<u>MODERN MILL: AN EXAMINATION OF ECONOMIC NUDGES AND MILL'S HARM PRINCIPLE</u>

by

Clayton Dana-Bashian

Submitted in partial fulfillment of the

requirements for Departmental Honors in

the Department of Political Science

Texas Christian University

Fort Worth, Texas

May 4, 2020

MODERN MILL: AN EXAMINATION OF ECONOMIC

NUDGES AND MILL'S HARM PRINCIPLE

Project Approved:

Supervising Professor: Dr. Samuel Arnold,

Department of Political Science

Dr. John Harris,

Department of Philosophy

Dr. John Harvey,

Department of Economics

ABSTRACT

John Stuart Mill purported to demarcate the limits of government power in his work *On Liberty*, which outlined The Harm Principle, the assertion that the government may only step in to limit someone's liberty when they would harm another. Since the time of its writing, Mill's Harm Principle has remained a powerful force in describing how governments may operate. One fundamental premise of this view is that individuals are capable of satisfying their stated preferences without government intervention. However, the work of behavioral economists Drs. Richard Thaler and Cass Sunstein show that individuals often fail to appropriately satisfy their own stated preferences. To correct for this, Thaler and Sunstein theorize a system of 'nudges' that change the context in which individuals make decisions in order to satisfy their stated preferences. The central question of this paper is whether Mill's Harm Principle can accommodate Thaler and Sunstein's nudges and if so, how. This paper will answer in the affirmative and defend a system that filters nudges through the various institutions that individuals are a part of, ending with a list of suggested and permissible nudges as well as a response to objections.

Introduction

It is relatively uncontroversial to say to a layperson that all individuals should be allowed to set and satisfy their own preferences. While a number of political philosophers would disagree with this, John Stuart Mill would wholeheartedly endorse that assertion. His hallmark contribution to political theory, the Harm Principle, asserts that the only time authority is justified in intervening in an individual's ability to set and satisfy their preferences is when that person would harm another. On the other hand, the theorists Richard Thaler and Cass Sunstein are wary of the statement, not because they disagree with allowing others to satisfy their own preferences, but because they are not convinced that people know how. These theorists have made significant contributions to the field of behavioral economics by developing a system of 'nudges' – actions that change the context in which individuals make decisions in order to help them better satisfy their preferences.

Despite the surface disagreement between Mill and Thaler/Sunstein, the two theories perhaps align in meaningful and instructive ways. A treatment of the two theories that puts them in conversation with one another, in order to discover the relationship between them, may have something meaningful to contribute to political theory. However, merely describing the relationship between the two theories is not particularly instructive in understanding what their implications are for the proper exercise of political power. The interplay between the two theories does, in fact, illuminate a particular role government may play in improving the lives of its citizens – namely that it is permissible for governments to nudge their citizens towards outcomes they would prefer. Exactly how instructive the interplay between the Harm Principle and nudges is the first question this paper will explore, followed by a survey of the ways in which governments can appropriately utilize nudges (as informed by Mill) in order to help their citizens lead better lives, as judged by the citizens themselves.

Review of Relevant Literature

Much of political philosophy is driven by the question of what governments have the legitimate right to force upon their citizens. John Stuart Mill, a nineteenth century thinker who is most famous for his work in helping establish the philosophical movement known as Utilitarianism, is concerned with this question in his work *On Liberty*. Utilitarianism, in short, is the theory that people should act in such a way that the consequences of those actions maximize pleasure (aka the good) and minimize pain (the bad). The theory turns on the so-called 'hedonistic calculus' that measures the outcomes of actions and assesses if they maximize the good. If so, Utilitarianism says that this action is moral. In short, Utiliarians believe that if everyone only ever acted in ways that conformed to this principle then the world would be the best it could be, because there would be the most positive and the least negative.

Mill translates this philosophy to government power in his work *On Liberty*. Mill's assertion in *On Liberty* is stark and powerful –

> 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. There are good reasons for remonstrating… persuading… or entreating him, but not for compelling him. To justify [compulsion], the conduct which it is desired to deter him from be calculated to produce evil to some one else.[1]

This statement, which has since been dubbed the Harm Principle, uses individual preferences and decision-making in order to translate Utilitarianism into a principle that can motivate government

---

[1] Dworkin, Gerald. *Morality, Harm, and the Law*. (Westview Press, 1994). Page 9.

action. Key in Mill's determination of the proper use of political power is the question of who should be sovereign in society.

Mill preserves the sovereignty of the individual in his elucidation of the Harm Principle, asserting "over himself, over [their] own body and mind, the individual is sovereign."[2] He does so because of a fundamental assumption of the hedonistic calculus – if someone is doing something, it must be because they enjoy it (i.e. it produces pleasure), and as long as no one else's sovereignty is implicated in the actions, the government should leave well enough alone. Otherwise interfering in an individual's decision-making would surely result in less enjoyable activities overall. This particular assumption of Mill's is particularly susceptible to criticism, as we will see later in this essay.

Nevertheless, Mill's Harm Principle has become one of the pillars of political philosophy. The common political consciousness, both implicitly and explicitly, accepts the fundamental constraint of Mill's view; while people and political parties may quibble over particular policy instantiations, a significant number of individuals labor under the assumption that the government should only step in when individuals would harm one another. To this end, then, it is necessary to examine the larger context of Mill's claim about harm and the implications it has for permissible government activity.

Moreover, it is important to note that the outcomes of one's actions are never truly contained to the actor themself: others may experience positive or negative feelings as a result of an individual's conduct. The quintessential example of such action is suicide – one's family and friends clearly experience serious emotional and mental anguish from an individual's decision to

[2] Ibid, 9.

commit suicide. Additionally, decisions to be unhealthy may indirectly increase a society's healthcare burden, personal energy usage may contribute to climate change, and so on. While some of these outcome costs may seem incredibly diffuse, nonetheless they are still external and societal costs that are levied as a result of individual action. However, Mill explicitly rejects these considerations, saying "when I say only [themself], I mean directly, and in the first instance… the objection which may be grounded on this contingency, will receive consideration in the sequel."[3] That is to say, Mill thinks these sort of external considerations are important and that a moral individual should consider them. However, they will not rise to the level of superseding an individual's preferences and certainly not to the level of government action.

Even so, Mill is not silent on the subject of individual preferences and their limits. In fact, he believes that *society* (not just government) can rightly place limits on individual conduct. It is important to note that societal and governmental interference constitute two separate levels of force (this is the difference between socialization and coercion, for example). Mill believes this is borne out in terms of reputational costs, saying:

> A person who shows rashness, obstinacy, self-conceit – who can't live within moderate means… must expect to be lowered in the opinion of others, and to have a less share of their favorable sentiments, but of this he has no right to complain.[4]

So Mill says that people may rightfully be judged by others, scorned for social transgressions, and the like. This is a perfectly acceptable social mechanism by which individuals may be judged and exhorted to better conduct. Mill is careful to distinguish these social choices from governmental coercion, saying that when "the evil consequences of [their] act fall on…society…

[3] Ibid, 11.
[4] Ibid, 14.

[it] must retaliate on him" and further that "it is not our part to inflict any suffering on him" (in reference to an individual inflicting harm).[5]

Finally, Mill devotes considerable space to the refutation of a competing conception of government power – legal moralism. Briefly, legal moralism is the theory that the government should use coercive means to criminalize and disincentivize behavior that it would consider morally reprehensible, as opposed to merely harmful. In fact, many governments have, now or in the past, enacted legal moralistic statutes in spades. For example, statutes that criminalize drug use, ban alcohol, or ban homosexuality are based in a legal moralistic principle. To Mill, these activities are all based upon an individual's preferences and have primary effects limited only to the actors. As such, Mill would strongly object to these laws – they do not rise to the level of interfering with another's exercise of their own liberty and thus cannot be seen as permissible interference by the government.

