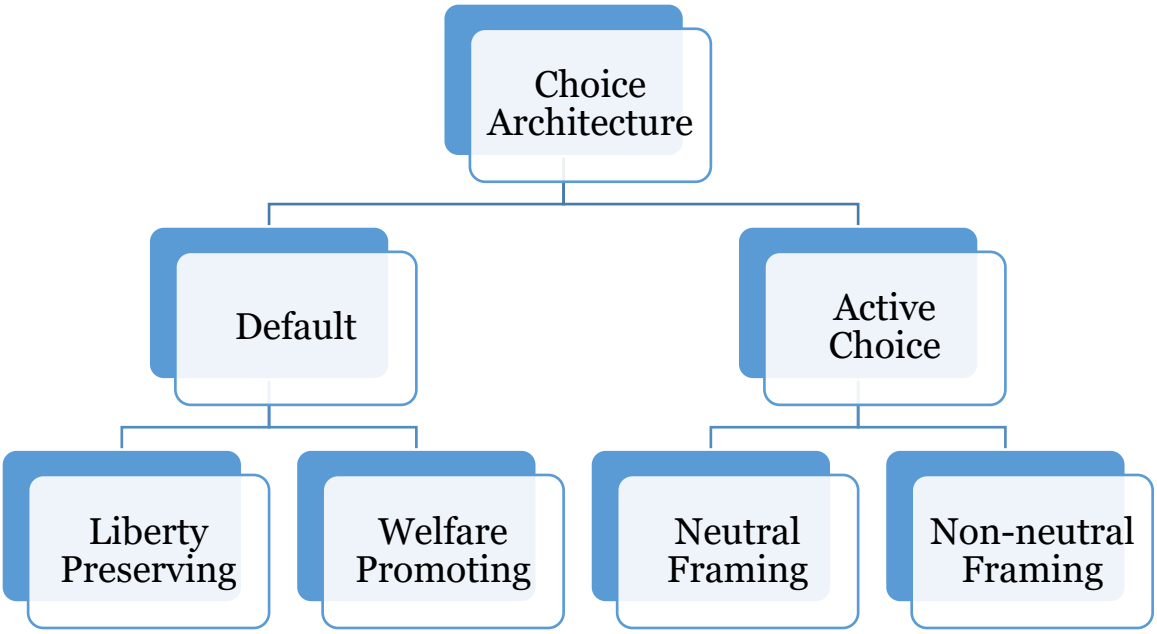
On the other hand, if we consider paying men to enter a ring and fight to the death a form of entertainment; nudging people to participate in the fight would still be an ethical form of redistribution. The 'rational' agents who were already fighting are doing so because they get paid money per fight, which in their head is worth the risk of death in the ring. It is possible that if more people begin participating, the rational agents get less money (which we assume is a finite resource). However, the more we nudge people to fight, the more entertaining it becomes for the public watching – which would indicate that the purpose of the good is being satisfied. The redistribution of resources is ethically justified in this case, but the inherent purpose of the good is clearly unethical. It's hard to argue that watching men fight and die for entertainment is ethical, much less something we as humans want for our society.

When considering the ethical aspects of implementing a nudge in the context of a good or service, understanding the purpose of the service concerned can help tremendously.

Ethical Implications of Defaults and Active Choice

To say that default options have a tendency to affect our decision-making would be akin to saying Michael Phelps has a tendency to swim well – it is a massive understatement. Setting defaults is an important decision, usually made by a government or central planner of some sort. An alternative to setting a default is to use active choice, which forgoes a default in favor of 'forcing' a choice about the matter at hand. For the sake of simplicity and to attempt to keep this

discussion more focused on real nudges and their implications, I focus specifically on the types of choice architecture outlined below.



We begin with setting ethical defaults, then proceed to ethical implementation of active choice.

Defaults

In determining ethical practices when defaults are involved, we must first decide whether setting a default is ethical in the decision making context.

In *The Art of Choosing*, Sheena Iyengar revealed that making choices is important to us as human beings. In fact, choice is an extremely powerful tool we use to mold and define ourselves (Iyengar, 2010). Schubert (2015) also claimed that active choice is autonomy promoting; that going through the process of preference formation by making active choices makes us more capable of navigating life and better human beings in general.

