

2019

An Ethical Examination of Libertarian Paternalism: The Impact of Retirement Savings Nudges on Autonomy and Welfare

Edward E. Dunn V
Bard College

Recommended Citation

Dunn, Edward E. V, "An Ethical Examination of Libertarian Paternalism: The Impact of Retirement Savings Nudges on Autonomy and Welfare" (2019). *Senior Projects Spring 2019*. 172.
https://digitalcommons.bard.edu/senproj_s2019/172

This Open Access is brought to you for free and open access by the Bard Undergraduate Senior Projects at Bard Digital Commons. It has been accepted for inclusion in Senior Projects Spring 2019 by an authorized administrator of Bard Digital Commons. For more information, please contact digitalcommons@bard.edu.

An Ethical Examination of Libertarian Paternalism: The Impact of Retirement Savings Nudges
on Autonomy and Welfare

Senior Project Submitted to
The Division of Social Studies
of Bard College

by
Edward Dunn

Annandale-on-Hudson, New York
May 2019

Acknowledgements

Thank you to Martin Binder for showing me what economics can be,

And Thank you to Mitra for not asking what if

Im Flugzeug gibt's Lust
In der Straße gibt's Fremdheit
Im Herz gibt's erdnussoße

PLAGIARISM STATEMENT

I have written this project using in my own words and ideas, except otherwise indicated. I have subsequently attributed each word, idea, and table which are not my own to their respective authors. I am aware that paraphrasing is plagiarism unless the source is duly acknowledged. I understand that the incorporation of material from other works without acknowledgment will be treated as plagiarism. I have read and understand the Bard statement on plagiarism and academic honesty as well as the relevant pages in the Student Handbook.

- Edward Dunn

An Ethical Examination of Libertarian Paternalism: The Impact of Retirement Savings Nudges on Autonomy and Welfare

“It is astonishing what foolish things one can temporarily believe if one thinks too long alone, particularly in economics” - John Maynard Keynes

Abstract

Behavioral economists have recommended the use of nudges to promote welfare through public and private policies as a form of ‘libertarian paternalism’. It remains a controversial topic and opponents of nudging claim that it violates individual liberty and autonomy. Even so, nudging has found success in helping people save more for retirement. Automatic enrollment and the Save More Tomorrow™ plan use defaults to encourage higher participation and contribution rates in 401(k) retirement savings plans. I argue that these policies violate individual autonomy and fail to determine individuals’ preferences. Ethical interventions prioritize the intelligibility of preferences, enhance individual autonomy, and promote welfare over time.

Keywords: Behavioral Economics, Nudging, Rationality, Ethics, Retirement Savings

JEL Classification: D91

Table of Contents

Introduction.....	1
Chapter 1.....	3
Chapter 2.....	12
Chapter 3.....	18
Chapter 4	25
Chapter 5	30
Chapter 6	41
Conclusion	48
References.....	50
Appendix.....	53

Introduction

Decisions are not made in a vacuum. Richard Thaler argues that, “the first misconception is that it is possible to avoid influencing people’s choices” (Sunstein & Thaler 2009, 12). If this is true, then a tremendous weight is placed on the shoulders of the ‘choice architects’ who influence — intentionally or unintentionally— how people make decisions. The design of the environment in which people make decisions has a strong effect on underlying cognitive processes that influence how decisions are made. That being so, it may not be the case that humans are always in control of their decisions. Most economic models of choice depict the *homo economicus*, or the economic man, who is able to “think and choose unfailingly well,” and “store as much memory as IBM’s Big Blue” (Sunstein & Thaler 2009, 7). Unfortunately, humans in the real world rarely exhibit these traits.

Since the rise of behavioral economics, the standard model of rationality has been criticized for its inability to explain anomalies in human decision making. Insights from psychology have identified a number of cognitive biases and heuristics that help explain these anomalies. Heuristics and biases are mental shortcuts, or “rules of thumb,” (Sunstein & Thaler 2009, 28) that we use to make decisions. They appear systematically and predictably, acting as cognitive errors that negatively affect judgement and decision making. In light of these discoveries, behavioral economists have found a way to apply the insights of their field: nudging.

Nudging is a method of modifying the environment in which decisions are made in order to “alter people’s behavior in a predictable way without forbidding any options or significantly changing their economic incentives” (Sunstein & Thaler 2009, 8). By relying on an understanding of cognitive biases and heuristics, nudges could help steer people to make better

decisions in situations where they are most likely to err. That being so, nudging could promote welfare through public and private policies as a form of libertarian paternalism. However, it remains a controversial topic and opponents of nudging maintain that it is “insidious and deeply manipulative” (Rebonato 2015, 12), exploiting cognitive weaknesses and taking away our freedom to choose.

With consideration of these ideas, this paper explores the ethics of libertarian paternalism and the economic implications of nudge policies in the context of retirement savings. Arguably, people around the world are not saving enough for retirement. To add, there are many people who express in interest in wanting to save more for retirement, but do not. This inconsistent behavior has been attributed to cognitive biases and decisional inertia. As such, this paper considers the ethical implications of libertarian paternalistic interventions aimed at helping people overcome these factors and save more for retirement. Over the course of this investigation, I discuss the ethical implications of libertarian paternalism and its applications in retirement savings. In this effort, I provide a balanced review of the criticisms and defenses of libertarian paternalism. Having considered these arguments, I then look at the ethical implications of automatic enrollment and the Save More Tomorrow plan — two popular libertarian paternalistic retirement savings interventions. With consideration of these policies, I then offer a conclusive recommendation of how individuals could be ethically helped to save more for retirement. In my view, I argue that ethical retirement savings interventions have the capacity to enhance individual autonomy and promote welfare over time.

Chapter 1

What is a Nudge?

Cass Sunstein and Richard Thaler, the two main advocates of libertarian paternalism, define a nudge as

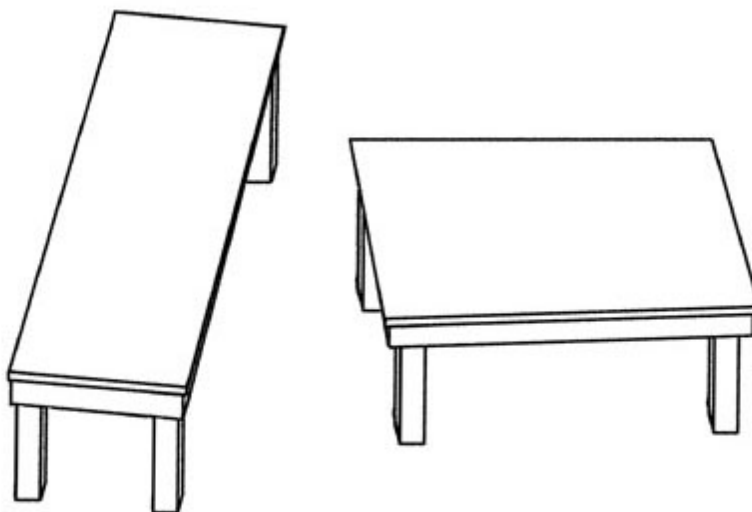
“any aspect of the [choice architecture](#) that alters people's behavior in a predictable way without forbidding any options or significantly changing their economic incentives. To count as a mere nudge, the intervention must be easy and cheap to avoid. Nudges are not mandates. Putting fruit at eye level counts as a nudge. Banning junk food does not” (Sunstein & Thaler 2009, 12)

To put the definition of a nudge in context, let us consider an example. In Chicago, there is a scenic road called Lake Shore Drive that overlooks the city's iconic skyline. A section of this road puts drivers through a series of S curves that has become infamously dangerous because of countless car accidents. The city of Chicago has addressed the issue by painting the road with white lines (that do not provide tactile information, only visual). When the stripes first appear, they are evenly spaced, but as the driver reaches the most dangerous portion of the curve, the stripes get closer together, creating the sensation that driving speed is increasing (see figure 1). Predictably, this causes most drivers to slow down, so we could call it a nudge. Surely, the driver could choose to drive quickly around the corner if they want to, but they are often nudged to go slower. This is a nudge that saves lives.

While the Lake Shore Drive example is not taking advantage of cognitive shortcomings, per se, it is example of how people can easily be nudged, especially in situations that are unfamiliar to them. Furthermore, it holds value as an example of how humans could be nudged in their own benefit. It seems fair to say that most people do not want to die in a car accident. While the Lake Shore Drive example demonstrates how a nudge functions generally, this paper investigates nudges that aim to correct irrationalities in economic decision making, particularly in retirement savings.

How do nudges work?

Consider the following image of two tables:



(Sunstein & Thaler 2009, 190)

At first glance, the table on the left will seem longer to most people. However, if a measurement is taken, it becomes evident that both tables are identical in their dimensions. Normally the human mind works remarkably well, but in some situations it can deceive us. This is a key insight that psychology has brought to behavioral economics. Thaler writes that even

“Einstein would probably have been fooled by those tables” (Sunstein & Thaler 2009, 21). From examples like this table illusion, we learn about the fallibility of the human mind — a mind that adopts sensible rules of thumb that sometimes “lead [people] astray” (Sunstein & Thaler 2009, 40). This does not mean that there is something wrong with humans, but it does mean that our understanding of the human mind can be improved by understanding how we consistently make the same errors in our decision making. Nudging could help us avoid making irrational decisions and judgements.