Mill couches his argument against legal moralism in terms of personal taste. He calls the legal moralistic principle "social rights", saying those who are proponents of them say "whosoever fails thereof in the smallest particular, violates my social right, and entitles me to demand from the legislature the removal of the grievance."[6] However, this "monstrous a principle is far more dangerous than any single interference with liberty; there is no violation of liberty which it would not justify."[7] Mill believes that we have an absolute right to our personal tastes – if an individual believes that intemperance, homosexuality, and/or drug use is immoral, then that individual need not engage in that behavior. However, it does not follow from this that one has a right to will that behavior or forbearance from any other person. Our own particular

---

[5] Ibid, 15-16.
[6] Ibid, 23.
[7] Ibid.

moral "taste is as much [our] own peculiar concern" and that any attempt to impose "the opinion of a similar majority, imposed as a law on the minority… is quite as likely to be wrong as right."[8] The government has no right to tell people what they should or should not be concerned with in their own personal tastes, because the government cannot rightfully tell us which of our tastes is correct. As such, legal moralism is defeated.

Mill's utilitarian principle is a significantly powerful tool in determining the limits of governance. Harm seems to be a clear delineation of the limits of governance – it is intuitively obvious that we have no right to harm others. However, Mill's principle of sovereignty goes even further than this, asserting not just that we have no right to harm one another, but that coercive means can *only* be used when individuals would harm one another. That is to say, people deserve autonomy both in their personage and the ability to pursue their own ends. On top of this, they do not have the right to tell other people how to live their lives and mediate their personal tastes, at least not through coercive means. While the jump between the lack of a right to harm others and the lack of the right to tell other how to live their lives is important, it is one Mill is happy to make, saying an individual "cannot rightfully be compelled… because it will be better for [them] to do so, because it will make [them] happier… [if] to do so would be wise, or even right."[9] This principle leaves some gaps for interpretation, particularly for a theory of nudges, which will be addressed in subsequent sections (notably along the lines of a non-coercive regime of nudges). At present, Mill is committed solely to only interfering with conduct which would harm another.

[8] Ibid, 19.
[9] Ibid, 9.

Mill's Harm Principle, then, seems to be an obvious principle to limit government intervention – the government (a coercive entity) should intervene only in those cases where someone would harm another. However, as we will see, this principle cannot provide direction for every course of conduct, even by Mill's own admission, and conduct whose coerciveness is debatable (such as providing information which is likely to change an individual's mind) is hard to quantify under the Harm Principle. There are behaviors that individuals and governments are morally permitted to engage in which are not explicitly directed by the Harm Principle. It prevents governments from interfering coercively except in cases of harm to others. But, it does not explicitly make determinations regarding government actions that are not coercive. So, coercive government actions against murder, rape, etc. fall directly out of the Harm Principle. However, government actions which are not coercive may yet be compatible with the Harm Principle even while they are not specifically endorsed by the theory. As we shall see, nudges are one such class of actions. They are not coercive, so the Harm Principle may readily accept their veracity within its framework.

In addition to the Harm Principle's intrinsic limitations, a more advanced understanding of human behavior has exposed further issues with Mill's justifications. Since Mill's work in the nineteenth century, a significant body of psychological research has advanced our understanding of the human psyche and why, in fact, some of Mill's conclusions seem to be false.[10] In particular, Mill seems to think that taste is an unassailable right that each individual has. Mill says governments cannot properly determine what an individual should or should not be doing

---

[10] For examples of such research, see:
Airely, Dan. *Predictably Irrational*. (HarperCollins, 2008).
Camerer, Colin, George Lowenstein, and Matthew Rabin, eds. *Advances in Behavioral Economics*. (Princeton University Press, 2004).
Kahneman, Daniel. *Thinking, Fast and Slow*. (Farrar, Straus and Giroux, 2011).

because people are sovereign in their ability to make choices. One important assumption in Mill's philosophy, then, is that individuals the actions that would be best for their own happiness. However, a body of research known as behavioral economics has shown that humanity predictably, systematically, fails to appropriately determine not only what's best for them, but what they actually want.

Drs. Richard Thaler and Cass Sunstein are pioneers of this movement and its determinants on political power. They describe themselves as 'libertarian paternalists', a school of political philosophy that uses government influence to allow people to better choose, based off their own preferences, while still retaining their absolute freedom to choose.[11] Libertarian paternalism is a system that seeks to recognize the reality of humanity's fundamental failures of logic and provide people the ability to more appropriately make decisions that align with their actual, enumerated preferences. Thaler and Sunstein believe that the libertarian paternalist conception of governmental authority can be implemented through a system of so-called nudges.[12]

Nudges are an aspect of choice architecture. Thaler and Sunstein define choice architecture as "the context in which people make decisions."[13] Obviously, this is a broad definition and rightfully so. Humans constantly make decisions in a number of different circumstances. Moreover, humans can only make decisions in circumstances. That is to say, people make decisions based on how things around them are organized – picking a vegetable from the top of the pile at the supermarket instead of digging through the entire pile, for

[11] Sunstein, Cass and Richard Thaler. *Nudge: Improving Decisions about Health, Wealth, and Happiness*. (Yale University Press, 2008). Page 5.
[12] Ibid, 6.
[13] Ibid, 3.

example. Thaler and Sunstein's view recognizes the inevitability that "everything matters" in terms of decision-making.[14] As such, they believe the best way to create a choice architecture that works for everyone is a libertarian paternalist principle, working within Mill's Harm Principle framework and opposing legal moralism.

Nudges themselves are the central feature of choice architecture. According to Thaler and Sunstein, nudges can be defined 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.[15]

There are a number of important pieces of this definition. Firstly, a nudge must systematically and predictably alter decisions. If a particular aspect of the choice architecture cannot have a consistent impact, its implementation cannot properly be called a nudge. Secondly, nudges are not requirements, they merely make it significantly easier (by accounting for human psychology) for people to make the choices they actually want to.[16] Thirdly, it is the case that nearly everything is a nudge in one way or another; all of our decisions are made within choice structures and nudges are a necessary feature of choice structures. Directed nudges acknowledge this fact and alter the choice architecture. Finally, their definition of nudge has something significant to say about economic incentives. Thaler and Sunstein do not want to allow any behavior that alters economic incentives in such a way as to actively encourage people to take a particular route of the architect's design. A fundamental tenant of their libertarian-paternalistic approach to nudges is that it is "liberty preserving" because "[l]ibertarian paternalists wants to make it easy for people to go their own way."[17] So while they could choose a more intrusive

---

[14] Ibid, 2.
[15] Ibid, 6.
[16] Ibid.
[17] Ibid, 5.

version of their theory, Thaler and Sunstein consciously want to remain as non-intrusive as possible, a point with which any good Millian can agree.

Remaining truly libertarian in the paternalism of nudges is the most difficult condition of nudging to satisfy: because everything matters when making a decision, it can be nearly impossible to tell if the choice architecture is actually helping people realize their tastes or instead leading them inexorably to choose something different. That being said, Thaler and Sunstein believe that there is a coherent statement to be made about nudges and that, in fact, it is possible for the government to act as choice architects and improve the lives of their citizens.

However, in order to appropriately determine when nudging can be effectual and beneficial to people, it is first necessary to determine the circumstances in which humans fail to make good decisions. These circumstances are wide-ranging and too numerous to fully list herein, but recounting a few of Thaler and Sunstein's most egregious offenders is instructive. They identify a number of heuristics to this effect, psychological rules of thumb we use to help make decisions and save cognitive real estate in a world that is constantly barraging us with information.[18] While extremely useful, heuristics nonetheless can form biases that may prevent people from accurately assessing information and making good decisions. For example, the availability heuristic explains why people are more afraid of terrorist attacks than they are of skin cancer, even though the latter is far more likely to kill them.[19] Other forms of systematic biases include a lack of self-control, the desire to fit in and follow a social 'herd', and the difficulty of making decisions with diffuse costs over a long time horizon.[20] By understanding the wide ranging and predictable ways in which human beings err, government choice architects can

[18] Ibid, 22-23.
[19] Ibid, 25.
[20] Ibid, 44, 54, 75.

create systematic nudges that correct these psychological errors. Thaler and Sunstein identify the best circumstances in libertarian paternalist nudges can be utilized to improve peoples' lives, saying:

> People will need nudges for decisions that are difficult and rare, for which they do not get prompt feedback, and when they have trouble translating aspects of the situation into terms that they can easily understand.[21]

These conditions are most easily understood through an example. Thaler and Sunstein identify a clear case that satisfies these conditions – mortgages. In the modern lending market, consumers are offered a smorgasbord of plans at differing lengths, interest rates, and payment plans.[22] Many people choose one mortgage over their lifetime, are rarely or never given the opportunity to choose a new one if the original turns sour, and mortgages require the understanding of a number of complex social and economics concepts. These factors combine to create a situation in which humans are not able to rely on their rational decision-making processes, instead falling back on the instinctive heuristics which can lead them astray.