I agree. Defaults may erode the natural benefits we get from making active choices. There is, nonetheless, plenty of evidence that making too many decisions takes a toll on our cognitive ability, and can negatively impact autonomy as a consequence (Schubert, 2015). Our discussion then advances to finding the line between which choices are 'important' enough to be active and which ones could use welfare or liberty promoting defaults.

Schubert (2015) believes that active choice should be promoted where Rawls' primary goods are at stake. These are goods that, according to Rawls, every citizen has a right to obtain, and that every rational human being values (Rawls, 2009). They are:-

- The basic rights and liberties;
- Freedom of movement, and free choice among a wide range of occupations;
- The powers of offices and positions of responsibility;
- Income and wealth;
- The social bases of self-respect: the recognition by social institutions that gives citizens a sense of self-worth and the confidence to carry out their plans (Wenar & Zalta, 2013).

While these goods may seem all-encompassing and extremely restrictive of the opportunities to use defaults at all, there is some merit to Schubert's argument. You would not want defaults that determine who occupies a position of power (for example, defaulting your vote to the candidate your parents picked if you do not vote actively). Unfortunately, since cases can be made for setting defaults that affect income and wealth (retirement plan defaults), among other things, we are constrained to determining the ethics of default vs. active choice based specifically on the case at hand.

In the case where defaults have to be chosen by a central planner, Mitchell (2005) argues that a liberty preserving default is an ethical alternative to a welfare promoting default. Liberty

preservation promotes setting a default that "if chosen mindlessly, will be least restrictive of individual liberty, while leaving mindful individuals to opt out of the default option and enter into greater entaglements if they so choose" (Mitchell, 2005).

This would mean setting a default that preserves the most future options, given the default set today. For example, a liberty promoting default in the context of employment would be at-will employment – which allows employers and employees to exit an employment contract without needing a reason. Of course both employers and employees can choose to enter a contract that is more binding, as they often do. But the default is set so that irrational people who enter into employment contracts do not sacrifice future liberty through the default (Mitchell, 2005).

A welfare promoting default, by constrast, could over-generalise the preferences of decision makers. The Indian National Pension System maintains a default investment choice, which modifies the allocation between equity and debt using the subscriber's age as a marker. It can be faulted for generalising the risk preferences of subscribers, even while seeming to promote long-term welfare and old-age security of the subscribers.

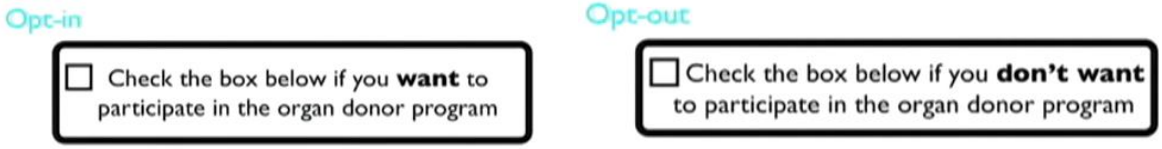
Cultural differences are a great way to showcase the need for casuistry in determining ethical defaults. Implementing the same default in different countries may have very different effects. This makes it extremely difficult to create ethical guidelines that will apply across nations, cultures, or even separate contexts within the same society.

I believe that utilitarianism can serve as guidance to determine the optimal default to set. The ethical default, in a case where we have determined that setting a default is preferable to promoting active choice, would be the option that maximizes societal utility. Given the limitations in measuring societal utility, and the attractiveness of liberty promotion as a cross-cultural solution (almost everyone likes liberty), it may seem that utility maximization is a dominated choice. I hope to offer some defence of utilitarianism in the context of default

settings through my subsequent example. I believe that the attempt to set ethical defaults begins with analyzing the ethics of the consequences of said defaults.

Liberty promotion can, of course, also be utility maximizing if the society involves places large value on liberty and gains utility from liberty promoting interventions. Any such co-existence is a happy accident, similar to Mitchell's example of a libertarian choice architect designing a cafeteria. Like the welfare maximizing paternalist, the libertarian would also place the most tempting food at the end, not because it promotes welfare, but because it enourages people to view all food options before making a decision (Mitchell, 2005). And now on to our example.