There are certain peculiarities of the human mind that allow nudges to work. Daniel Kahneman in his book *Thinking, Fast and Slow* argues that humans have a “System-I,” which is fast, intuitive, and emotional, and a “System-II,” which is slow, deliberate, and logical (Kahneman 2015, 13). Cass Sunstein describes our system 1 to be “a bit like Homer Simpson,” while our System-II is “more like Spok from *Star Trek*” (Sunstein 2015, 26). The majority of the cognitive heuristics occur systemically in system-I. Depending on the situation, we are more or less likely to be at risk of making a cognitive error, but we are particularly vulnerable in unusual or unfamiliar situations (Sunstein 2015). Nudging, therefore, is most useful in these types of situations because that is when people are most likely to rely on our system-I to make a decision, and hence the most likely to be affected by a cognitive bias.

As an example, of a cognitive bias let us consider loss aversion. Loss aversion is a cognitive bias that says we dislike losses more than we like corresponding gains. Consider a simple experiment: half of the students in a class were given mugs with their university insignia and asked how much money they would be willing to sell them for. The other half, who were not given mugs, were asked for what price they would be willing to buy a mug. Predictably, the students who were given mugs demanded roughly twice as much to sell their mugs as the others

were willing to pay to acquire one (Tversky & Kahneman 1991). These results also demonstrate that framing a problem in terms of gains and losses has a strong effect on how people make decisions. As an example of the framing effect, Tatiana A. Homonoff explored the impact of framing on the use of plastic bags. She found that if people must pay a five-cent tax for using a plastic bag, they are more likely to be affected than if they are given a five-cent bonus for bringing their own bag (Homonoff 2018). In this case, losses are valued more than gains. This is loss aversion at work.

Interestingly, framing effects seem to disappear if these types of problems are dealt with in a different language. When people are asked to consider questions like those of the mug experiment in a foreign language that they speak proficiently (but not necessarily comfortably), they are less likely to show framing effects or loss aversion (Keysar et. al. 2012). When people are speaking their native language, they think quickly and effortlessly using system-I, but when people are speaking another language they are less comfortable with, their system-I becomes overwhelmed and system-II is activated. The authors of the relevant study call it the ‘foreign language effect.’ A reduction in emotional resonance is associated with speaking a foreign language, and emotional reactions are often responsible for inducing poor decisions. Indeed, the human mind may be fallible, but this experiment reveals a silver lining: when humans think slowly, system-II takes over and we become less likely to make cognitive errors.

With consideration of these examples, the fallibility of the human becomes recognizable. This is problematic from an ethical point of view because it directly challenges the notion of what it means to be an autonomous decision maker. In many situations, people may not be as in control of their own decisions as they think they are. Politicians and policymakers are more than

aware of this fallibility, and have turned to libertarian paternalistic nudge interventions that purportedly influence people's behavior to promote their welfare.

One area where nudging has been arguably most successful is in retirement savings plans. Some would argue that retirement savings rates in the U.S. are too low. Notably, these low savings rates can be attributed cognitive biases that prevent people from joining retirement savings plans. In order to address this problem, behavioral economists have recommended the use of nudge policies to encourage participation in retirement savings plans and to increase contribution rates.

Nudging in Retirement Savings

The entire U.S. retirement system stands on a three-legged stool. Americans must rely on Social Security, pensions, and personal savings to provision for the future. In *When I'm Sixty-Four* by Ghilarducci she argues that a comfortable retirement plan requires a replacement of 65% to 80% of pre-retirement income (Ghilarducci 2018, 26). Relying on anyone of these sources alone would not be sufficient to meet this goal. Also, there has been a national shift from defined benefit pension plans to defined contribution pension plans. Among full-time employees in medium and large private establishments, the percentage of those covered by a defined benefit plan has fallen from 80% in 1985 to merely 33% in 2003 (Ghilarducci 2018, 26). As a result of this shift, the responsibility to save for retirement mostly rests on the employee's shoulders.

Despite it being one of the primary vehicles for individuals to save for the future, defined contribution pension plans are seeing enrollment rates that "are far from 100 percent" (Benartzi & Thaler 2007, 82). There is even a common plan where employers match 50 percent of

employee's contributions up to some threshold. In this case, it would only be rational for employees to participate in such a plan with a "sure profit opportunity" (Benartzi & Thaler 2007, 82), but we still observe that some people do not join. As an extreme example of employees' apparent reluctance to join a retirement plan, data on 25 defined benefit plans in the United Kingdom report that only "half of the eligible employees signed up" (Benartzi and Thaler 2007, 82). This example in particular begs the question: what is preventing people from signing up?

It has become apparent that people do not save like standard economic theories of savings predict. The life-cycle model and permanent income model have a hard time explaining such low enrollment and savings rates. Even so, that is not to say that people don't want to be saving more. In a relevant study, Choi et. al found that two-thirds of sample participants in 401(k) plans viewed their savings rate as too low. Among these people, 35 percent expressed an interest in increasing savings rates in the next few months, but 86 percent of them never ended up doing so (Choi et al. 2004). From this finding it could be argued that there are people who have expressed a preference to be saving more, but are not.

One reason why people are not acting on their expressed preferences is that they procrastinate. Behavioral economists argue that people discount hyperbolically, which leads to procrastination. Hyperbolic discounting says that, as time passes, people become more patient. The perceived costs of waiting are exceeded by the perceived benefits of a larger reward in the future. When considering a decision today, on the other hand, people show time-inconsistent behavior and "[weigh] current and near-term consumption especially heavily" (Benartzi & Thaler 2004, 81). This behavior leads to impulsivity and procrastination — two factors that behavioral economists argue prevent people from saving more for retirement. Coupled with a bias for the status quo these two factors thought to generate enough inertia to prevent people

from saving the money they want to save (Madrian & Shea 2000). With consideration of these factors, behavioral economists have designed a few nudge policies aimed at increasing retirement plan enrollment rates and savings rates.

Automatic Enrollment

Since low enrollment is thought to be the result of inertia, economists have recommended the use of a nudge that leverages it. This method, known as automatic enrollment, changes the default to automatically enroll employees into a retirement savings plan. Instead of having employees opt-in to a retirement plan, they would be automatically enrolled unless they actively choose to opt-out. This is a nudge because it supposedly leverages the employees' inertia for their own good. This strategy has shown to substantially increase enrollment in U.S. defined contribution plans. At one firm where automatic enrollment was adopted, participation rates jumped to 90 percent immediately and increased to 98 percent after 36 months (Madrian & Shea 2000). These findings indicate that, under automatic enrollment, employees start saving sooner and more participants eventually join. Automatic enrollment, therefore, is thought to help people overcome procrastination and a lack of willpower, helping them realize their expressed preference to save.

Additionally, since automatic enrollment automatically enrolls people into a plan, the plan must have a default savings rate and asset allocation plan. The defaults are supposed to be conservative, so that employees start saving at a low rate with low risk. This gives employees the freedom to increase their savings rate, if they choose, or reallocate how their assets are invested. In Madrian & Shea's study, they also recommend, under automatic enrollment, the large portion of contributions should be directed into the money market fund "where there is much less

variation in returns than is the case for stock funds” (Madrian & Shea 2000, 38). Still, though, the asset allocation is conservative. Automatic enrollment, in this way, is similar to organ donation nudges that set defaults and actively require participants to opt-out (Dalen & Henkens 2014). This approach is useful for bringing new participants into retirement plans, effectively providing positive fund balances to those who previously had zero fund balances in their 401(k) plans.

Save More Tomorrow™

Next, Thaler and Sunstein argue that, even though automatic enrollment is effective at increasing participation rates, people “tend to stick to the default contribution rate” (Sunstein & Thaler 2009, 122), which are typically quite low. In order to address this problem, Thaler and Benartzi created a program called the Save More Tomorrow (SMT) program that nudges employees into saving more than they might save if they were just automatically enrolled. The main feature of the Save More Tomorrow plan is that employees commit in advance to timed contribution increases that coincide with pay raises. Because of hyperbolic discounting, the lag between when employees commit and the first timed contribution increase should be “as long as feasible” (Benartzi & Thaler 2004). Second, by synchronizing pay raises with savings increases, participants will “never see their take home amounts go down” (Sunstein and Thaler, 2008). This helps employees not view increases in their savings contributions as losses to their income. Since people are loss-averse, the SMT program prevents people from seeing a decrease of their income in nominal dollars. Third, the contribution rate increases steadily until it reaches a preset maximum, allowing inertia and the status quo bias “[keep] people in the plan” (Benartzi and Thaler 2004, 170). Lastly, libertarian paternalists argue that this plan, when combined with

automatic enrollment, increases enrollment rates *and* savings rates. For that reason, SMT is often coupled automatic enrollment.

Chapter 2

Ethical Frameworks

So far, we have seen two well-known examples of libertarian paternalism: automatic enrollment and SMT. As it is my primary intention to engage with the ethical implications of libertarian paternalism, it is first important to understand the ethical frameworks from which it emerged. Keeping automatic enrollment and SMT in mind, I first outline the important features of libertarianism and paternalism, after which I explore how they have come together as a reconciliation of these philosophies.

Libertarianism

Libertarianism is the view that “[ascribes] an intrinsic value to the individual’s ability to express and enact her choices without hindrance” (Rebonato 2015, 19). This means that libertarianism is not consequentialist by nature. In other words, advocating freedom of economic choice because productivity will be enhanced is not libertarianism. Keeping this in mind, I would argue that the most salient feature of libertarianism says that an individual’s freedom to choose should be preserved, and it should not be preserved with an instrumental justification. As an extension of this idea, a policy would not be libertarian in nature if it impedes an individual’s right to choose. Any policy that exercises power over another individual is unjust if it acts against their will. As such, most forms of government intervention are anti-libertarian because there is the implicit assumption in libertarianism that says that people are capable of making decisions for themselves.