The important thing to note here is that Thaler and Sunstein are making an argument that directly refutes how Mill grounds his Harm Principle. Nudge theory states, explicitly, that individuals are not good at choosing how best to maximize their own utility.[23] As discussed previously, Mill makes a fundamental assumption in developing the Harm Principle – that individuals are actually able to make decisions in their best interest. So, if Thaler and Sunstein are right, Mill's argument is incomplete. Mill's assertion that we should defend the Harm

---

[21] Ibid, 74.
[22] Ibid, 135.
[23] It is critical to understand here that Thaler and Sunstein do not doubt that individuals can set their preferences in a way that will make them happiest. Rather, they are asserting that individuals can fail to satisfy those preferences once they are set. Nudges are not for telling us how to set out preferences, they are helping us satisfy them.

Principle via an argument of individual sovereignty is deficient in some very real respect, and nudges arguably offer a better way in which we can satisfy the Utilitarian principle.

Nudges offer a more nuanced supplement to Mill's Harm Principle. Clearly not all conduct can be regulated and ensured solely through a system of nudges. In fact, Thaler and Sunstein's libertarian paternalism only serves to regulate a small minority of conduct. For example, libertarian paternalism has nothing to say about crimes of really any kind – murder, assault, etc. cannot be appropriately regulated and punished under a system consisting solely of nudges. Thus, the question to be considered here is not whether libertarian paternalism supplants Mill's conception of the Harm Principle. Rather, we must determine if Mill's Harm Principle still holds valid when considered in light of the psychological failings of humans.

<u>The Unsafe Bridge – Limitations of the Harm Principle</u>

The first thing that must be determined is if Mill's Harm Principle can accommodate Thaler and Sunstein's concept of nudges. To ground this, we must rely on a passage from *On Liberty* known as The Unsafe Bridge. As a reminder, the Harm Principle can be described as the idea that the only time that anyone is justified in limiting another's actions is when that action would cause harm to another person.[24] This is a very strong anti-paternalist principle. Mill seems to commit to the idea that the state (or other people, for that matter) cannot act in a parental capacity, which would involve limiting one's actions to prevent harm to themself. The Unsafe Bridge is best known because it seems to undermine the absolute commitment Mill makes to his anti-paternalism.

[24] Dworkin, Gerald. *Morality, Harm, and the Law*, 9.

Mill's Unsafe Bridge is an example of a time that Mill thinks it is permissible to interfere with a person's liberty, even though they are not harming another person. Imagine a bridge that you know is extremely rickety – you see a person blithely walking towards the bridge, clearly about to cross it. Mill says that, in this instance, you are allowed to grab the person and pull them from the edge of the bridge, preventing them from crossing a bridge that you can reasonably assume they do not know is very unsafe.[25] While you do not prevent them permanently from crossing the bridge (which would be a strong paternalist example), you nonetheless detain them temporarily, which can be understood as an espousal of a weak paternalist principle.[26] Thus, Mill's Harm Principle is undermined, perhaps fatally. However, Samuel V. LaSelva takes another tack on this passage, arguing that instead "the bridge passage is not a case of paternalism, but its antithesis, good Samaritanism."[27] Good Samaritans, then, are individuals who want to help individuals satisfy their preferences by providing them with new information or filling in gaps of knowledge relevant to allowing individuals to make decisions, such as the status of the bridge as unsafe. According to LaSelva, Mill is not advocating a paternalist principle in the bridge passage. Rather, it is an espousal of a principle that secures an individual's sovereignty over themselves by making sure they do not inadvertently make a mistake which would be against their preferences.

LaSelva's argument turns on the distinction between positive and negative liberty. Traditionally, Mill is attributed to hold only the belief that freedom exists only as an absence of restraint on one's will. While it is clearly the case that Mill does believe that a component of

---

[25] Mill, John Stewart, *On Liberty* (Project Gutenberg, 2011). Page 182-183. Accessed from: https://www.gutenberg.org/files/34901/34901-h/34901-h.htm
[26] LaSelva, Samuel, "'A Single Truth': Mill on Harm, Paternalism and Good Samaritanism." *Political Studies, 36(3)*, 486–496 (1988). https://doi.org/10.1111/j.1467-9248.1988.tb00244.x. Page 488.
[27] Ibid, 491.

freedom is this negative freedom to be allowed to do whatever it is that one wants, LaSelva holds that Mill's bridge shows that "the connection between individual sovereignty and negative freedom completely breaks down."[28] That is, an individual's ability to do what they want (which is clearly the kind of freedom that Mill seeks to defend) cannot always be couched in terms of an absence of restraint. If it were indeed to be the case that sovereignty can be cashed out only in terms of negative freedom, "then accidents would simply be one of the consequences of individual sovereignty and to prevent accidents, even if justified, would be to restrict freedom or sovereignty."[29] This seems to be an almost indefensible position for Mill to hold and a commitment to it would be deeply problematic. This position would be indefensible for Mill because, logically, people can't be assumed to want to walk across a bridge that lead to their doom. If Mill were to bite the bullet and assent to the inevitably of accidents under a system of absolute individual sovereignty, it would be equivalent to saying that maximizing pleasure is served by allowing people to die due to preventable accidents that run counter to their own stated preferences.

LaSelva threads this needle by saying that Mill believes "to secure that sovereignty it is sometimes necessary to give priority to freedom from restraint, sometimes to the freedom to do certain things and at other times to the freedom of other agents."[30] So, then, to appropriately cash out freedom in a rigorous way, it is necessary to keep in mind that what matters the most is an agent's ability to, over a long time horizon, achieve all the ends that they actually want to. What "really is of importance, not only what [people] do but also what manner of [people] they are that do it" meaning that one's ability to actually execute their desires is equally as important as not

---

[28] Ibid, 490.
[29] Ibid, 491.
[30] Ibid.

being hindered in such aims.[31] The practical result of this distinction is that "good samaritans (*sic*) come to the assistance of others without insisting that their assistance be accepted."[32] If Samaritans would insist that their advice be taken, then they would cease to be Samaritans and would instead be paternalists. LaSelva is navigating the gap between providing individuals with the information necessary to make informed decisions and using coercion to ensure individuals make particular choices.

LaSelva's account provides a coherent and powerful view that squares Mill's strong commitment against paternalism with the temporary restriction of liberty contemplated in The Unsafe Bridge passage. However, LaSelva's demarcation of Samaritanism is different from nudges in a few key respects. To robustly defend the incorporation of nudges into Mill's Harm Principle these key differences must be addressed.

To begin with, let's return to The Unsafe Bridge. If good Samaritanism involves the actual restraint of a person's actions (say, by pulling them back from the edge) long enough to educate someone and determine if their preferences would actually align with crossing the rickety bridge, nudges are certainly less objectionable than Mill's elucidated example. While nudges operate similarly to Samaritans insofar as they do not insist that their prescriptions be taken as gospel, they do so in a different way. Nudges are necessarily impersonal in nature, operating on a level more categorical than personal. Whereas the Good Samaritan can only affect an individual's decision-making calculus through interpersonal communication, the nudge operates through different means. To stick with the bridge metaphor, a nudge would be a big sign reading "DANGEROUS BRIDGE AHEAD" and a chain roping off the bridge's entrance. This is

---

[31] Mill, John Stewart, *On Liberty,* 110.
[32] LaSelva, Samuel. "A Single Truth", 491.

necessarily less restrictive on the will than the Unsafe Bridge, because it involves no physical restriction of an individual's desires (i.e. pulling the individual away from the edge of the bridge).

Additionally, Mill contemplates only how a private citizen may be allowed to interfere outside the bounds of 'harm to others'. For our purposes, it is necessary to establish that nudges which are permissible for an inherently coercive institution (i.e. government) can operate in the same manner as an individual Samaritan. To answer this question, we must return to the purpose of nudges and examine the intent of the Harm Principle itself, adapted to the understanding of both positive and negative freedom.