One of the most discussed implementations of the default nudge is the organ donation opt out. To determine voluntary consent to donate organs after death through a default setting, there are two possible systems. An opt-in, which means only those citizens who have given explicit consent are donors, and an opt-out, which means every citizen who has not refused is a donor. Changing the default choice to presumed consent (opt-out) has been shown to raise the amount of available organs by a large amount.



(Watanbe, 2014)

To determine the ethical practice in the domain of organ donation, we would first need to determine whether donation should be an active choice or be entitled to a default setting. I would argue that organ donation falls very firmly in the 'set a default' category.

Implementing active choice in organ donation begs the question 'when?' There is a significant difference between donation decisions made in sickness and in health. The apparent trend seems to be that people are passionate about making a decision to donate (or not donate) organs

only when they are ill or in the hospital. This is not to mention the active choice that the family and loved ones have to make immediately after the patient has just passed away. Neither of these situations is particularly conducive to rational decision-making. Encouraging people (and by extension family members) to make choices about organ donation when they are at their most vulnerable may not be beneficial or optimal.

If we decide that the active choice should be made when the individual is healthy (for example when they renew their driver's license), we must justify that organ donation is important enough to be an active choice. I believe it fails this test for two reasons. One, it does not qualify as a primary good, and two, it does not affect the decision maker's life (donation happens only in case of death). Active choice in organ donation therefore works very actively against utility maximization. While utility maximization is not the theory that determines ethicality here (we discussed its importance in determining which default to choose, not whether to implement one), recall that it is a salient consideration in policy regardless.

If we agree that organ donation fits into the 'set a default' category, we can move on to determining which default is the ethical one to select. Presumed consent (opt-out) has tremendous welfare benefits, while opt-in tends to be the safer practice that is less likely to be deemed unethical. The significant welfare benefits (more than 80 percent higher donation using presumed consent, across culturally similar countries) seem to indicate that presumed consent is the right move (Thaler & Sunstein, 2009).

Raihani (2013) postulated that presumed consent in organ donation qualifies as an ethical nudge, since the nudged individual gets to share in the collective benefit that the nudge provides. In the organ donation scenario, this would imply that nudging individuals toward donating organs will make more organs available in case of accidents, and the nudged individual gets to directly share in this benefit since he is more likely to obtain an organ during a time of need. This obviously translates into a positive effect on societal well-being as well.

Unfortunately, opt-out defaults are not always ethical, and have different consequences in developing nations. Nagral (2009) has pointed out that given the current state of health care in India, implementing an opt out system would do little other than provide rich and well-connected people an expanded donor base. This is primarily because the health care system in India is significantly less transparent and more prone to corruption. He also points out that given that India lacks affordable and universal health care, under-privileged individuals that are encouraged to donate organs do not actually get to participate in the collective benefit of the nudge. "The poor are also implored to donate but will not get organs when they need them. That, by itself, is a scandal but of course is not perceived as one" (Nagral, 2016).

The presumed social benefit from the opt-out default may thus yield a skewed social outcome, where the rich that can afford to pay for organs receive a plentiful supply that reduces their price, while the poor participate as donors but fail to receive the intended matching benefits. Technically, increasing organ availability for the rich while the poor receive no more or less organs than they did before is a pareto improvement. However, these consequences of setting a presumed consent default in this scenario are likely to be unethical in almost everyone's eyes. There is a tradeoff between the (quantifiable) higher number of organs available to the rich and the less quantifiable decrease in societal utility from the unequal consequences of the default.

In western nations with a transparent and universal health care system, however, the positive consequences of an opt-out default are easier to see. Any small decrease in societal utility can be countered with the massive increase in utility from lives saved. Western societies also tend to be more open about having a political discussion about the implications of presumed consent, making it easier to have a conversation to 'arrive' at presumed consent as the right answer, instead of forcing it on society.

I hope the organ donation example served to highlight that the choice between defaults, whether liberty promoting, welfare promoting, or any other, is sensitive to concerns that we cannot hope

to encompass through overreaching ethical guidelines. Governments and central planners need to determine a usable counterfactual to understand the ethical implications of default settings using societal utilitarianism as a guideline.

There are, of course, people who will agree that a welfare promoting default can increase utility but still be unethical. They are not wrong. However, an attempt to use Kantian ethics or a similar framework here would judge the ethicality of default setting based on intrinsic rules, while ignoring the consequences of the default setting chosen. I do not believe governments and central planners can afford to think like this. We arrive at the decision to set a particular default only after judging that setting a default is the ethical practice in the first place. The area for rule-based ethics was in the first decision, to make sure that implementing a default cannot affect basic rights like freedom of speech. Once we have arrived at the consensus that setting a default is the right way to go, welfare promotion and utility maximization should be at the forefront in making the selection.