Autonomy in Libertarianism

At the heart of libertarianism is the notion of autonomy. Individual autonomy is the idea that people can choose for themselves and, more importantly, that individuals are self-governing agents. An autonomous agent would make decisions directed by considerations, desires, conditions, and characteristics that are a part of one's "authentic self" (Christman 2007). Furthermore, I would argue that autonomy should be distinguished from freedom, which generally emphasizes the importance of the ability to act without external or internal constraints. Autonomy, therefore, differs from freedom in that it concerns the independence and authenticity of the desires that make someone do something in the first place. If autonomy is not merely the absence of internal and external constraints, it may not be the case that autonomy is solely the freedom to choose. As mentioned, an autonomous decision reflects the will of the agent's authentic self. Authenticity, however, is hard to measure or recognize. Therefore, the notion of authenticity imposes a fragility on individual autonomy that I would like to pay close attention to. That being so, it becomes difficult to determine what an authentic decision truly is. In order to substantiate this view of autonomy, I also turn that of Dworkin, which says that autonomy is not only the nominal freedom to choose but "the capacity of a person to critically reflect upon, and then attempt to accept or change, his or her preferences, desires, values, and ideals" (Dworkin 1988, 48).

Paternalism

Generally, paternalism is 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 2017). While there are many forms of paternalism, the unifying idea is that there is a reduction of an individual or group’s freedom for a benefit of greater proportion. The loss of freedom is the cost for making an individual or society better off in some way. In *On Liberty* by John Stuart Mill, he writes that the government should not exercise power over people unless it aims to prevent harm to others. To add, he insisted that, as a rule, the government may not exercise power over an individual if its only goal is to protect an individual from themselves. Mill argues that,

“the only purpose for which power may 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 mental, 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. The only part of the conduct of anyone, for which he is amenable to society, is that which concerns others. In the part which merely concerns himself, his independence is, of right, absolute. Over himself, over his own body and mind, the individual is sovereign” (Mill 1869, 8).

Still, there are numerous examples of laws and regulations that call Mill’s Harm Principle into question. For example, there are laws that require people “to get a prescription before obtaining certain medicines,” while others “forbid employees from working in unsafe

workspaces,” (Sunstein 2015, 4) even if they would voluntarily take the risk. In some cases, power is not only exercised over people to prevent harm done to others, but over individuals “to promote their own good” (Sunstein 2015, 4). As such, there many interventions that are designed to protect people from themselves. In certain contexts, people are particularly vulnerable to error, giving paternalistic interventions a “strong and emphatically moral argument” (Sunstein 2014, 5) to improve an individual’s well-being. Even so, paternalistic policies will always come at a cost: a loss of freedom. That being so, policy makers like Richard Thaler and Cass Sunstein believe they have discovered the ‘real third way’ — a way that preserves individual autonomy while simultaneously exercising a soft form of paternalism.

Since libertarian paternalism is thought to be a form of soft paternalism, then understanding the difference between soft and hard paternalism becomes important. Soft paternalism is the view that interventions are justified only under conditions when it is necessary to determine whether an individual or group is acting voluntarily and knowledgeably. To use Mill’s famous damaged bridge example, consider the following quote:

If either a public officer or any one else saw a person attempting to cross a bridge which had been ascertained to be unsafe, and there were no time to warn him of his danger, they might seize him and turn him back, without any real infringement of his liberty; for liberty consists in doing what one desires, and he does not desire to fall into the river. (Mill 1869, 88)

This example illustrates the notion of soft paternalism quite well. In this situation, the person attempting to cross the bridge faces sure danger. Secondly, since there is not enough time

to warn him, forcibly stopping him would not only prevent them from being harmed, but it would be an opportunity to determine whether or not he is acting voluntarily and knowledgeably. If it is determined that the person is aware of the bridge's unsafe condition yet still wants to cross, then a soft paternalist must allow him to continue. On the other hand, a hard paternalist would say that even if the person is aware of the risks and wants to, say, commit suicide, then it may be permissible to prevent him from crossing the bridge. In this case, a hard paternalist would be entitled to prevent voluntary suicide. With consideration of these differences, it is interesting that nudging is often promoted as a form of 'soft' paternalism because, as I will argue later in this paper, some nudges may not be so soft, after all.

Libertarian Paternalism

Richard Thaler and Sunstein write that libertarian paternalism is “an approach that preserves freedom of choice but authorizes both private and public institutions to steer people in directions that will promote their welfare” (Sunstein and Thaler 2009, 5). They continue by writing that they strive to design policies that “maintain or increase freedom of choice” (Sunstein and Thaler 2009, 5). In their understanding, they argue that libertarian paternalistic policies try to influence choices in a way that will “make choosers better off, as judged by themselves” (Sunstein and Thaler 2009, 5). In many situations, Sunstein and Thaler argue that individuals make poor decisions — decisions that they would not have made if they were properly informed. To add, Thaler and Sunstein insist that it is a light, nonintrusive form of paternalism because choices are never “blocked, fenced off, or significantly burdened” (Sunstein and Thaler 2009, 6).

Keeping this in mind, the goal of libertarian paternalism is supposedly to promote welfare. So, in order to promote welfare, libertarian paternalists rely on ‘choice architecture’ to supposedly influence people in their own benefit. In their view, choice architects have the “responsibility of organizing the context in which people make decisions” (Sunstein and Thaler 2009, 3) in hopes of steering them in a particular direction. Choice architecture generally manifests in two different forms. The first form of choice architecture describes that which is created intentionally, i.e. with the intention serving a particular purpose. While these intentional forms of choice architecture are not all created with the intention of nudging, per se, they are still features of the context in which people make decisions, and thus have an effect on how people make decisions. A door handle, for instance, is designed to be grabbed, so most people will grab the door handle in order to the open a door. In this example, the door handle influences behavior, so it can be considered an intentional form of choice architecture. The second form of choice architecture manifests unintentionally and is the product of uncontrollable factors, such as the weather. A study by Busse et al. found that weather has a significant impact on consumers who are purchasing vehicles. When the weather is warm and sunny consumers may fall victim to the projection bias, i.e. the tendency of individuals “to over predict the degree to which their future tastes will resemble their current tastes” (Busse et al. 2012, 1), and purchase a car with a sunroof. With consideration of these ideas, the supporters of libertarian paternalism argue the choice architecture is inevitable, rendering many forms of paternalism unavoidable, so people might as well be nudged in a positive direction.

Chapter 3

Criticisms of Libertarian Paternalism

In this chapter I outline the most common criticisms of libertarian paternalism. In this effort, I hope to highlight the depth of the debate surrounding the ethics of libertarian paternalistic interventions. Furthermore, I explore how these criticisms apply to automatic enrollment and Save More Tomorrow.

Libertarian Criticisms

First and foremost, it is important to remember that libertarian paternalism is a form of paternalism. As such, libertarianism naturally opposes paternalism, so we will begin with the libertarian arguments against paternalism. Libertarianism has two primary oppositions to paternalism: the instrumental critique and the autonomy-violating critique.

The instrumental objection to paternalism says that paternalism is ineffective. In other words, paternalistic policies — policies that we will consider as government interventions — have a tendency to bring about worse outcomes than if individuals had the freedom to choose for themselves. Whether this failure is attributed to an ignorance of preferences or the diverging agendas of the state and its citizens, instrumental libertarians maintain that paternalism is consequentially ineffective. In a market context, for example, instrumental libertarians argue that “the common good is indirectly achieved through the actions of well-informed, rational and... self-regarding economic agents” (Rebonato 2015, 92). This consequentialist objection opposes any form of paternalism and, by extension, libertarian paternalism.

The second libertarian objection emphasizes the dignity of the autonomous choice. This is not a consequentialist argument. In this view, paternalism restrains an individual's ability to choose autonomously, thereby diminishing their dignity as human beings. The autonomy-autonomy-inspired libertarian may even concede that the outcome of a paternalistic intervention may be better than if the individual chose for themselves, but to them, the freedom of choice is what is important. As an echo of this sentiment, Mill writes that "it is possible that an individual might be guided in some good path, and kept out of harm's way,... but what will be his comparative worth as a human being"(Mill 1869, 55)? Therefore, for the autonomy-inspired libertarian, the value placed on autonomy is absolute.

With consideration of these two libertarian points of view, any form of paternalism is wrong, whether it judged by its means or ends. In regards to the autonomy-inspired objection, the question remains as to whether or not libertarian paternalistic policies like automatic enrollment allow individuals to retain their autonomy, as the program suggests it does.

Is the Interpersonal Intelligibility of Preferences Possible?

One of the most common lines of criticism against libertarian paternalism takes issue with how policymakers can possibly understand peoples' preferences. In order for a policy-maker to nudge someone in the direction that 'she herself would have chosen,' the policymaker must be able to understand her preferences. Libertarian paternalists largely accomplish this goal by referring to what a rational actor would do. Therefore, libertarian paternalists refer to the *Homo Economicus*, who is "unencumbered by cognitive limitations" and tries to "maximize his own welfare" (Rebonato 2015, 154). Rebonato argues that utility maximization and self-interest are

not necessary, however, in the determination of an individual's preferences. He continues points out the importance of revealed preferences in determining what an individual's rational preferences may look like. Similarly, Binmore writes that "if [an individual] acts consistently, she will act as though maximizing a utility function tailored to her own behavior" (Binmore 2009, 20). In this view, if an individual is behaving consistently, they are acting rationally, and actions are representative of rational preferences. Rebonato argues that libertarian paternalists fails because they must "guess [the] preferences by filtering out the 'noise' produced by the System-I self" (Rebonato 2015, 152), which is arguably a difficult task. Even so, in the context of retirement savings, libertarian paternalists have undertaken this challenge with policies like automatically enrollment, where they assume that people's true preference is to save more money.