Nudges are necessarily designed to help an individual to align their stated preferences with their ability to execute those preferences. They do so in a non-coercive manner, through encouragement and shaping of the choice architecture surrounding individuals. If the Harm Principle is incorporated with positive freedom, then nudges are permissible insofar as they direct individuals to being better able to execute their stated preferences. Mill's beliefs on government operation in this arena are very straightforward and, in some ways, a precursor to nudges as understood here. Mill says "it is a proper office of public authority to guard against accidents" insofar as crossing the bridge is regarded as an accident.[33] Moreover, Mill also endorses something that looks very much like nudges, saying "such a precaution, for example, as that of labelling the drug with some word expressive of its dangerous character, may be enforced without violation of liberty."[34] However, he couches this endorsement in terms of reference to criminal conduct (i.e. a drug that may be used as a poison). Mill simply did not have access to

---

[33] Mill, John Stewart, *On Liberty*, 182.
[34] Ibid, 183.

the level of research conducted by Sunstein and Thaler that could've identified heuristic failings to him. As such, he couldn't have considered nudges in terms of individual psychology.

Despite having access to no direct textual evidence where Mill signs onto the concept of nudges as such, his endorsement of The Unsafe Bridge shows how the intuitive inclusion of nudges within the Harm Principle framework may be robustly defended. However, we must consider the possibility that Mill would restrict the conduct of governments more stringently than the conduct of individuals. While individuals have limited impact, they can only pull a single person away from the bridge, governments can impact the behavior of a much wider swath of people, anyone who sees the sign. To hedge against this concern and expand the nudges that Thaler and Sunstein contemplate in *Nudge*, in the next section we will consider the difference between personal and institutional nudges and how the latter may better align with Mill.

The Role of Government

So, then, since we have seen how Mill can accommodate nudges into his Harm Principle, the next question to be answered is if governments have a responsibility to include nudges into their governance and, if so, what sort of nudges governments should encourage. For although it is the case that Mill may be okay with nudges, it does not necessarily therefore follow that governments are morally obligated to use them. However, it seems to be the case that a government that is invested in the well-being of its citizens has a responsibility to those citizens to institute nudges that reliably reflect their interests. In encouraging better behavior, as identified by the citizens themselves, governments fulfill their mandate to represent the well-being of their people. To support this, consider again the definition forwarded by Thaler and Sunstein for nudges: "any aspect of the choice architecture that alter people's behavior in a predictable way without forbidding any options or significantly changing their economic incentives."[35] In providing such a weak mandate, those who would choose to implement nudges have a wide latitude in how specifically to implement those nudges. As a result, nudges need not be onerous to either the government or the citizenry that is being nudged. More importantly, however, governments have the responsibility to nudge their citizens because the government is in a better position to understand the complex choices in which citizens find themselves. Thaler and Sunstein identify a number of complex choices that people are faced with throughout their lives and which the government is better equipped to deal with.

Consider, for example, retirement. Thaler and Sunstein note that "68 percent of 401(k) participants say that their savings rate is 'too low'" and further that "few of the participants who say they should be saving more make any changes to their behavior."[36] They therefore (correctly) draw the conclusion that people want to save more, but their psychological failings prevent them from actually changing their behavior to match. While Thaler and Sunstein mainly focus on the role of the private sector in encouraging their employees to save more, they make two salient points about the pension system and Social Security.

Firstly, they laud the example of the Pension Protection Act, a 2006 law that offers regulatory incentives to automatically enroll their employees into retirement plans.[37] The specific role of regulatory and tax-based incentives to nudging private institutions will be returned to later as a key point in the implementation of nudges, but the Pension Protection Act is important to flag here because it represents a way in which government can use its power to help make decisions on behalf of their citizens. This law enrolls employees into retirement plans without actually ever consulting them. However, both companies and the government can be pretty

---

[35] Thaler and Sunstein, *Nudge*, 6.
[36] Ibid, 109.
[37] Ibid, 118.

certain that employees will want to save something and that they will be amenable to a savings plan in which their contributions are automatically matched by their employers.[38] So, then, the government levies its power to change the outcomes in an uncontroversially positive direction.

Secondly, Thaler and Sunstein discuss Social Security as a way in which government intervention has failed to produce good results, even when well-intentioned. Social Security, although a single binary choice (withdraw or don't withdraw money), includes myriad factors that influence when the optimal time to withdraw that money is.[39] Most government materials on Social Security offer mere barebones information about what may influence the optimal time to begin withdrawing money from Social Security. This approach to the system leaves many people unable to make informed decisions about their retirement and can leave them destitute in their old age. However, this is not an indicator of the government's ability to effectively manage the system (i.e. an argument for privatization), but rather an admission of the failure of this particular government's approach to management. Thaler and Sunstein entirely agree with this, arguing that "if the SSA [Social Security Administration] had a designated choice architect, she could surely do better."[40] In other words, the failings of the SSA are the result of individual poor management.

This leads to the second point in favor of government nudges, as opposed to the full privatization of the enterprise. Unlike governments, private enterprise is invested in the expansion of profits. While in an economic system this is seen as a positive, as competition and desire for profits is supposed to drive innovation and increase the market's ability to satisfy

---

[38] Crucially, individuals who do not wish to be a part of this retirement scheme, or wish to change their contribution, may do so. The basic, default scheme is only what most people would desire, based on their own stated preferences.
[39] Thaler and Sunstein, *Nudge*, 118-119.
[40] Ibid, 119.

consumer demand, behavioral economics provides a counterpoint to this view. The ability to satisfy market demand is only as effective if the consumers' interests reflect their real desires. Thaler and Sunstein make a clear nod towards this problem, saying:

> Markets provide strong incentives for firms to cater to the demands of consumers, and firms will compete to meet those demands, whether or not those demands represent the wisest choices.[41]

So, then, markets have no interest in satisfying what is the best for people, or even what they think or say is this best for them. In the insatiable drive for profit, firms care only about what will create profits for them, even if doing so is sometimes against the wishes of the individual consumer. One only need to look at the deluge of advertising intended to create unnecessary desire in consumers for evidence of this phenomenon.

Governments, on the other hand, do not derive their legitimacy from profit. They have no need to compete with other entities for supremacy, because they are necessarily the only and most powerful entity in the state. As such, then, their main responsibility is directly to the citizens themselves, rather than their own profit. Interestingly enough, both businesses and politicians are interested in their own continued survival, but the mechanism by which their survival is ensured is different. Businesses are able to survive only if they continually profit. Politicians (and thereby governments), on the other hand, survive by getting reelected, which requires improving the lives of their constituents.[42] So while the government may fail in its charge to appropriately discharge its duties, it operates (at least ostensibly) outside of the systemic perverse incentives that private industry is susceptible to.

---

[41] Ibid, 49.

[42] Obviously there are legitimate concerns and criticisms over the real extent to which money runs the government, rather than voter preferences. While an in-depth discussion of these concerns is relevant and may change the instantiation of government-implemented nudges, that is a conversation for a different paper.

Moreover, not all examples of government programs that act as nudges have been failures. In fact, Texas has created one of the most successful examples of nudged behavior, the Don't Mess With Texas campaign. This campaign to stop littering has reduced roadside trash over 70% in less than a decade, all without making any changes in laws.[43] Instead, the state government played to the legendary pride of Texans, letting that be the catalyst for better behavior.

So, then, governments have two tools that make them a good candidate to utilize nudges for the betterment of their citizenry. Firstly, governments are in a good position to synthesize complex information and make it available for extremely large numbers of people. Individuals are extremely unlikely to know all the information they need in order to make informed decisions about complex decisions like retirement, information which can be collected and disseminated by the government. Additionally, governments are not afflicted by the same set of incentives that private companies are. While governments are not always effective, they are undoubtedly capable of providing nudges that can improve the lives of their citizenry. That being said, the question of government waste and ineffectiveness is a genuine and powerful concern for this discussion, which will be addressed in the next section.

Why Institutional Nudges?

It is important to note that government is not a hegemonic entity. It, like any other institution, is made up of individuals. Individuals, as we have already seen, are fallible in deep, problematic ways. In fact, the whole project of nudges is to help direct imperfect people into making better decisions. So, there seems to be a clear contradiction here: how can we rely on

---

[43] Thaler and Sunstein, *Nudge*, 60.

individuals to fix the systematic mistakes that individuals make? The answer to this question lies in the relationship that individuals have to institutions and governments. Like nudges, rules and laws act as checks on human behavior. Laws, in particular, are negative checks on human behavior – they demarcate what behavior may not engage in. Nudges, on the other hand, are a positive check on behavior – they improve the ability of individuals to engage in the behavior they want to. As such, unlike laws, nudges are less likely to implicate the concerns of justice that laws do. Choice architects certainly influence individual behavior and may do so in profound ways. At the same time, however, there are no punitive concerns to worry about – nudges by definition cannot be associated with punishment. While both laws and nudges should be taken seriously as ways to influence behavior, the difference between the two is important to answering the question of fallibility.