Active Choice

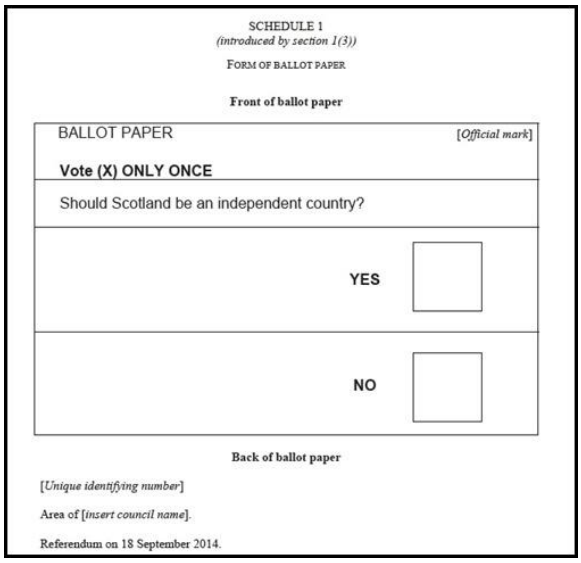
In contexts where defaults should not be set, where making a choice is either unavoidable (choosing food in a cafeteria), or clearly the ethical practice (voting in a democracy), the choice architect is again confronted with two options - a neutral frame, or a non-neutral frame. Neutral framing promotes choice architecture that displays information in the most unbiased manner possible. Non-neutral framing promotes choice architecture that consciously works to nudge the decision maker toward an option, be it welfare promoting, liberty promoting, or even profit promoting.

Enhanced active choice, a type of choice architecture that forces people to make a decision without a default option, but highlights one of the available options as the ‘good’ or ‘right’ one, is

an example of non-neutral framing. Active choice without the highlight is an example of neutral framing.

When considering the benefits and viability of neutral framing, two recent examples come to mind regarding referendums held in European countries. In 2014, Scotland voted on whether to stay in, or leave the UK. The question on the ballot was phrased “Should Scotland be an independent country?” with the options “Yes” and “No” (Murray, Treanor, Chan, & Martin, 2013). As we know now, the “No” vote won, and Scotland stayed in the UK. However, the choice architecture in the question is far from neutral. ‘Independence’ carries a certain empowering notion to it, and it is entirely possible that the framing of the question affected votes in a very real way (Rajda, 2016).

The referendum on Britain leaving the EU (or Brexit, as it is affectionately called) was a different story. Care was taken to ensure that the choice architecture was as neutral as possible, to prevent any bias (Saiidi, 2016). The question on the ballot was “Should the United Kingdom remain a member of the European Union or leave the European Union?” The options were "Remain a member of the European Union" or "Leave the European Union” (Watt & Syal, 2015). The contrast is clear to see, and regardless of which side of Brexit one supports, the effort to maintain neutrality is laudable. I would even go so far as to say that regardless of whether the outcome was welfare promoting or not, the Brexit referendum frame was more ethical than the one used in the Scottish referendum.



SCHEDULE 1
(introduced by section 1(3))
FORM OF BALLOT PAPER

Front of ballot paper

BALLOT PAPER [*Official mark*]

Vote (X) ONLY ONCE

Should Scotland be an independent country?

YES ☐

NO ☐

Back of ballot paper

[*Unique identifying number*]

Area of [*insert council name*].

Referendum on 18 September 2014.

Referendum on the United Kingdom's membership of the European Union

Vote only once by putting a cross ✗ in the box next to your choice

Should the United Kingdom remain a member of the European Union or leave the European Union?

Remain a member of the European Union ☐

Leave the European Union ☐

Scottish Referendum (Taylor, 2014) UK Referendum (Heffer, 2016)

For important votes like this one, non-neutral framing is a difficult choice to defend from an ethical perspective. A bias would exist, for example, in the Brexit vote, because the status quo is that Britain stays in the EU. Leaving the EU is a prospect filled with uncertainty, and most people are naturally averse to uncertainty (a bias known as ambiguity aversion). Conscious effort could potentially be made to use non-neutral framing to de-bias the voting population, either by informing them that they are prone to ambiguity aversion and the status quo bias, or by changing the architecture of the ballot to counter the bias. I have no doubt that either of these measures, if implemented by electoral commission, would be widely protested and justifiable deemed unethical. There are other times, however, when non-neutral framing is the ethical thing to do.