A Reluctance to Debias

Rebonato asks the question, "are our cognitive limitations hard-wired" (Rebonato 2015, 210)? One of the central assumptions of the libertarian paternalism program is that the biases and cognitive inadequacies of an individual are treated as an "unchangeable biological given" (Rebonato 2015, 210). Also, as we have seen, libertarian paternalists argue that these aspects of our decision-making ability lead us to make poor decisions. In order for the nudge agenda to maintain an ethical justification, it is, therefore, important to determine the extent to which these cognitive inadequacies are biologically hard-wired. In other words, libertarian paternalists, in Rebonato's view, would need to explain why they prefer 'nudging' to 'teaching,' or another way of way of organizing choice settings so that it brings to fore the rational faculties of system-II (Rebonato 2015, 211). Applying this arguments to retirement savings nudges, one

could argue that automatic enrollment and SMT treat procrastination and the status-quo bias to be biologically hard-wired biases, making little effort to make people less susceptible to choice manipulation.

Rebonato continues and argues that Debiasing techniques could help people make better decisions, and even learn. Rebonato cites several papers that recommend how we can help people overcome these cognitive biases. When dealing with contextualization and framing problems, he cites Mitchell, who argues that we could encourage people to reframe choice problems into more personally meaningful frames, or encourage greater reflection and deliberation to help people overcome contextualization problems (Mitchell 2005, 9).

Rebonato also argues that when considering statistical information, especially conditional probabilities, cognitive biases have a very strong impact. In a relevant study, Gigerenzer and Edwards report that doctors with an average of 14 years of professional experience failed to determine the hypothetical probability that someone who tests positive for colorectal cancer actually has cancer. After being given the sensitivity and specificity of the cancer test, the doctors' estimates grossly overestimated the probability of cancer, with most of them providing answers of 45% and higher (Gigerenzer & Edwards 2003, 741). The actual probability of cancer in the hypothetical subject was about 5%. This result can largely be attributed to the base-rate neglect of the doctors, who failed to consider *how low the prevalence of cancer really is*. Using this example as evidence, Edwards and Gigerenzer argues that, yes, humans are vulnerable to cognitive biases, but this vulnerability is not strictly an internal cognitive weakness. He argues that "every piece of statistical information needs a representation... some forms tend to cloud minds, while others foster insight" (Rebonato 2015, 213). In other words, Rebonato argues that when information is presented as misleading, it will be misleading, but if it is presented in a way that helps people

overcome their cognitive biases, then it could help people learn, therefore “[fostering] insight” (Gigerenzer & Edwards 2003, 742).

The Slippery Slope Argument

One of the most common lines of criticism against libertarian paternalism, known as the slippery slope argument, says that if libertarian paternalistic policies are allowed, it could lead to more and more paternalistic policies and an overly powerful government. Edward Glaeser in article entitled *Paternalism and Psychology* argues that:

“Advocating soft paternalism is akin to advocating an increased role of the incumbent government as an agent of persuasion. Given how attractive it is to use persuasion for political advantage, an investment in soft paternalism seems to carry great risks...” (Glaeser 2005, 38)

Here Glaeser implicitly raises the question of how the governing party of a democracy should be kept in check by its opposition and citizens. Often times, a defense against overly paternalistic interventions is “enhancing visibility and transparency,” which allows citizens to understand the “nature and reach of the paternalistic intervention” (Rebonato 2015, 104). The trouble is that there is a fundamental difference between paternalistic interventions and those of the libertarian paternalist. Libertarian paternalistic policies are often implemented subtly and covertly, making them particularly dangerous. Glaeser and other opponents of libertarian

paternalism would argue that these supposedly ‘soft’ paternalistic interventions could strengthen an overly-powerful government.

Excessive Faith in Policymakers and Social Planners

The notion of bounded rationality has been a main justification for libertarian paternalistic interventions. Edward Glaeser uses the same argument against libertarian paternalism, stating that “consumers have a better incentives to overcome errors than government decision-makers” (Glaeser 2005, 133). In other words, he argues that if we accept that all human beings have a bounded rationality (i.e. display cognitive deficiencies), then politicians will also make decisional mistakes. Furthermore, he argues that humans have better incentives to overcome these decisional mistakes than government officials who are, too, bounded by their own rationality, as well as their preferences, goals and ideology. For these reasons, Glaeser argues that “recognizing the limits of human cognition may strengthen the case for limited government (Glaeser 2005, 98). This criticism of libertarian paternalism aligns closely with that of the epistemic argument, which insists that “people should be allowed to act on the basis of their own judgments” because policymakers and social planners “lack the information that individuals have” (Sunstein 2015, 91).

Having excessive faith in social planners is also dangerous because it may negatively impact social learning. In the context of defaults, for example, a decision-maker may be inclined to actively switch to a default because they consider the default option to be the expert opinion (Carlin et al. 2010). In some situations, defaults may be optimal, but in others, they may reduce social welfare. If people do not invest in gathering new information themselves, then people may

rely on the opinion of the social planner. The greater reliance on the social planner may reduce the “positive externalities from social learning” (Rebonato 2015, 123) and lead to unproductive homogeneity, especially if the social planner ‘gets it wrong’.

Libertarian Paternalism Is Not Soft Paternalism

Bringing us to our last criticism, in *The Manipulation of Choice*, Mark White argues that libertarian paternalism “is hard paternalism disguised as soft paternalism” (White 2013, 90). Libertarian paternalism purports to be soft paternalism, but it fails because it violates individual autonomy. As was previously outlined in chapter two, soft paternalism is the view that interventions are justified only under conditions when it is necessary to determine whether an individual or group is acting voluntarily and knowledgeably. Because it is impossible to know that person’s true interests, there is an assumption being made about what an individual truly prefers. Additionally, since nudges work by exploiting cognitive biases and heuristics, they work coercively and covertly, often without the person being nudged ever realizing they are being nudged. As a result, libertarian paternalists prevent the opportunity for poor decisions to be made, rendering libertarian paternalism hard paternalism. With consideration of this objection, automatic enrollment may not be entirely unobjectionable.

Chapter 4

The Libertarian Paternalistic Rebuttal

In this chapter several libertarian paternalistic rebuttals to the criticisms outlined in the previous chapter will be considered. This will, hopefully, capture the depth of the debate surrounding the ethics of nudging interventions and the extent to which libertarian paternalists have attempted to address some of the most noteworthy objections.

Rebuttals to Libertarian Objections

As we saw, libertarians argue that paternalism is ineffective and that, instrumentally, welfare is maximized without interventions. One could interpret this argument as a Hayekian challenge to paternalism. Hayek argued that the “dispersed nature of human knowledge and the informational advantages of markets” (Sunstein 2015, 93) are superior to that of any social planner or policymaker. This competition of knowledge, in effect, promotes welfare that any paternalistic intervention would hold back. The libertarian paternalist response to this argument says that “good choice architecture... should promote rather than squelch competition” (Sunstein 2015, 113). Sunstein continues and argues that we should keep “Hayekian objections in mind” as long as “choice architects provide information about warnings and risks” (Sunstein 2015, 113). Keeping this in mind, libertarian paternalists acknowledge the importance of competition and unregulated informational exchange, but emphasize the importance of ‘good’ choice architecture in stimulating competition and not restraining it. Richard Thaler in *Nudge* mentions that some critics “object to any forced exchanges” and that they don’t like “any government policy that

takes resources from some in order to assist other” (Sunstein & Thaler 2009, 248). In this view, paternalistic interventions do may require them to pay for a program that doesn’t benefit them, but instead someone in need. Paradoxically, this may only hinder competition. If people who need help are also imposing a cost on society — for example, through higher health costs — than helping them out a little bit might help them be more productive and competitive.

The other libertarian objection to paternalism argues that autonomy should be preserved. In this view, any paternalistic intervention limits an individual’s freedom to choose. Libertarian paternalists, however, argue that their interventions are freedom preserving and allow individuals to retain their autonomy because they are easily reversible. By making an intervention easily avoidable, the “low-cost opt out rights” (Thaler and Sunstein 2009, 236) are designed to help retain freedom of choice. Even so, I would argue this poses a difficult challenge. While many interventions may be nominally reversible, nudges are not altogether avoidable, especially if they on the exploitation of heuristics and biases.

Expressed Preferences

Plainly, some critics of libertarian paternalism argue that the only way to determine an individual’s preferences is to look at their revealed preferences. The libertarian paternalist would argue that revealed preferences may not be representative of an individual’s true preferences. So, they look at an individual’s expressed preferences, i.e. what people say that want to do, but cannot do. Libertarian paternalists argue that cognitive biases and heuristics explain why people cannot realize these expressed preferences. For example, inertia and procrastination are thought to prevent people from joining retirement plans.

Generally, behavioral economists take issue with revealed preference theory as a method of determining what constitutes an individual's preferences and, by extension, a rational decision. In fact, this very disagreement contributed to the development of behavioral economics in the 20th century. In 1938, Paul Samuelson attempted to re-construct consumer choice theory without reference to preferences or utility. He wanted to “develop a theory of consumer's behavior freed from any vestigial traces of the utility concept”(Hédoin 2014, 1). He defined a consistency condition which is now known as the weak axiom of revealed preference (WARP). WARP says that if a bundle of goods A is chosen over bundle B when both are affordable, then the consumer shows (reveals) that they prefer A over B . From this logic, WARP says that if these are the consumer's true preferences, then under no circumstances would the consumer prefer bundle B to bundle A . Sen disagreed with these basic axioms.