To illustrate this strength of nudges, consider the distinction between laws and nudges. Laws have punitive consequences, which can severely damage or even lead to the death of those who run afoul of them. Because of this, people are likely extremely wary of laws and take significant steps (such as the presumption of innocence and the massive appellate court systems) to ensure that injustice is not proliferated, especially when injustice would be the result of human error. However, nudges are only as effective as the nudged person allows them to be. For this reason, it would be basically impossible to nudge someone into doing or refraining from an act that they feel strongly about. So, if the nudger goes too far afield in their construction of the choice architecture, then the nudged person themselves would act as a check. For example, if the government set the default option to enroll employees in a retirement plan that automatically saves 95% of their income, it is doubtful that even the psychological power of default options would result in continued enrollment in such a plan for most people. A more extreme example is reflected in the popular *Purge* film franchise.[44] The government can reorganize the choice architecture in favor of murder, but no amount of nudging could force individuals to kill unless they actually wanted to. Human error in the construction of nudges is at least in part answered by humans themselves.

That being said, there are some nudges which operate subconsciously and are less susceptible to the natural check of human desire. If the option to reduce your automatic contribution to retirement is buried under two weeks of bureaucratic red tape, even the most dedicated individual would find it difficult to change their automatic retirement contribution. This strong of a nudge identifies a problem endemic to nudges – if they are made too onerous to overcome, they begin to approach the level of coercion that implicates mandates or laws. As Thaler and Sunstein say, "nudges count as such… only if any [cognitive] costs are low."[45] A retirement scheme that requires significant investment of time and cognitive resources is a coercion disguised as a nudge. Some of this can be ameliorated by proper design of nudges (i.e. making it relatively easy to break out of the nudge), but it is still possible that nudges would overstep and in those cases, there would be no check to rein in government overreach. This leads to an alternative, less onerous method of constructing nudges.

Importantly, nudges can be constructed in such a way so they don't necessarily always and only apply to individuals themselves. Constructed this way, they mirror laws. An important subcategory of the law is regulation – those laws that govern how institutions such as companies, nonprofits, and even governments themselves are allowed to operate. Regulatory action

[44] For those unfamiliar, the basic conceit of this film series is that once a year, all crime is legal for a day and citizens are invited to 'purge' their criminal penchants. Many of the heroes/heroines of the franchise are individuals who resist the temptation to purge and listen to the better angels of their nature.

[45] Thaler and Sunstein, *Nudge*, 8.

necessarily does not directly affect individuals. Instead, regulation shapes the surrounding environment in which people make decisions. Under regulatory-style nudges, individuals can be protected from the failures of government by nudges not directly impacting individuals themselves, but rather the institutions that help construct individuals' behavior. In other words, institutional nudges, rather than individual nudges. This approach to nudges has two significant benefits.

The first significant benefit of approaching nudges in terms of government-to-institution, rather than government-to-individual, is to save time and resources. As a general rule, individuals hold widely different preferences. Therefore, they need different nudges in order to help them achieve the particular outcomes they desire. Creating fine-grained nudges that work for each individual, or even classes of individuals, is likely to be extremely burdensome for the government. Moreover, the more burdensome an initiative is, the less likely it is to be successful as failures or inefficiencies in design and implementation stack up. On the other hand, institutions which attract particular types of people are better equipped to implement nudges that satisfy the preferences of their members. Consider the following example.

Imagine the government has discovered that a great number of people want to donate blood but do not and decided that a nudge is the most effective way to correct this irrationality. The government can choose between two different nudges, either 1) a public relations campaign that tries to convince citizens that others like them are donating, so they should too, or 2) a tax break for companies that bring host a certain number of successful (whatever that means), voluntary blood drives per year.[46] Incidentally, it is worth noting that tax breaks and waived

---

[46] The example of a public relations campaign is chosen here because of the power of conformity elucidated on page 62 of *Nudge*. Thaler and Sunstein show clear evidence that people can be nudged by trying to make them conform to

regulations do alter the economic incentives of companies. However, because the important unit of measurement to evaluating a nudge is the individual, rather than the institution, some leeway in the alteration of economic incentives for institutions is acceptable. While the power of a public relations campaign is undeniable, there are also a number of potential pitfalls in such a broad campaign, such as backlash to being told what to do, an inability to reach people, and so on. The nudge targeted at companies, however, gives those companies significant leeway in how to implement the nudge. Nudges which are effective for certain classes of individuals may not be for others. So, at least arguably, when the government is able to avoid having to paint with so fine a brush, nudges are more likely to be successful.

The second benefit of institutional nudges is that they have to compete against other strong incentive pressures. While individuals, too, have a number of competing interests and preference sets that would compete against nudges, Thaler and Sunstein's research shows clearly that nudges can compete with any of them. Many institutions, however, are subject to marketplace pressures that balance against nudges far more strongly than competing individual pressures. Consider the situation of three different institutions likely to be affected by nudges – private companies, nonprofit organizations, and branches of government.

The influence of the marketplace is most strongly felt by private companies.[47] Companies have no choice but to profit or die, a fact alluded to as a potential negative both in this essay and Thaler and Sunstein's work. However, the necessary attachment to profit by companies holds a different relationship when the private companies themselves are the ones being nudged. Like any other actor, private companies will (imperfectly) assess their various incentives and make

---

what their peers are doing. Moreover, both options require little government overhead and relatively easy to implement.

[47] The author is writing here in the context of a capitalist system.

decisions on the balance between those incentives. Nudges, the marketplace, commitments to social and environmental responsibility, for example, can all affect a private company's decision-making process. However, unless a company is profiting it will die, period. So, then, governmental nudges cannot place private companies in a place where they will become unprofitable and die. Just like nudges would not be able to lead an individual to kill, they would not lead a company to commit to a course of action that would lead to their demise.

If the government sees value in increasing the number of blood donors and wants companies to help achieve this goal, but the marketplace dictates that its achievement would be economically prohibitive, then the nudge will fail. This has two ramifications for a governmental objective. Firstly, it indicates that the nudge is particularly weak because it cannot overcome the pressures of the marketplace. Additionally, it gives the government more information about how to achieve their goal. Consider another example of government objectives – the reduction of CO_2 emissions. If the government implements a nudge to have companies reduce their emissions and the nudge fails, then the government may want to turn to more coerce uses of power, such as increased regulation and laws. Nudges are not the upper limit of government power and in cases where the government objective is to be prized enough to justify coercion, nudges do not require that the government forfeits this power.

Nonprofits, too, are subject to the marketplace. Rather than profit, charities are competing for donations. As such, their incentive structure is similar to companies – receive donations or die. While they need not be as rabid in their pursuit of donations as companies are for profits, they are nonetheless competing for limited monetary resources; in fact, the shoestring budgets that are common among nonprofits mean that actions which could adversely impact their donations are likely not to be adopted. Similarly here, then, nudges would not lead them down the path to self-destruction. Nonprofit entities would balance the influence of nudges against their desire to collect donations and make themselves attractive to donors.

The situation of these two institutions reflect the second significant advantage that institutional nudges have over individualized nudges. Unlike individuals, who face very few decision points which threaten their very existence, companies and nonprofits must always make their decisions with an eye towards their continued survival. This lends itself to a natural and convenient set of checks on nudges that would be lacking in individuals. If the government finds that its nudges are constantly being ignored in favor of other actions, it knows that its nudges are a failure and can redesign their nudges, advance to an action that implicates coercion, or abandon the program altogether. In the case of individual nudges, there is more uncertainty whether nudges are too strong, meaning individuals incentives are changed and the nudge acts more as a mandate, or actually helping people achieve their stated preferences. Companies and nonprofits cannot be nudged to change their incentives too strongly, so that question is no longer up for debate.