When we consider the benefits of neutral choice architecture, ethicality is primary among them. It seldom helps decision making to frame things as neutrally as possible. Sometimes, it may not even help to simplify the decision. All it really does is ensure that our intentions are in the right place, that we respect people as rational beings, and do not interfere (even unintentionally) with their autonomy. Kant would be proud.

I would argue that decisions about when to utilize non-neutral framing should be made keeping Rawls' publicity principle in mind. In its simplest form, Rawls' publicity principle states that a government should be banned from implementing policy that it is unable or unwilling to defend publicly to its own citizens (Wenar & Zalta, 2013). Thaler and Sunstein bring this principle up as a guideline to ethical nudging themselves (Thaler & Sunstein, 2009).

In the context of active choice and non-neutral choice architecture, I believe this principle is particularly salient. Its application would imply that implementing non-neutral framing in the context of active choice is ethical only in situations where the choice architect can defend his decision to the nudged public. This does not imply that the government should be consistently defending its non-neutral choice architecture to its citizens, many of whom may not be familiar with nudging as a concept. The idea is that the choice architect should be capable of defending their decision to nudge (with a non-neutral framework), if they had to do so.

Whereas the idea of what a government could defend to its citizens is subjective, the publicity principle provides a much needed line between architecture that must clearly be neutral (voting ballots), and that can be justifiably biased (retirement savings). It is admittedly a thick line – there is definitely policy that can be implemented in that intermediate gray area. Nevertheless, I believe the publicity principle works well to promote ethicality in the context of active choice.

Chapter 4 - Ethical Nudging in Action

As part of a research seminar during my master's degree, I worked with a group of students from both Erasmus University and Harvard Business School on a project for the Dutch Ministry of Finance and Pension Federation (*pensioenfederatie*). We were tasked with designing an online tool that would represent people's pensions to them accurately, while maintaining the trust that they had in the Dutch pension system.

With pension plans shifting from defined benefit to defined contribution, the amount of money people were going to receive as pension was both lower than they might expect, and more uncertain than they might expect. We quickly realized that a possible way to counter this would be to 'anchor' people to the lower values – a technique that would de-bias them and encourage them to both confront the reality of their situation and save more for retirement.

Our other option was a neutral framework – one that represented the possible pension outcomes without emphasizing any, or exploiting any biases. It was designed to aid comprehension, not directly or systematically influence behavior. We decided to use the anchoring framework.

From a utilitarian perspective, non-neutral framing was the right move. It promoted both individual and societal welfare, and accomplished the objective that we set out to achieve. Unfortunately, it didn't help us to realize this. We were left feeling queasy about it for weeks after, wondering if presenting the options more neutrally and taking the hit on effectiveness on the nudge would have been worth it.

The pension platform was organized to enforce active choice (a decision that was made for us by our client). If prompted, any member of our team would be more than happy to defend our non-neutral architecture to any nudged individual – in line with Rawls' publicity principle. Our intentions were set in stone through the project – to aid comprehension and trust in the system, and increase savings; both clearly beneficial for the user. We had no rational agent cost or redistribution concerns, since the pension plan was defined contribution. Each user's savings only affected their own payout. We were not transparent about our intentions, however, since we were too concerned that indicating the purpose of the nudge would result in a sacrifice on effectiveness. Perhaps we could have afforded to be more transparent, but 'designed to aid your comprehension' serves no real purpose, and 'designed to improve your trust in the system' would instantly make people suspicious.

All said, I believe we nudged ethically. I hope that having read this thesis; you can make your own judgements about my claim.

May ethical nudging be ever salient in our quest toward becoming a better society.

References

(n.d.).

Ariely, D., & Loewenstein, G. (2006). The heat of the moment: The effect of sexual arousal on sexual decision making. *Journal of Behavioral Decision Making*, 19(2), 87-98.