In the 1970's Sen began his criticism of Samuelson's revealed preference hypothesis, claiming that “man is a social animal” and his choices are “not rigidly bound to his own preferences only”(Sen 1973, 253). Sen argued that, every time an individual makes a decision, it is context-dependent. This means that the axioms of Revealed preference theory are “conceptually meaningless” (Hédoin 2014, 8) unless preferences are understood as mental states referring to an evaluation of personal welfare at a particular point in time. Also, Sen points out that consumers' preferences are often menu dependent. This means that the preference ordering of a person changes depending the set of available alternatives. In this effort, Sen effectively demonstrated that consumer preferences are not consistent. If, then, preferences are not stable and consistent, libertarian paternalists would likely argue that the task of interpersonal intelligibility is necessary if a nudge is to promote their welfare.

Is the Mistake Worth the Cost?

Rebonato proposed that most libertarian paternalistic policies are one-shot methods of preventing mistakes and do not help people learn. Since people learn by making mistakes, they should be allowed to make irrational decisions as a part of a learning process. Thaler and Sunstein argue that they are not preventing people from making mistakes. Instead, they argue that if people “really want to invest their entire retirement portfolio in high-tech Romanian stocks” (Sunstein & Thaler 2009, 240), they should certainly be allowed, but there is also little harm in putting “warning signs up along the way” (Sunstein & Thaler 2009, 240). In their view, learning is important, but there are some situations when the cost of making the mistake is so high, that a nudge may help prevent someone from making a terrible decision. For example, it may not be wise to let a child fall into the pool in hopes that he will emerge a great swimmer. In the effort of aiding in the learning process, libertarian paternalists claim that they are not trying to infantilize people. If interventions provide truthful information or helpful warnings, “it is not impeding learning” (Sunstein 2015, 114). Forms of hard paternalism, on the other hand, may impede learning because they might prohibit decision-makers from making the bad decision even if they are aware of the consequences.

Rebuttal to the Slippery Slope Argument

The slippery slope argument says that if some paternalistic interventions are allowed, then more will follow, until there is “an onslaught of intrusive forms of paternalism” (Sunstein & Thaler 2009, 235). Libertarian paternalists have three responses to this line of criticism. First,

they argue that supporters of the slippery slope argument are “[ducking] the question of whether [libertarian paternalistic] proposals have merit in and of themselves” (Sunstein & Thaler 2009, 235). Sunstein and Thaler believe that their proposals really do make people better off, whether it is by helping them save more, eat better, or choose better insurance plans. In short, libertarian paternalists believe that the slippery slope argument fails to criticize the proposals directly.

Second, libertarian paternalists argue that their proposals are “emphatically designed to retain freedom of choice” (Sunstein & Thaler 2009, 236). By maintaining the belief that proposals should be easily reversible, they argue that low-cost opt-out rights “reduce the steepness of the slippery slope” (Sunstein & Thaler 2009, 236). This line of defense is similar to their third rebuttal to the slippery-slope argument, which says that “it is pointless to ask the government to just stand aside” (Sunstein & Thaler 2009, 236). If choice architecture is inevitable, it would be unwise to let it impact people negatively, when it could be altered to someone’s advantage. In many ways, the onslaught of interventions has already begun, but it may just be hard to recognize. They argue that these interventions already exist, but they “so obvious and so sensible that that they do not see them as rules at all” (Sunstein & Thaler 2009, 237).

Chapter 5

The Ethics of Nudging in Retirement Savings

Having discussed the criticisms and defenses of libertarian paternalism at length, I would now like to discuss the ethics of libertarian paternalism policies in retirement savings. In this effort, I aim to contextualize the arguments that have been explored so far, and explore the extent to which nudges are ethical in the context of retirement savings. In the context of the ethical debate, I explore the ethical implications of automatic enrollment and the Save More Tomorrow plan. In doing so, I ask whether or not these interventions can successfully understand individual preferences, respect autonomy, and promote welfare in the long run.

The Ethics of Automatic Enrollment

First, it becomes necessary to determine the extent to which libertarian paternalists can successfully determine people's retirement savings preferences. Using revealed preferences as a method of doing so, it would seem as if people are already acting rationally, since people around the world are consistently revealing their preference to not save. This behavior could be further rationalized by arguing that people are choosing to actively consume today because any shift of expenditure into the future would not necessarily make them better off. The consumption expenditure of budget constrained households (and young households, especially) may have high marginal values (Zywicki 2017, 919). If this is true, then it may not be in a household's interest to shift expenditure into the future when they need it for consumption today. This argument not

only resonates with the revealed preference notion of rationality, but with that of the epistemic argument which says that households and individuals understand their preferences better than policymakers, and are thus better equipped to promote their own welfare in the context of retirement savings.

However, as was already mentioned, some people may not be saving as much as they want to. In this view, libertarian paternalists are looking at the expressed preferences of individuals rather than their revealed preferences. Automatic enrollment, therefore, is thought to help people realize their expressed preferences and overcome the negative consequences of procrastination, a bias to the status quo, and/or a lack of willpower. I would argue that, for the people who have expressed an interest in wanting to save more, automatic enrollment is a good method of helping them save more for retirement. As anyone who has lived a day on this planet knows, procrastination and a lack of willpower often prevent optimal decisions from being made. Still, how can a policymaker identify who wants to save more but can't, exactly?

Automatic enrollment is promoted as an asymmetrical form of paternalism, meaning it is designed to help the least sophisticated members of society, while imposing few costs on the most sophisticated (Benartzi & Thaler 2004). This argument would mean that those who are saving hardly any money would be the least sophisticated, and are thus deemed the people who should be nudged into saving more, and whose informed preferences would be to save more. But, as was already mentioned, it may be the case that these "unsophisticated" households are budget constrained and, therefore, would only be impacted negatively by shifting consumption into the future. So, in this way, automatically enrolling people would asymmetrically affect some households negatively.

Overall, it seems as if automatic enrollment takes individual preferences in retirement savings to be expressed preferences, or the preferences that would make an individual better off, ‘as judged by themselves’. Still, the trouble is that, without asking every single person whether or not they want to save more money, the task of determining an individual’s preferences is very difficult, if not impossible. In consideration of automatic enrollment as a method of addressing low enrollment rates, I would argue that, while it may help some people overcome a lack of will power and procrastination, it makes unfair assumptions about people’s preferences — preferences that cannot be understood by a planner or policymaker. Instead, automatic enrollment makes value substituting assumptions about people’s preferences (White 2013). For these reasons, I believe there are better alternatives to automatic enrollment that do a better job at understanding people’s preferences.

Next, Libertarian paternalism claims to be a form of soft paternalism. However, upon closer inspection, I would argue that this claim is disputable depending on the context. In this case of automatic enrollment, I argue that it is not soft paternalism. Let us revisit Mill’s example of the uninformed man who is about to cross a damaged bridge. In this example, the man who is about to cross the bridge lacks valuable information — information that could save his life. Thus, he could benefit from a paternalistic intervention. A soft paternalistic intervention would involve stopping the man because it is necessary to make sure that he is aware of the dangers that lie ahead of him. If, in light of receiving this information, he still wants to continue, then the soft paternalist should allow him to exercise his autonomy and continue, *even if this means certain danger and possible death* await him. On the other hand, a hard paternalist would prohibit the man from crossing even he were aware of the risks. In this case, the hard paternalist would be substituting

their own values for that of the man who wants to take risks, consequently preventing him from exercising his individual autonomy. In this hard paternalistic view, an assumption is being made that the man's true preference is to not cross the bridge. Even if he voluntarily wants to, say, commit suicide, the hard paternalist is assuming that this is not what he really wants.

With consideration of the bridge metaphor, I would argue that automatic enrollment is not soft paternalism. Soft paternalism argues that interventions are justified only under conditions when it is necessary to determine if an individual or group is acting voluntarily and knowledgeably. In the context of automatic enrollment, I would not say that it is *necessary* to determine whether or not an individual is voluntarily choosing to not participate in a retirement plan. By automatically enrolling an individual into a retirement plan, the policymaker is implicitly claiming that it is not possible to determine whether or not the person is acting voluntarily or knowledgeably. Moreover, since sure-danger lies ahead of this person, an intervention is justified. In this case, libertarian paternalists take low-participation rates and expressed preferences alone to mean that individuals must be mistaken or confused, and are thus unsure of whether or not they want to join a retirement plan. That being so, I would argue that automatic enrollment is hard paternalism because it may not be necessary to determine whether or not an individual is acting or voluntarily or not.

Secondly, under automatic enrollment, it is impossible to determine whether or not an individual is voluntarily choosing not to participate in a retirement plan. Even if it *were necessary* to determine whether or not an individual is voluntarily not participating in a plan, how could an opt-out nudge determine if an individual is acting voluntarily? Going back to Mill's bridge example, a soft paternalist must let the man cross the bridge if he is aware of its dangers and still wants to continue. In this case, it may be deemed unwise to cross the bridge, but the man is

entitled to exercise his autonomy. However, automatic enrollment — by design — is incapable of determining whether or not an individual is acting voluntarily. Plainly, it prevents people from making a decision.

In response to this argument, a libertarian paternalist may argue that, since the intervention is easily reversible, an individual's freedom is not being violated. At any point, the person who is automatically enrolled could ostensibly opt-out of the plan. However, I would argue that, since automatic enrollment functions by taking exploiting cognitive heuristics, and biases, it are sneaky and covert. In this way, individual autonomy is violated. Referring back to Gerald Dworkin's view of autonomy, it is "the capacity of a person to critically reflect upon, and then attempt to accept or change, his or her preferences, desires, values, and ideals" (Dworkin, 1988, p. 48). Using this definition of autonomy, we see that people are not being given the opportunity to accept or change their preferences because they are covertly steered into joining the plan via an exploitation of cognitive heuristics and biases. Often times, people may be unaware that they have been nudged into a retirement savings plan and are thereby robbed of an opportunity to critically reflect on their decision. As a result, they are coercively manipulated into doing something they may or may not want to do. Furthermore, once they are defaulted into the retirement savings plan, the status quo bias, that was previously preventing them from joining, is now working to keep them in.