Branches of government are in a separate and special situation for a couple reasons. Firstly, more so than corporations or nonprofits, government offices are filled with individual people, many of whom are transient between administrations or elections. So, then, nudges targeted at government itself are more likely to implicate individuals, rather than hegemons. Moreover, because of government's special status in society above individuals and the marketplace, it is above the concerns of survival. While individuals within the government (especially politicians) are concerned with their survival in terms of job security, neither the government as a whole nor its various branches (such as Congress and the courts) and bodies (such as the EPA, IRS, and Justice Department) are constantly worried about collapse. It takes much more extreme circumstances to threaten the very survival of government, which means that it is less likely to provide the kind of feedback that is advantageous. It also is notoriously slow and cumbersome, which means it is less responsive to nudges. Because its livelihood is never threatened, government has developed into a massive, largely unresponsive entity. As such, it is less likely to be affected by nudges at all.

So, then, we have seen the advantages that institutional-level nudges have over individualized nudges. While there is significant benefit to individual nudges, the necessity of tailoring such nudges to the individual's avowed preferences represents an overwhelmingly complex task to be undertaken by the government, especially at the federal level. In essence, delegating the work of the actual behavioral changes to the institutions themselves, by nudging them, is a policy that better protects individuals from ineffective overreach. The built-in system of checks that constantly allows for nudges to be refined and made more effective is a positive side-effect of institutional nudges that should not be discounted. The next section of this essay will be more practical, discussing particular nudges that the government can take to lead to better outcomes for people.

Some Proposed Nudges

With an understanding of why nudges are important to good governance and how they can be used to align citizens' stated preferences with their actions, it is now possible to proffer some potential nudges that the government can take to nudge institutions in a direction that is ultimately beneficial for its citizens. While this section will not provide an exhaustive list of nudges that can be used, it is a useful starting point for a regime of nudges that can be implemented in the United States. A few examples have been used in previous sections, although those should be understood as mere exemplars to prove a point and will be set aside here.

Additionally, these nudges should be understood with a few caveats. Firstly, and most importantly, it is necessary to realize that nudges are only as effective as the choice architect designs them to be. Any nudge, if poorly crafted, can fail in achieving its goals or actively prevent people from aligning their preferences. For example, if it is determined that people should be enrolling in health insurance at a higher rate and to achieve this end the government creates a detailed (tedious) list of the benefits of various insurance plans, it may very well be the case that the government sees enrollment drop even further. In fact, the current insurance regime under the ACA has a tendency to act this way. However, the failure of this nudge to improve enrollment should be taken as a failure in implementation, not in concept. So, then, the examples in this section should be understood to be reasonably crafted to achieve their ends. Additionally, it will seem like the institutional nudge is a one-trick pony, with the nudge in most cases being tax credits for the institution in question. This is not a serious issue, because as we have noted previously the marketplace is a powerful tool to influence institutions, and tax breaks position entities, especially companies, much better than their competitors. So, while it may take a variety of strategies by institutions to nudge their members, it is not a bad thing that the government has one primary tool by which it nudges. While the government has as many tools in its arsenal of nudge-crafting as it is creative enough to dream up, including less incentive-altering ones such as public relations campaigns (like Don't Mess With Texas) and bureaucratic guidance, tax breaks that reposition institutions economically are easy to institute and intuitive. We will begin with a nod towards some of the example nudges outlined in *Nudge*. Thaler and Sunstein posit a number of nudges that are quite good and while the majority of their work focuses on individual nudges, some of their examples either are or can nonetheless be adapted well to institutions.

The first institutional example that Thaler and Sunstein outline relates to saving for retirement. As they say, data indicates “68 percent of 401(k) participants” are saving at a rate that “is ‘too low’.”[48] Their preferences clearly indicate that they would like to be saving more, so Thaler and Sunstein outline a couple of nudges that would help improve savings rate. The first is, in fact, an institutional nudge. They say that employers should simply make the default choice (a powerful economic incentive) to opt into saving X% of an employee’s income.[49] While this nudge could (and should) be implemented directly by companies in order to nudge their employees, it is also permissible to nudge those companies to nudge their employees. So, the government can offer tax incentives to companies who change their default option to enrolling employees in retirement accounts. Alternatively, the government could waive regulations for companies that have automatic savings enrollment or simply engage in a campaign that attempts to convince companies that they have a moral responsibility to help their employees save for retirement. In this way, the integrity of the original nudge (which encourages but does not require additional retirement savings) is retained.

Thaler and Sunstein also note a current government nudge that operates very similarly: the Pension Protection Act of 2006 waives a regulation for employers who automatically match the contributions of their employees and raise contributions over time.[50] Operating in conjunction with the proposed nudge above (perhaps by amending the Pension Protection Act), the government can both incentivize and retain employer interest in increasing retirement savings among their employees.

---

[48] Thaler and Sunstein, *Nudge*, 109.
[49] Ibid, 111.
[50] Ibid, 117-118.

They offer another set of good institutional nudges when discussing how to improve environmental impact of institutions and individuals. They note that, following the Chernobyl nuclear disaster, Ukraine created a program called the Toxic Release Inventory, which mandates that companies disclose how many pollutants they release.[51] While the disclosure itself is a mandate, the results of the disclosure are a nudge. No company has to change their behavior, but as groups and consumers note who is polluting the most, companies stand to lose customers and profits. This, in turn, nudges individuals who are environmentally conscious consumers to change their purchasing patterns to reflect their desire to consume responsibly. Thaler and Sunstein note two other important 'green nudges' in terms of consumer products. The first is a tool that visually shows homeowners how much energy they are consuming at any given time.[52] While this wouldn't require a change in behavior, it makes visual the amount of energy a person is using and lets them modify their behavior accordingly. Secondly, they note that some hotels require guests to use their room card to activate lights and air conditioning and automatically shutting those off when the room is empty.[53] In such scenarios, guests are incentivized to save energy in scenarios where they otherwise wouldn't.

These two nudges, in particular, are easily translated into government nudges at institutions. Consider that even though individuals ultimately bear the benefits of these energy-saving tools, it is possible for the government to fund institutions to make them up front. Home design, as Thaler and Sunstein note, is particularly tricky to encourage cost-saving architecture because "the costs of making a home energy efficient are borne up front by the builder, whereas the costs of heating and cooling are later paid by the owners."[54] However, these costs can be

[51] Ibid, 192.
[52] Ibid, 196.
[53] Ibid, 195.
[54] Ibid, 194-195.

inverted. By offering taxpayer funded grants and subsidies to builders (or municipalities for that matter) for building more greenly, the government can make it economically viable for construction companies to build the more expensive, better homes. Moreover, the costs of these grants can be funded by the savings from lower energy costs. Even if the individual never actually saves money, the flip in incentives can help individuals align their preferences to be more responsible consumers of energy.

Other institutional nudges can reflect people's assumed preferences, much like the Unsafe Bridge reflects the presumed preference of the bridge-crosser not to be harmed. One such example was recently implemented at my institution – Texas Christian University. The university has had a history of students crossing the street dividing campus at random locations, which represents an extreme danger to both pedestrians and drivers who may not be paying attention. The university has solved this problem by redesigning the curb to be lined with waist-high shrubs that nudge students to cross only at the designated crosswalks, where drivers are much more likely to be on the lookout for pedestrians. Governing bodies can co-opt this solution themselves by providing grants to institutions for such shrubs or high curbs that prevent individuals from jaywalking.

The final nudge that will be discussed is perhaps the most ambitious nudge in this paper, both in terms of cost to the institution and complexity of design. However, it is important to mention because it represents a radical shift in what nudges can be and how they can improve individuals' lives. Thaler and Sunstein mention a number of small design flaws that make people's lives vastly more inconvenient over the aggregate. For example, push doors with big handles are often pulled instead, as humans' unconscious brains equate handles with pulling.[55] While this doesn't represent a big inconvenience once, or even a few times, when aggregated over the course of months or years, this simple design flaw becomes incredibly frustrating. How can institutional nudges affect this problem?

Nudges can be helpful in this arena by incorporating what Dan Norman identifies as "Human-centered design (HCD)," a philosophy of building that considers how the design of a product interacts with the capabilities of the people that are attempting to use said object.[56] There is a laundry list of objects that are designed poorly to be understood by humans, whether for design reasons (i.e. designed to showcase a particular aesthetic value) or because of plain ignorance in the design process.[57] Ultimately, products can be made better or worse, and thus augment or diminish the experience of the individual using the product. The government can solve these small design flaws by offering a bank of designs to companies and in exchange for creating their product in accordance with this design the companies can receive tax credits or regulatory exemptions. In theory, the marketplace could be counted on to innovate products that align with HCD. However, the mere fact that we can point to so many design failures in everyday objects is evidence that we cannot always trust the marketplace in this arena. In nudging companies in this way, governments can ensure that individuals are more likely to encounter products that reflect HCD and thus are more user friendly.