Baude, W. (2014, February 16). *Glenn Beck, Sarah Palin, and others on Cass Sunstein*. Retrieved from Washington Post Web site: https://www.washingtonpost.com/news/volokh-conspiracy/wp/2014/02/16/glenn-beck-sarah-palin-and-others-on-cass-sunstein/

Bentham, J. (1879). *An introduction to the principles of morals and legislation*. Clarendon Press.

Bhanot, S., & Violante, A. (2016, January 20). *Why Don't People Take Free Cash?* Retrieved from Misbehaving: http://www.misbehavingbook.org/blog/2016/1/20/why-dont-people-take-free-cash

Bhargava, S., & Loewenstein, G. (2015). Behavioral economics and public policy 102: Beyond nudging. *American Economic Review*, 396-401.

Blogs, M. B. (2010, March 7). *Choice Architechture at the kwik-e-mart*. Retrieved from Nudge Blog: http://nudges.org/tag/homer-simpson/

Boaz, D. (2016). *Libertarianism*. Retrieved from Encyclopædia Britannica: https://www.britannica.com/topic/libertarianism-politics

Brown, Z. (2013, November 9). *Kant*. Retrieved from Stickmen with Martinis: https://stickmenwithmartinis.com/category/kant/

Bruns, H., Kantorowicz-Reznichenko, E., Klement, K., Luistro Jonsson, M., & Rahali, B. (2016). Can Nudges Be Transparent and Yet Effective? *Social Science Research Network*, 2816227.

Busse, M. R., Pope, D. G., Pope, J. C., & Silva-Risso, J. (2012). Projection bias in the car and housing markets. *National Bureau of Economic Research.*, w18212.

Calhoun, S. (2012, April 29). *SquareOne Explorations*. Retrieved from SquareOne: http://squareone-learning.com/blog/2012/04/

Camerer, C., Issacharoff, S., Loewenstein, G., O'donoghue, T., & Rabin, M. (2003). Regulation for Conservatives: Behavioral Economics and the Case for" Asymmetric Paternalism". *University of Pennsylvania law review*, 151(3), 1211-1254.

Dayan, E., & Bar-Hillel, M. (2011). Nudge to nobesity II: Menu positions influence food orders. *Judgment and Decision Making*, 6(4), 333.

Dworkin, G. (2016, June 21). *Paternalism*. Retrieved from Stanford Encyclopedia of Philosophy (Summer 2016 Edition): http://plato.stanford.edu/archives/sum2016/entries/paternalism/

Fischer, M., & Lotz, S. (2014). Is Soft Paternalism Ethically Legitimate?-The Relevance of Psychological Processes for the Assessment of Nudge-Based Policies . *Cologne Graduate School in Management, Economics and Social Sciences.*, No. 05-02.

Heffer, G. (2016, Jan 27). *REVEALED: Here's what the EU referendum ballot paper will look like - ready for Brexit?* Retrieved from Sunday Express: http://www.express.co.uk/news/politics/638210/EU-referendum-ballot-paper-Brexit-vote-June-23

Iyengar, S. (2010). *The Art of Choosing*. London: Little, Brown.

Kahneman, D. (2003). Maps of bounded rationality: Psychology for behavioral economics. *The American economic review*, 93(5); 1449-1475.

Kahneman, D. (2011). *Thinking, fast and slow*. New York: Farrar, Straus and Giroux.

Kant, I., Wood, A. W., & Schneewind, J. B. (2002). *Groundwork for the Metaphysics of Morals*. . Yale University Press.

Loewenstein, G., Bryce, C., Hagmann, D., & Rajpal, S. (2014). Warning: You are about to be nudged. *Behavioral Science and Policy*, 35-42.

Mill, J. S. (1966). *On liberty*. *In A Selection of his Works*. London: Macmillan Education UK.

Mitchell, G. (2005). Libertarian paternalism is an oxymoron. *Northwestern University Law Review*, 99(3).

Murray, L., Treanor, S., Chan, V., & Martin, C. (2013, January 24). *Testing of the Proposed Question for the Referendum on Scottish Independence.* Retrieved from The Electoral Commission: http://www.electoralcommission.org.uk/__data/assets/pdf_file/0005/153689/Ipsos-MORI-Scotland-question-testing-report-24-January-2013.pdf

Myers, D. (2014). *Exploring Social Psychology; 7 edition*. New York: McGraw-Hill Education.