Overall, I would argue that automatic enrollment is a poor example of soft paternalism because 1) there is not sufficient justification to determine whether or not people are acting voluntarily by not joining, 2) by defaulting people into a plan there is no way to determine whether or not they don't want to be defaulted, and 3) their autonomy is violated because it

exploits people's bias to the status quo. In consideration of these factors, I would say that it is merely paternalism with a mask. It isn't a nudge, it's a shove.

Finally, Libertarians would argue that people learn by making mistakes, so they should be allowed to make irrational decisions as a method of learning. Thaler and Sunstein respond to this common objection by arguing that there are some situations where the costs of making a mistake are so high that a nudge may help prevent a terrible decision. Since I find insufficient reason to believe that everyone needs to be nudged into saving more for retirement, it is not obvious that forgoing participant in a 401(k) plan is a 'terrible decision.' As I've mentioned, there are plenty of reasons why an employee may not want to join a retirement savings plan — reasons that a policymaker cannot understand. Moreover, Cass Sunstein writes himself that “with respect to learning, some forms of hard paternalism might run into strong objections” (Sunstein 2015, 114). As I have argued, I believe that automatic enrollment is not a form of soft paternalism so I agree with him entirely. Under automatic enrollment people are nudged away from making a mistake. Since people are covertly manipulated in a certain direction, it effectively prevents them from having the option of choosing what they want to choose, even if that decision would be a suboptimal decision. By defaulting people into a retirement savings plan, people are infantilized and denied of their right to make the decision that a policymaker has deemed irrational. In this view, automatic enrollment prevents people from having the opportunity to learn because the opportunity is blocked from the start.

Let us now pretend that not participating in a retirement savings plan is a terrible decision. Even if automatic enrollment were to increase an individual's well-being, there is reason

to believe that it could affect an individual's future preferences. Even though I have stressed the futility of determining an individual's true preferences, I would not argue that people's preferences are stable and consistent like revealed preference theory suggests. That being so, it is important to understand how actions not only reveal preferences, but create them (Norton & Ariely, 2008). A nudge at time t is likely to influence an individual's preferences at time $t + 1$ and onward (Binder 2014). Accordingly, automatic enrollment could have inter-temporal effects on people's preferences.

First, people who are automatically enrolled would lose the opportunity to learn about their preferences through cognitive reflection, thereby affecting what their future preferences could have been. Second, since individuals most often learn unknowingly through associative learning (Witt 2001), automatic enrollment could stealthily impact preference learning. Preferences learned via associative learning tend to be difficult to unlearn because of the low conscious involvement and repeated reinforcement over time (Zajonc and Markus, 1982). That being so, automatic enrollment may create preferences for sticking with defaults and putting faith in the 'expert' opinion of policymakers. This could lead to a dangerous situation where individual preferences become deeply entrenched and no longer subject to doubt. Also, as was discussed in chapter 3, this may lead to homogeneity and excessive faith in the policymaker that would negatively impact social learning and the positive externalities that come from it (Carlin et. al. 2010).

Also, on the topic of learning, I would argue that automatic enrollment is a great example of how many libertarian paternalistic interventions demonstrate a reluctance to de-bias decision making. By defaulting people into a retirement plan, policymakers are taking procrastination, hyperbolic discounting, a lack of willpower, and a bias to the status quo to be intrinsic and hard-

wired aspects of human cognition. In this view, not only could automatic enrollment affect preference learning, but it would not help people overcome their biases, effectively shaping new preferences and strengthening the impact of cognitive biases. Individuals could end up learning to be dependent and inactive, resulting in a reduction of their autonomy and active use of critical system II decision making.

Aside from learning, I would argue that by increasing participating rates in retirement savings plans, there could be several positive impacts. First, it would put less strain on social security — a program that is arguably inadequate as a primary means of provisioning for the future. Hypothetically, if someone were to fail to save for retirement because inertia and procrastination, they may end up economically disadvantaged. Moreover, it may also place a financial burden on that person's family, who now must forgo consumption, saving, or investment in order to take care of them, in turn affecting their ability to save for retirement. Keeping this in mind, it is definitely a good idea to encourage people to save, but automatic enrollment may not be the most ethical way to do it.

The Ethics of Save More Tomorrow™

In many ways, I would argue that Save More Tomorrow is much less manipulative than automatic enrollment. Under the SMT plan, employees are nudged into saving at a preset default contribution rate and asset allocation. Furthermore, employees' savings rates are escalated automatically to account for hyperbolic discounting. Since these contribution increases are synchronized with pay raises, people do not experience the negative effects of loss-aversion. Lastly, the same inertia that prevents people from joining the program is now leveraged toward

keeping them in the program. All of these features make up the SMT plan, making it markedly different than automatic enrollment, whose mechanism is more simple and, as we have seen, more coercive.

In regards to the intelligibility of people's preferences, the Save More Tomorrow plan is notably different than automatic enrollment. As a part of the SMT plan, "employees are approached" (Thaler & Benartzi 2004, 170) by a financial consultant and asked whether or not they want to join the plan. In this way, since employees are directly asked about their preferences, the task of determining their preferences becomes possible. The financial consultant tells the employee about the program and, through this process, full information is given to the employee. This means that when employees do make a decision on whether or not to join the SMT program, they are actively making a decision, thus revealing their preference to join or not join. They are asked to commit in advance to account for the negative effects of hyperbolic discounting, helping people overcome a strong time preference for current consumption. One could argue that this time preference is rational since employees have revealed their preferences by not joining. However, I would argue that procrastination is a powerful force, so by requiring people to make a decision, they can express their true preferences.

Unlike automatic enrollment, SMT is a clear example of soft paternalism. Employers intervene at an early stage and determine whether or not employees want to join the plan. In doing so, the employer would be able to determine whether or the employee wants to voluntarily make a suboptimal decision, which they are entitled to. Referring to Dworkin's view of autonomy, we see that employees have the opportunity to "attempt to accept or change their preferences" (Dworkin 1988, 48) at this stage. As such, employers are not using deception or coercion to influence behavior. Rather, they are encouraging critical reflection, allowing people to

engage their system II and make an autonomous decision. Applying Mill's bridge example to SMT, we see that employees are intercepted preemptively and warned of the dangers of the bridge — that is, the danger of not saving for retirement. Thus, individual autonomy is not violated.

In regards to promoting welfare over time, I would argue that, since the Save More Tomorrow program gives employees the opportunity to decide, they could learn via cognitive reflection about their decision. However, just like with automatic enrollment, those who do choose to participate in the program would be subject to associative learning, where their preference for the default would be reinforced over time. This would have the same effect on inter temporal preferences as automatic enrollment, functioning also as a one-shot intervention not aimed at helping people make a critical decision and overcome cognitive biases. By enrolling an employee into the program, they may passively develop a preference and dependence on the default. In effect, this would lead to an “inert acceptance of a default option” in which the brains of decision makers will “become ill-equipped to reason rationally and critically about the next choice” (Rebonato 2015, 219). Since SMT succeeds by exploiting how people fail to distinguish between nominal and real income, it is merely changing “the optics of the scheme” by making sure that employees never see a decrease in their income in nominal terms, even though “the purchasing power of the take home pay check decreases year after year” (Rebonato 2015, 161). People often fail to understand the effects of inflation, which is a deeply ingrained bias (Shiller 2008). As such, SMT fails to help people overcome this bias. There may be ways, however, to help people overcome these biases and learn to make better decisions over time that still incorporate the automatic escalation of pay raises.

Overall, SMT is a more ethical approach to promoting retirement savings because it is more successful in determining preferences and promoting autonomy. Even so, it may still lead people to develop a dependence on defaults. Lastly, if SMT is combined with automatic enrollment, as it is often found, then I would argue the merits of the intervention are compromised. In this way, the plan would not only be coercive and deceptive, but it could exacerbate dependence on default plans, thereby reducing future decisional autonomy even more than automatic enrollment does on its own.

Chapter 6

Ethical Interventions in Retirement Savings

In this chapter, I aim to provide a conclusive recommendation on how employees could be ethically helped to save more for retirement. In this effort, I first emphasize the importance of knowing people's preferences and not making value substituting assumptions. Second, I emphasize the importance of making sure that individual autonomy is not violated, i.e. ensuring that people have the opportunity to accept or change their preferences and make a reflective system-II decision. Lastly, I emphasize the importance of promoting individual welfare over time by helping people learn, not make 'inert' decisions, and overcome cognitive biases. Overall, I argue that there are ways to ethically help people to save more for retirement, where their autonomy is not violated, but rather, enhanced.

To start, we must return to the central question of how a policymaker can determine an individual's preferences. Making sure that an intervention does not make value-substituting assumptions about individual preferences is key. Thaler and sunstein argue that automatic enrollment and Save More Tomorrow are aimed to help people whose expressed preferences are to save more, even though their revealed preferences say otherwise. In order to help these people, I would argue that there are better ways than using opt-out default nudges.