---

[55]Ibid, 83-84.

[56] Norman, Dan. *The Design of Everyday Things, Revised and Expanded Edition*. (Basic Books, 2013). Page 8.

[57] Throughout his book, Norman identifies a seemingly inexhaustible number of objects that can be made better by HCD, from doors and windows to watches, stoves, automobile seat adjusters, nuclear power plants, thermostats, and paper money. The points here is that poor design is rife within our society and can be affected or even corrected through the use of HCD. Reader are encouraged to read Norman's book more closely in order to better understand the full implications of HCD, although for the purposes of this nudge this understanding should suffice.

Poor design is a prime candidate for nudging because people can reasonably be assumed to want less frustration and inconvenience in their lives. Poorly designed doors, stoves, and watches all increase frustration, so why not nudge them in a better direction? There are two potentially serious issues with this ambitious of a program: costs to implementation and interference in the marketplace of ideas.

The first issue is relatively simple – all that is required in these situations is to evaluate the frustration a given object causes and the cost to create and nudge a particular design. For example, a nuclear power plant with better design would have high creation costs, but also an extremely high upside of better design. Similarly, doors with clear indications of how to open would reflect low reduction of frustration, but also low design. On the other hand, objects like keyboards with illogical arrangement of letters represent a low reduction in frustration, but an extremely high cost in redesign and relearning, as individuals are used to and propagate the illogical skill.[58] So, the government can offer HCD blueprints for those objects where the cost of redesign is commensurate to the reduction in frustration.

The costs of interfering in the marketplace are much more dangerous. By offering HCD blueprints to companies, the government is arguably interfering with the natural process of innovation that is the hallmark of capitalism. In theory, those companies that create products that incorporate HCD are more likely to gain customers and be more successful. The government interfering in market innovation by innovating on behalf of companies seems like an impermissible violation of the libertarian paternalist principle. However, this argument trades on a misunderstanding of what motivates consumers. In the arenas where un-HCD products are

[58] Incidentally, this phenomenon is also seen in America's failed attempt to shift to the metric system during the 1970s-80s.

rampant, consumers are unlikely to notice the difference between products incorporating HCD and those without. In such scenarios, companies are not incentivized to make better-designed products, and individuals are hurt. For example, people are unlikely to make a decision about which television to purchase based on the construction of the remote control. So while it may seem that the government is giving away a competitive advantage in designing remote controls on behalf of companies, the reality is that without the government, remote controls would remain poorly designed. And while the government may not always design products that are superior to the marketplace, that is not an issue. Companies do not have to accept the designs proffered by the government if they believe they can innovate a superior one, but consumers are provided an increased possibility that they have access to a product designed with HCD in mind.

This section should not be considered a comprehensive list detailing the entirety of those nudges that the government can implement on behalf of individuals. However, it is a good sample of those situations that are ripe for institutional nudges and offers a potentially radical reimagining of how the government can operate on behalf of consumers to create better products.

Concerns and Objections

In the penultimate section of this essay, a few objections shall be considered and disposed of. Additionally, Thaler and Sunstein raise and answer a number of objections in *Nudge*. Their reasoning should considered to be adopted here, as they deal with some of the more fundamental objections to a regime of nudges. The objections here are more narrowly tailored to a system of institutional nudges, for the most part.

Firstly and crucially, it is necessary to discuss how a regime of nudges differs from an authoritarian system of legal moralism. Legal moralism describes a system of governance in which the legal system enforces laws that uphold a particular moral code. Legal moralism is often juxtaposed against libertarianism, because the former seeks to restrict conduct on moral grounds, while the latter encourages unrestrained conduct. Mill, Thaler, and Sunstein all oppose this kind of governmental control. The Harm Principle, of course, is deeply libertarian in nature, seeking only to restrain people when their conduct would interfere with others. As avowed 'libertarian paternalists', Thaler and Sunstein also reject legal moralism. However, it seems very easy for nudges to slide into encouraging a certain moral system. A nudge towards particular actions (say, for example, to increasing charitable donations) can arguably be said to shift people away from their preferences towards the moral system that the government prefers.

This objection is answered by the nature of nudges in themselves. Nudges by themselves have no moral content; they are merely changes in the choice architecture that align avowed preferences with actual actions. On principle, the government's use of nudges to mandate a particular moral system would be a misapplication of the principle. Take a recent (somewhat controversial) example, the Supreme Court cases *Obergefell v. Hodges*, which legalized same-sex marriage in the United States. While this case implicates questions of morality and some may argue that it is a form of legal moralism, the case itself is both a mandate and a nudge. It is a mandate in that enforces a particular set of protections within the United States – it legalizes the right of same sex couples to marry. So, *Obergefell* acts as a mandate in terms of securing rights for a particular group of people. However, it is also a non-coercive nudge towards accepting the LGBTQ+ community. Legitimization of the community in the eyes of government acts as a nudge towards the citizenry – if the government deems same sex marriage legitimate, others may as well. But, no one is forced to be more accepting.

The complaint that *Obergefell* is used as an enforcement mechanism for LGBTQ+ acceptance is nonsensical, because its properties as a nudge have no teeth of their own. That functionality is drawn from the one being nudged, because that is when nudges are implicated within choice. If someone is dead set on opposing same sex marriage on moral grounds, *Obergefell* can't change that. It may seem like this is a banal response to the objection. However, definitions matter, even if they are not of themselves a defense against bad policy. Certainly, the government could create a moralistic mandate and call it a nudge, but that does not make it a nudge. Similarly, calling a nudge a law does not mean that the nudge is backed by the same coercive force as a law. In the end, then, the structure of what a nudge is belies any attempt to use nudges as coercive means, in service of legal moralism or any other government end. So while nudges can implicate particular moral principles and nudge towards them, no one is coercively forced to accept those principles. This is what separates nudges that endorse a particular moral framework from a coercive mandate of a particular morality.

A second potential objection here is that many of the examples cited herein are individual nudges, rather than institutional ones. For example, the Don't Mess With Texas campaign, which should be upheld as an example of a successful nudge, is not directed at institutions. Rather, it is very clearly an individual nudge. Moreover, *Nudge* cites no nudges directed at institutions; all of their discussions are direct from government-to-individual. These examples should be primarily understood as proof of concept for nudges writ large. It is uncontroversial that nudges work – Thaler and Sunstein cite numerous examples of the success of certain individual nudges. Institutional nudges, on the other hand, are not treated within the work. However, this should not be taken as proof positive that institutional nudges will fail. Rather, all that can be said is that nudges directed at institutions are an open questions. Perhaps they will be effective, perhaps not, yet this can only be determined by actually implementing institutional nudges.

Additionally, there is some question about the efficacy of government nudges. The argument against privatized nudges turns in part on the marketplace's lack of concern with what is best for consumers, instead caring only about profit maximization. Similarly, it may be argued that governments are only concerned with what's politically popular, rather than what is best for constituents. For example, consider the strong opposition to soda taxes, which are in line with people's avowed desire to be healthier.[59] If government were truly better equipped than private companies to implement nudges (like the soda tax) that align with citizens' avowed preferences, then plans such as soda taxes should not be so strongly resisted.

However, the failure of soda taxes is not a strike against nudges, but rather one in favor of them. Governments proposed soda taxes, which individuals then mobilized to defeat so they could continue to enjoy sugary beverages, despite the health risks. They made a cost/benefit analysis and determined that the cost of the tax was not worth it. *Individuals made the decision they wanted to, even when nudged to the contrary*. The whole point of the project is shown here – to incentivize better behavior without mandating it. The soda tax incentivized a change in behavior that would align with nominal preferences, which individuals then chose to discard. It is true that the government did no better than the private market in promoting well-being, but it allowed individuals to choose via a good faith effort to improve lives, rather than attempting to drive profits.

---

[59] Grynbaum, Michael M, and Marjorie Connelly. "60% in City Oppose Bloomberg's Soda Ban, Poll Finds." *The New York Times* (New York, New York). Aug. 22, 2012. Accessed from: https://www.nytimes.com/2012/08/23/nyregion/most-new-yorkers-oppose-bloombergs-soda-ban.html; Thaler and Sunstein, *Nudge*, 44.