Nagral, S. (2009). Will presumed consent make transplantation accessible, ethical and affordable in India? *Indian Journal of Medical Ethics*, 6(3).

Nagral, S. (2016, August 3). *Better buy than die? The unfortunate enduring saga of organ sales in India*. Retrieved from Scroll.in: http://scroll.in/pulse/812795/better-buy-than-die-the-unfortunate-enduring-saga-of-organ-sales-in-india

Raihani, N. J. (2013). Nudge politics: efficacy and ethics. *Frontiers in psychology*, 4.

Rajda, V. (2016, March 7). *To B or not to B, that is the question*. Retrieved from FinalMile | Behaviour Architecture: http://finalmile.in/behaviourarchitecture/to-b-or-not-to-b-that-is-the-question

Rawls, J. (2009). *A theory of justice*. Harvard university press.

Roese, N. J., & Vohs, K. D. (2012). Hindsight bias. *Perspectives on Psychological Science*, 7(5), 411-426.

Rozin, P., Scott, S., Dingley, M., Urbanek, J. K., Jiang, H., & Kaltenbach, M. (2011). Nudge to nobesity I: Minor changes in accessibility decrease food intake. *Judgment and Decision Making*, 6(4) 323.

Saiidi, U. (2016, June 22). *The Brexit ballot wording wasn't always so simple*. Retrieved from CNBC : http://www.cnbc.com/2016/06/22/brexit-question-ballot-wording-framing.html

Sandel, M. J. (2010). *Justice: What's the Right Thing to Do?* New York: Farrar, Straus and Giroux.

Schnellenbach, J. (2012). Nudges and norms: On the political economy of soft paternalism. *European Journal of Political Economy*, 28(2), 266-277.

Schubert, C. (2015). On the ethics of public nudging: Autonomy and agency. *Social Science Research Network*, 2672970.

Selinger, E., & Whyte, K. (2011). Is There a Right Way to Nudge? The Practice and Ethics of Choice Architecture. *Sociology Compass*, 923-935.

Spielberg, S. (Director). (1993). *Jurassic Park* [Motion Picture].

Sticker 'Smile! You're on Camera'. (2016, 1 August). Retrieved from Foscam: http://www.foscam.nl/index.php/eusticksmile.html

Sunstein, C. (2014, November 29). *The Ethics of Nudging*. Retrieved from Social Science Research Network: http://ssrn.com/abstract=2526341

Sunstein, C. R., & Thaler, R. H. (2003). Libertarian paternalism is not an oxymoron. *The University of Chicago Law Review*, 1159-1202.

Taylor, S. L. (2014, September 18). *Today in Comparative Ballots ("Independence or not" Edition)*. Retrieved from Outside the Beltway: http://www.outsidethebeltway.com/today-in-comparative-ballots-independence-or-not-edition/

Thaler, R. H. (1988). Anomalies: The ultimatum game. *The Journal of Economic Perspectives*, 2(4), 195-206.

Thaler, R. H. (2015, October 15). *The Power of Nudges, for Good and Bad*. Retrieved from The New York Times: http://www.nytimes.com/2015/11/01/upshot/the-power-of-nudges-for-good-and-bad.html

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

Thaler, R. H., Sunstein, C. R., & Balz, J. P. (2014). Choice architecture. *The behavioral foundations of public policy*.

Tversky, A., Slovic, P., & Kahneman, D. (1990). The causes of preference reversal. *The American Economic Review*, 204-217.

Watanbe, B. (2014, March 27). *Lessons on U.X. from Louis C.K.* Retrieved from A Medium Corporation Web Site: https://medium.com/learning-startups-stuff-from-other-stuff/lessons-on-u-x-from-louis-c-k-21b9390ddba6#.28n0pe8fm

Watt, N., & Syal, R. (2015, September 1). EU referendum: Cameron accepts advice to change wording of question. *The Guardian*. Retrieved from The Guardian: http://www.theguardian.com/politics/2015/sep/01/eu-referendum-cameron-urged-to-change-wording-of-preferred-question

Wenar, L., & Zalta, E. N. (2013, December 21). *John Rawls*. Retrieved from The Stanford Encyclopedia of Philosophy (Winter 2013 Edition): http://plato.stanford.edu/archives/win2013/entries/rawls/