Naturally, the first way to address the issue of determining individual preferences would be to require employees to make a decision. This method is known as required active choosing and it requires new employees to choose whether or not they want to be in the plan. If they choose to be in a plan, they would also be required to choose a contribution rate. This method is useful because it gives employees the freedom to choose the plan that they prefer. Also, required

active choosing “encourages agents to think about an important decision and thereby avoid procrastinating” (Carroll et. al. 2005, pg. 3). Therefore, required active choosing may be effective in helping people overcome decisional inertia preventing them from realizing their expressed preferences. It has also been found to be more effective than standard opt-in approaches. For instance, one company that switched from the standard opt-in approach to required active choosing saw a 25 percent increase in participation rates (Carroll et. al. 2005). On the surface, this may seem like an ideal solution, but as Cass Sunstein would argue, required active choosing is a still form of paternalism because some people may choose not to choose (Sunstein 2014). For one reason or another, some people may prefer to not make a decision. Some people might fear that they will make a bad decision. Others may be aware of their lack of information or fear the complexity of the task. Some people might even hate the task of choosing. For these reasons, requiring people to choose is still a form of paternalism that could distort individual preferences.

As an alternative to required active choosing and automatic enrollment, I would argue that the best way to help people whose expressed preferences are to save more is to make the enrollment process easier. This can be done by reducing the complexity of the decision process and by giving people an opportunity to easily join the plan they want to join. In a 2006 study, researchers looked at the effectiveness of a low-cost manipulation designed to simplify the enrollment process called Quick Enrollment™. Under this approach, employees are given the option of opting-in to a default contribution rate and asset allocation. This design reduces complexity and helps employees evaluate a smaller subset of options (Choi et. at. 2006). Using this method, people can easily realize their expressed preferences and overcome procrastination and inertia. At one company, this method tripled participation rates among new employees within three months of being hired (Choi et. at. 2006). This finding shows that Quick

Enrollment™ is an effective method of helping people save, and it does not make value substituting assumptions about employees' preferences.

Under Quick Enrollment™, however, employees are enrolled at a preselected contribution rate and asset allocation. In this way, some decisions are still being made for the employee. Perhaps a good way to address this problem would be to give employees — while still keeping the enrollment process simple — a few options with variable contribution rates. Even more, employees could be assisted by a financial advisor who could help them pick the plan that best fits their expressed preferences.

In regards to asset allocation, this turns out to be a much more complex task, so I would argue that it not a bad idea for employees to stick with the default. Benartzi and Thaler investigated whether participants are good at picking a portfolio, as judged by themselves (Benartzi and Thaler 2002). In this study, subjects were presented with three unlabeled distributions of projected retirement income and asked to rate them on a scale of one to five, with five being the most attractive. The authors found that self-constructed portfolios received the lowest average rating, while professionally managed portfolios received the highest. To add, among the people who expressed a preference to pick their own portfolios, 80 percent of them found the professionally managed accounts more attractive (Benartzi & Thaler 2002). This finding highlights the difficulty of constructing portfolios and that people would likely be better off having someone else make the decision for them. Still, some employees may still insist that they want to construct their own portfolio. For these individuals, it may be useful to offer “lifestyle” funds that “blend stocks and bonds in a way designed to meet the needs of different levels of risk tolerance” (Benartzi & Thaler 2004, 88). For instance, employees could choose from a conservative, moderate, or aggressive portfolio.

With consideration of these approaches, I maintain the assertion that an ethical method of determining people's preferences would be to allow them to choose for themselves, while emphasizing simplicity of the enrollment process. If people are truly so strongly impacted by a lack of willpower, procrastination, hyperbolic discounting, etc. then making the enrollment process easier would certainly help them overcome these hindrances. In doing so, value substituting assumptions would not be made and policymakers could reasonably determine employee preferences. In practice, this could promote individual welfare and even enhance individual autonomy.

Next, I argue that interventions should not violate individual autonomy. Even more, they should not only aim to not violate autonomy, but encourage it. In Dworkin's view, if autonomy is to be preserved, individuals should be given the opportunity to accept or change their preferences. Therefore, in the context of retirement savings interventions, efforts should be made to encourage critical cognitive reflection, thereby giving people an opportunity to accept or change their preference. One of the best ways to do this would be to increase transparency, thereby preventing deception or coercion. Another way to encourage individual autonomy would be to enact policies that encourage system II decision making and reduce the negative effects of cognitive biases and heuristics.

Under automatic enrollment, decision-makers are subtly manipulated into joining retirement savings plans through the exploitation of heuristics and cognitive biases — a process that largely lacks transparency. Rather, it coercively manipulates individuals into joining plans they may not want to join because, by design, automatic enrollment is covert and relies on the

“yeah, whatever heuristic” (Sunstein & Thaler 2009, 35) to increase 401 (k) participating rates. Keeping this in mind, increasing transparency could encourage individual autonomy.

There are several ways policymakers could increase transparency. First, policymakers could use a method like Quick Enrollment™, where employees are not tricked into enrolling. Next, policymakers could increase transparency of goals. In the context of retirement savings, policymakers could inform employees that savings rates are typically quite low and it may be in their interest to join. Lastly, policymakers could increase transparency of means. This could entail informing employees of the enrollment method that is being used. If a Quick Enrollment™ policy were being used, for instance, employers could inform employees of how the system works and that, if they do join, they will be enrolled at a default contribution rate and asset allocation.

Another way to increase decisional autonomy in 401(k) enrollment is to use de-biasing techniques. These techniques could reduce the negative impact of cognitive errors on decision making, thereby increasing autonomy. One of the best ways to do this is to simplify the enrollment process. By presenting information in a simple and easy to understand format, people can more effectively call upon the rational faculties of system-II. In many situations, “information needs a representation — that is, a form” (Gigerenzer & Edwards 2003, 741) that is not misleading, but rather, easy to understand. By reducing the complexity of a problem, people may have more success in exercising their autonomy. Another method of promoting autonomy is to reduce the choice set to a manageable size such that a decision-maker’s ability to autonomously make a decision is not compromised. Since autonomy is not merely the nominal freedom to choose, it may not be the case that providing every choice option would increase autonomy. In fact, “freedom of this sort is ironically constraining” (Trout 2005, 416), as it is likely to

overwhelm decision makers and make it harder for them to authentically accept or change their preferences. In a relevant study, Sheena Iyengar explored the effects of limited choice on 401(k) enrollment rates and found that the more retirement plan options employees are offered, the lower the participation rates (Lyengar et al. 2004). Keeping this in mind, I would argue that by reducing the choice set and making the enrollment process easier, autonomy can be enhanced.

Meanwhile, another method of promoting autonomy in retirement savings would be enact policies that foster system-II decision making. Policymakers could implement policies such as required active choosing that call to the fore the rational faculties of system-II. However, as I previously mentioned, since required active choosing is still paternalistic, it is important that simplification of the enrollment process is maintained. I would argue that something like Quick Enrollment™ could be implemented, but instead of opting-in to a default, employees could pick from a few simplified options, thereby retaining their freedom to choose while simultaneously encouraging critical, system-II thinking.

Along the same lines of promoting decisional autonomy, is the importance of helping people make better decisions in the long-run, thus creating opportunities for people to promote their own welfare in the future. In this view, interventions aimed at increasing retirement savings should not be one-shot interventions aimed at helping an employee at time t , but also at time $t + 1$ and onwards. As it turns out, one of the best ways to promote welfare in the long run would be to employ similar methods used to promote individual autonomy at time t , as we have covered. These methods are largely aimed to help people reduce the effects of their cognitive errors as well as promote the activation of system-II faculties, therefore preventing an inert dependence on defaults and the expert opinion of policymakers.

One method of helping people make decisions over-time would be to give them the opportunity to change their retirement savings plan without penalty in the future. These are known as “cooling off” periods, and would give employees not only the opportunity to consider the salient features of a retirement plan at time t , but “also those that will more import when the outcomes are experienced” (Johnson et al. 2012, 498). As a policy recommendation, perhaps employees could be approached a few months after a retirement savings decision has been made and asked about whether or not they want to stick with it, drop the plan, or change it. In this way, employees would not develop a bias to the status quo that may develop during the cooling off period. This process, in its entirety, could encourage thoughtfulness and active cognitive reflection their retirement savings plan, therefore fostering insight and helping employees learn.

Finally, another way of helping promote welfare in the long run could be to directly teach employees about cognitive biases that have the potential to affect decisions at time t and beyond. For instance, even if employees were enrolled into a plan through something like Quick Enrollment™ or automatic enrollment, they could be taught about the effects cognitive biases, such as the exponential growth bias. Exponential-growth bias is a perceptual bias that manifests from an underestimation of how value grows exponentially with compounding interest. In a relevant study, researchers argued that a misunderstanding of these biases, particularly the exponential-growth bias, significantly contributes to low retirement savings (Goda et. al. 2015). If people could be reminded of the compounding effects of exponential growth, the effects of the biases could be mitigated, thereby increasing the likelihood of increased retirement plan participation rates.

Conclusion

To summarize what has been discussed thus far, I would begin by saying that nudging is not without its merits. That being said, it is important to remember when it is useful and when it is not. In the context of retirement savings, as we have seen, there is undoubtedly room for people to make better decisions — decisions that could have a lasting impact on their well-being. In this effort of helping people make better decisions, I maintain that it should be done ethically.

As I have emphasized over the course of this paper, it is essential that policymakers be successful in determining people's preferences. In retirement savings interventions, libertarian paternalistic policies have taken the expressed preferences of some people to represent the true preferences of most. Surely, this view is not without justification, as humans are undoubtedly guilty of fallibility. In this fallibility, people become victims of biases and blunders, neglecting the vital task of provisioning for the future. Even so, I maintain the assertion that, in policymaking, value substituting assumptions of individuals' preferences should be avoided at all costs. If and when possible, a decision-maker should be asked to think of her preferences, for she is most keenly aware of what is important and necessary in her life.