Another salient question about institutional nudges might be in terms of passing along costs to members of the group. For example, reconsider the example of the government offering tax breaks for corporations that host regular blood drives. It is a simple matter for those corporations to pass along that nudge to their employees in the form of a mandate – donate blood or be fired. Clearly in this case the point of the government nudge (to align people's desire to donate more blood with their actions) has been commandeered by the institution to become a mandate for the group that matters, the individuals.[60] So, then, an important hurdle to the appropriate implementation of institutional nudges is preventing those nudges from turning into de facto mandates to individuals.

The answer to this concern is, unfortunately, to fight fire with fire. As has been noted previously, nudges lack the power that mandates do. So, if an institution decides to bulldoze the purpose of the nudge for its own benefit, a nudge cannot prevent this from happening. As a result, the only way to counter a mandate is with a mandate. If institutions want to force the nudge upon their members, the only way to counter this with any sort of certainty is to forcibly prevent them from doing so. Ensuring nudges do not morph into mandates requires mandating that they remain nudges when passed to the individual. This is permissible because, while mandates should be avoided in the service of nudges, as long as the integrity of the nudge is retained at the level of the individual, the nudge is still operating as intended. So, then, the previous nudge could be amended to have tax breaks for certain amounts of blood donation by a corporation *as long as the corporation doesn't require its members to donate*. While this blends

[60] As a relevant aside, it should be noted here that the end of all discussions of nudging should ultimately be considered in terms of the individual unit. Institutional nudges should not be pursued at the benefit of the institution itself, but of the individuals that make up those institutions. Nudges in this paper are all geared at the benefit of the individual. While some nudges may be appropriate to improve institutions themselves, those style of nudges are beyond the scope of this discussion.

the idea of nudges and mandates in a somewhat confusing way, it is certainly not outside the realm of reasonableness for governments to take this sort of approach.

It is important to note that this is a unique feature of nudges with an intermediary. Individual nudges do not require formulation that protects individual choice, because without intermediaries there is no danger of nudges transforming into mandates. If the government were to nudge individuals directly into donating more blood, there would be no danger in the nudge being commandeered for less scrupulous purposes. While formulating nudges in a way that prevents them from morphing into mandates as they trickle down is a way to ensure that institutional nudges maintain their integrity, it seems a rather convoluted way to approach nudges. This brings up another salient concern with institutional nudges as advanced by this paper – why not simply stick with the less complicated and better tested individual nudge?

The answer to this issue lies in the costs of implementing individual nudges. As mentioned previously, one of the main concerns that should be kept in mind by choice architects is the efficacy of their nudges. Those nudges that are ineffective should be modified to better achieve their ends, or eschewed entirely in favor of better nudges. However, individual nudges lack robust counterbalance of the sort that institutions have in the form of the marketplace. So, while there are no doubt extra costs associated with implementing institutional level nudges, those costs must be weighed against the additional feedback that is gained when setting nudges against the power of the marketplace. It should be noted that perfectly crafted individual nudges and perfectly crafted institutional nudges will ultimately result in the same outcome – individuals who make choices that better align with their preferences. However, it is arguably more difficult to achieve well-designed nudges at the individual level.

Additionally, there is an important conversation to be had between how granular particular nudge regimes should be at different levels of government. Consider the difference in proximity between local, state, and federal governments. Many people know their local officials as friends and neighbors and vice versa, while state and especially federal officials are often many steps removed from their constituents. While this may be a benefit from a lawmaking perspective, but when a policy is designed to help people align their preferences to their actions, proximity is extremely important. So, it is more likely that nudges created by local governments would be more efficiently designed than federal nudges, which means that the crucible of institutions is less important.

Moreover, it should also be admitted that institutional nudges are not a panacea for all government objectives. Just as laws and mandates have their place alongside nudges, so too do both institutional and individual nudges have a place as well. As the Don't Mess With Texas campaign shows, not all nudges need to be filtered through the crucible of institutions to be effective and better incentivize people. In fact, it is entirely likely that, as nudges are refined through the feedback of the marketplace, they can then be shifted away from institutions back to individuals. For example, if the government implements a nudge regime through employers to encourage increased retirement savings, there is no specific need to keep that regime through the employers forever. Once the government figures out what works well, they are more than capable of shifting the same nudge system directly to individuals, thus attempting to better secure the individual liberty so important to Mill, Thaler, and Sunstein.

Conclusion

Thaler and Sunstein's work on nudges represents a potential watershed for political philosophy. While not necessarily created to be understood in such terms, Thaler and Sunstein nonetheless have designed a system of economic incentives that has significant implications for how we understand the role of government. Historically powerful political philosophies – notably Mill's Harm Principle – are strongly affected by Thaler and Sunstein's insights and must be understood in light of the revised and deepened understanding of human psychology that we now have.

The Harm Principle stands as one the most powerful ways to understand governmental authority. In theorizing that government authority is limited to preventing harm to others, Mill makes some key assumptions about human nature. Rightfully, Mill theorizes that individuals should be allowed to set their own preferences and craft their lives in order to fulfill those preferences. However, he also assumes that individuals are capable of crafting plans that reasonably direct their lives towards their preferred ends. Thaler and Sunstein show that this assumption is far from guaranteed, and in fact may be one of the most unreasonable assumptions in political philosophy. What they show is that the assumption of perfect rationality, which undergirds so much of political theorizing, is unable to reflect the reality of human behavior.

Mill's anticipation of irrationality, in the form of the passage of the Unsafe Bridge, is not an indictment of his Harm Principle, but rather an understanding of how individuals' stated preferences can be anticipated and furthered by responsible actors. While he remains committed to his view that the government cannot interfere except in cases where one actor could harm another, he does recognize that people are not omniscient and that, in cases where individuals would be actively harming themselves, Mill recognizes the reality of intervention. It is not inconsistent with Mill's principles, but rather an endorsement of them in the fullness of their implications.

While the jump from intervention in cases of harm to securing of preferences is not trivial, it is rather short. From the perspective of the actor, the difference between prevention of active harm and furthering of preferences is small. Moreover, the result of nudging is even less burdensome than most intervention. Because individuals are never forced to comply with the terms of the nudge, intervention in this small way only is not prohibitive to any individual actions. Mill likely never contemplated the concept of nudges, nor did he have access to the sophisticated theoretical and economic frameworks that exist in today's world. Ultimately, then, the Harm Principle is more than happy to accommodate nudges and welcomes them as a valid exercise of government power and knowledge.

Moreover, the inclusion of nudges as a tool in government does not lead us down the slippery slope to paternalism. Granted, Mill's belief that individuals are best able to satisfy their preferences is refuted by the work of behavioral economists like Thaler and Sunstein; their work shows that people often make irrational decisions that work at cross purposes to their stated preferences. However, this is no reason to doubt that individuals do not set their preferences, merely that irrationality in their achievement is inevitable. If an individual wants to smoke or drink sugary soda, despite the health implications, then they should be free to do so. If an individual wants to quit smoking, but addiction prevents them from doing so, then nudges that incentivize quitting are warranted. If the nudge doesn't work, it is because the nudgee really has satisfied their preferences. Mill's Utilitarian principle is still satisfied – the individuals are allowed to choose the behavior that best maximizes their happiness, as indicated by their ability to set and satisfy their preferences. Nudge simply allow individuals to do so.

While Thaler and Sunstein do outline a well-realized regime of nudges and how they can implemented in today's economic reality, they do not exhaust all the promise that nudges represent. Institution-level nudges, in particular, are not contemplated by the two because they are concerned with those that have a direct impact on the individuals. Direct nudges are powerful tools, as they allow direct impact on the individual without concerns over passing along costs or the inefficiencies that inevitably develop when bureaucracy is layered. At the same time, however, direct nudges pass along any problems in design directly to the nudgee.

Institutional nudges hold promise not in directness of impact, but rather in efficiency of implementation. Many of the problems that Thaler and Sunstein note in *Nudge* with government design of nudge regimes is due to problems with proximity of the nudger. Successful government nudges, on the other hand, are by and large implemented closer to home, often through state governments. Institutional nudges trade more explicitly upon this success, relying on the proximity of the institutions to their members to decide how they are best nudged. While not always the right decision for a given objective, institutional nudges are another tool in the arsenal of a just government.