Furthermore, it remains imperative to respect the dignity of individuals whose right to autonomy is absolute. In regards to retirement savings plans, people should be given the chance to make a thoughtful decision. In this effort, they should not be coerced or deceived through an exploitation of biases that already stand to threaten autonomy. If anything has become clear as a result of this investigation, it is the fragility of individual autonomy. The act of making an autonomous decision is a challenging endeavor, since one must reflect upon their "preferences, desires, values, and ideals" (Dworkin 1988, 48), and then attempt to accept or change them. That

being so, decision makers should never be denied the opportunity to exercise their autonomy, as it not an easy task. Instead, decision-makers should be urged to reflect on their preferences, desires, values and ideals. In this effort, people are gifted the opportunity to learn. After all, we all have the right to be foolish” (Ludwig von Mises, 1979).

References

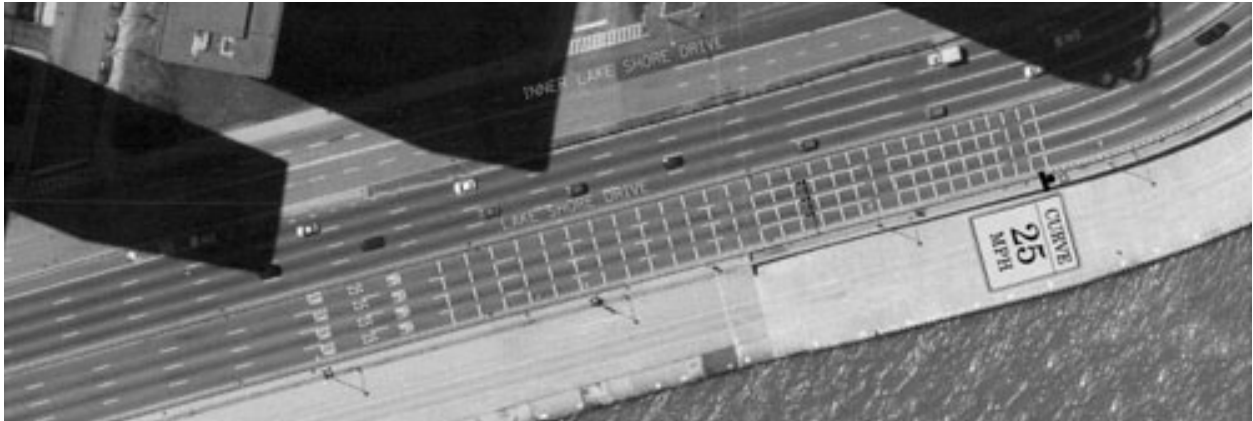
- Ariely, D., & Norton, M. I. (2008). How actions create – not just reveal – preferences. *Trends in Cognitive Sciences*, 12(1), 13-16. doi:10.1016/j.tics.2007.10.008
- Benartzi, S., & Thaler, R. H. (2002). How Much Is Investor Autonomy Worth? *The Journal of Finance*, 57(4), 1593-1616. doi:10.1111/1540-6261.00472
- Benartzi, S., & Thaler, R. H. (2007). Heuristics and Biases in Retirement Savings Behavior. *Journal of Economic Perspectives*, 21(3), 81-104. doi:10.1257/jep.21.3.81
- Binder, M., & Lades, L. K. (2014). Autonomy-Enhancing Paternalism. *SSRN Electronic Journal*. doi:10.2139/ssrn.2436928
- Binmore, K. G. (2009). *Rational decisions*. Princeton: Princeton University Press.
- Busse, M., Pope, D., Pope, J., & Silva-Risso, J. (2012). Projection Bias in the Car and Housing Markets. doi:10.3386/w18212
- Carlin, B. I., Gervais, S., & Manso, G. (2010). Libertarian Paternalism, Information Sharing, and Financial Decision-Making. *SSRN Electronic Journal*. doi:10.2139/ssrn.1570158
- Carroll, G., Choi, J., Laibson, D., Madrian, B., & Metrick, A. (2005). Optimal Defaults and Active Decisions. doi:10.3386/w11074
- Choi, James J., David Laibson, Brigitte Madrian, and Andrew Metrick. 2004. “For Better or For Worse: Default Effects and 401(k) Savings Behavior.” In *Perspectives in the Economics of Aging*, ed. David Wise, 81–121. University of Chicago Press
- Choi, J., Laibson, D., & Madrian, B. (2006). Reducing the Complexity Costs of 401(k) Participation Through Quick Enrollment(TM). doi:10.3386/w11979
- Christman, J. (2007). Autonomy, History, and the Subject of Justice. *Social Theory and Practice*, 33(1), 1-26. doi:10.5840/soctheorpract200733133
- Dalen, H. P., & Henkens, K. (2014). Comparing the effects of defaults in organ donation systems. *Social Science & Medicine*, 106, 137-142. doi:10.1016/j.socscimed.2014.01.052
- Dworkin, G. (1988). *The nature of autonomy*. Cambridge: Cambridge University Press.
- Dworkin, G. (2017, February 12). Paternalism. Retrieved April 30, 2019, from <https://plato.stanford.edu/entries/paternalism/>

- Ghilarducci, T. (2018). *When I'm sixty-four: The plot against pensions and the plan to save them*. Princeton: Princeton University Press.
- Gigerenzer, G., & Edwards, A. (2003). Simple tools for understanding risks: From innumeracy to insight. *Bmj*, 327(7417), 741-744. doi:10.1136/bmj.327.7417.741
- Glaeser, E. (2005). Paternalism and Psychology. doi:10.3386/w11789
- Goda, G. S., Levy, M., Manchester, C. F., Sojourner, A., & Tasoff, J. (2015). The Role of Time Preferences and Exponential-Growth Bias in Retirement Savings. doi:10.3386/w21482
- Hédoin, C. (2016). Sen's criticism of revealed preference theory and its 'neo-samuelsonian critique': A methodological and theoretical assessment. *Journal of Economic Methodology*, 23(4), 349-373. doi:10.1080/1350178x.2016.1218530
- Homonoff, T. A. (2018). Can Small Incentives Have Large Effects? The Impact of Taxes versus Bonuses on Disposable Bag Use. *American Economic Journal: Economic Policy*, 10(4), 177-210. doi:10.1257/pol.20150261
- Johnson, E. J., Shu, S. B., Dellaert, B. G., Fox, C., Goldstein, D. G., Häubl, G., . . . Weber, E. U. (2012). Beyond nudges: Tools of a choice architecture. *Marketing Letters*, 23(2), 487-504. doi:10.1007/s11002-012-9186-1
- Kahneman, D. (2015). *Thinking, fast and slow*. New York: Farrar, Straus and Giroux.
- Keysar, B., Hayakawa, S. L., & An, S. G. (2012). The Foreign-Language Effect. *Psychological Science*, 23(6), 661-668. doi:10.1177/0956797611432178
- Madrian, B., & Shea, D. (2000). The Power of Suggestion: Inertia in 401(k) Participation and Savings Behavior. doi:10.3386/w7682
- Mill, J. S. (2001). *On liberty*. Kitchener, Ont.: Batoche Books.
- Mitchell, G. (2005). Libertarian paternalism is an oxymoron. *Northwestern University Law Review*, 99(3), 1245-1278.
- Mises, L. V. (1979). *Economic policy: Thoughts for today and tomorrow*. Liberty fund: Indianapolis.
- Rebonato, R. (2015). *Taking Liberties A Critical Examination of Libertarian Paternalism*. New York, NY: Palgrave Macmillan.

- Sen, A. (1973). Behaviour and the Concept of Preference. *Economica*, 40(159), 241. doi: 10.2307/2552796
- Sethi-Iyengar, S., Huberman, G., & Jiang, G. (2004). How Much Choice is Too Much? Contributions to 401(k) Retirement Plans. *Pension Design and Structure*, 83-96. doi: 10.1093/0199273391.003.0005
- Shiller, R. J. (2008). *The subprime solution*. Princeton: Princeton Univ. Press.
- Strotz, R. H. (1955). Myopia and Inconsistency in Dynamic Utility Maximization. *The Review of Economic Studies*, 23(3), 165. doi:10.2307/2295722
- Sunstein, C. R. (2014). Requiring Active Choosing is a Form of Paternalism. *SSRN Electronic Journal*. doi:10.2139/ssrn.2400532
- Sunstein, C. R. (2015). *Why nudge?: The politics of libertarian paternalism*. New Haven: Yale University Press.
- Thaler, R. H., & Sunstein, C. R. (2009). *Nudge: Improving decisions about health, wealth, and happiness*. New York, NY: Penguin Books.
- Thaler, R., & Benartzi, S. (2004). Save More Tomorrow™: Using Behavioral Economics to Increase Employee Saving. *Journal of Political Economy*, 112(S1). doi:10.1086/380085
- Trout, J. D. (2005). Paternalism and Cognitive Bias. *Law and Philosophy*, 24(4), 393-434. doi: 10.1007/s10982-004-8197-3
- Tversky, A., & Kahneman, D. (1991). Loss Aversion in Riskless Choice: A Reference-Dependent Model. *The Quarterly Journal of Economics*, 106(4), 1039-1061. doi:10.2307/2937956
- White, M. D. (2013). *The manipulation of choice: Ethics and libertarian paternalism*. New York: Palgrave Macmillan.
- Witt, U. (2001). Learning to consume — A theory of wants and the growth of demand. *Escaping Satiation*, 29-42. doi:10.1007/978-3-662-04528-2_3
- Zywicki, T. J. (2017). Do Americans Really Save Too Little and Should We Nudge Them to Save More? The Ethics of Nudging Retirement Savings. *SSRN Electronic Journal*. doi:10.2139/ssrn.2901173

Appendix

Figure 1



(Source: Sunstein, Cass R., and Richard H. Thaler. *Nudge: Improving Decisions About Health, Wealth and Happiness*. Penguin, 2012.)