

University of Missouri, St. Louis
IRL @ UMSL

Theses

Graduate Works

9-24-2010

Evaluating Libertarian Paternalism

Michelle Elizabeth Catalano

University of Missouri-St. Louis, mekfyb@umsl.edu

Follow this and additional works at: <http://irl.umsl.edu/thesis>

Recommended Citation

Catalano, Michelle Elizabeth, "Evaluating Libertarian Paternalism" (2010). *Theses*. 213.
<http://irl.umsl.edu/thesis/213>

This Thesis is brought to you for free and open access by the Graduate Works at IRL @ UMSL. It has been accepted for inclusion in Theses by an authorized administrator of IRL @ UMSL. For more information, please contact marvinh@umsl.edu.

Evaluating Libertarian Paternalism

Michelle E. Catalano

MBA, Business Administration, Southern Illinois University Edwardsville, 2005

BS, Business Administration, Southern Illinois University Edwardsville, 2004

A Thesis Submitted to The Graduate School at the University of Missouri – St. Louis in
partial fulfillment of the requirements for the degree
Master of Arts in Philosophy

September, 2010

Advisory Committee

Anna Alexandrova, Ph.D.
Chairperson

Stephanie Ross, Ph.D.

Eric Wiland, Ph.D.

ABSTRACT

The purpose of this thesis is to evaluate a form of soft paternalism put forth by Richard Thaler and Cass Sunstein known as libertarian paternalism. On their view, criteria are provided for distinguishing between justified and unjustified libertarian paternalism policy proposals. The central philosophical issue surrounding libertarian paternalism is determining when limitations on autonomy for the sake of beneficence are justified. However, the criteria given by Thaler and Sunstein are inadequate because they do not effectively overcome problematic case examples. I contend that an additional autonomy-based condition should be added to their criteria. Specifically, autonomy may trump beneficence in favor of rejection of a policy if certain crucial autonomy considerations are violated.

Introduction

A growing body of literature in economics and law is challenging the notion that people are generally competent decision makers with regard to their own wellbeing. Failures in rationality at the level of judgments and choices are being systematically documented. In response to this, there are optimistic perspectives, within both economics and philosophy, embracing the possibility that paternalist measures can counteract limitations on rationality and bring about welfare improvements in society. The overall purpose of this thesis is to explore a form of soft paternalism, called *libertarian paternalism*, which has been the topic of many recent discussions on the direction of future public policy. Libertarian paternalism has been popularized by a prominent economist Richard Thaler and a no less prominent legal scholar Cass Sunstein in their book, *Nudge*. The specific aim of this thesis is to develop a way of distinguishing justified libertarian paternalist proposals from the unjustified ones in specific policy decisions by investigating two problematic case examples. My view is that Thaler and Sunstein's criteria are inadequate because they do not effectively overcome problematic case examples and that an additional condition should be added to their criteria. The condition is that autonomy trumps beneficence to reject policies if certain crucial, but not all, autonomy considerations are violated.

Thaler and Sunstein refer to their libertarian paternalist proposals as nudges. The theory behind libertarian paternalism is similar to other contending theories by different authors including "asymmetric paternalism" (Camerer, et al., 2003) and "new interventionism" (Lowenstein & Ubel, 2008). As defined by Thaler and Sunstein, a

nudge is “any aspect of choice architecture that alters people’s behavior in a predictable way without forbidding any options or significantly changing their economic incentives” (Nudge, 6). They classify their theory as libertarian paternalism because it attempts to unite a traditional economists’ commitment to consumer sovereignty and free markets with an element of institutional intervention. They contend that a libertarian variety of paternalism is coherent, even though seemingly contradictory, because choices “are not blocked, fenced off, or significantly burdened” (Nudge, 5). Actually, their original, seminal article on the subject boasts about libertarian paternalism that it is “not an oxymoron” (Sunstein & Thaler, 2003). On their view, libertarian paternalism consists of interventions that are “easy and cheap to avoid” and that are never mandates (Nudge, 6). At the same time, however, libertarian paternalist interventions are designed to paternalistically “steer people’s choices in directions that will improve their lives” by the standards that they themselves would endorse through “self-conscious efforts” implemented by private and public institutions (Nudge, 5).

Some examples of nudge policies are put forth by Thaler and Sunstein. A commonly referenced example provided in the introduction of their book pertains to the arrangement of food choices in a school cafeteria. In this example, a cafeteria director would arrange the food choices such that the healthy options are more frequently chosen over the unhealthy options. The cafeteria director deliberately chooses the arrangement that will encourage students to choose the healthy options rather than an arrangement that is random, that encourages students to select the most popular food choices, or that maximizes profit. In this way, the cafeteria director is nudging to encourage healthier

eating habits. Another prototypical example of a nudge is called “Save More Tomorrow” which encourages employees to save more money for retirement via defined contribution plans through their employer. When employees enroll in this program, they commit to allowing an automatic increase in their retirement contribution rate for every subsequent pay raise. Employees are enrolled by default but they are not required to join the program and do not incur penalties for opting out. Essentially, employees are nudged towards earmarking more of their earnings for retirement. A third example of a nudge proposal is to increase organ donations by changing the standard (or default) rule from explicit consent (or opt-in) to either presumed consent (or opt-out) or mandated choice (everyone is required to make a choice one way or another). Either way, making the change away from explicit consent would be a nudge towards saving the lives of many awaiting organ transplants.

The remainder of this thesis is divided into the following sections. First, the concept of paternalism as well as the different types of paternalism will be explained in detail. Second, the rationality findings from behavioral economics related to libertarian paternalism will be summarized. Third, the general objections to libertarian paternalism will be discussed. Fourth, the specific case problems of subliminal messages and the cocaine vaccine, which are of primary interest, will be described in detail. Next, I will present the conditions which must be evaluated in order for a nudge to be justified and show how the specific problems facing libertarian paternalism can be overcome. Last, this thesis will end with concluding remarks.

Philosophical Issues

The all-encompassing definition of paternalism is “the interference of a state or an individual with another person, against their will, and justified by a claim that the person interfered with will be better off or protected from harm” (Dworkin, 2005). At this point, it will be helpful to more narrowly specify the sort of libertarian paternalism that nudge policies fall under. To put the rest of this discussion in context, libertarian paternalism can be defined by reference to the standard distinctions in the philosophical literature on paternalism.

Types of paternalism

The first distinction is between hard and soft paternalism. Hard paternalism allows for paternalistic acts that prevent a person from doing harm to themselves even if they are acting voluntarily and with full information about the negative consequences of their action (Dworkin, 2005). On the other hand, soft paternalism only allows for intervention up to the point of determining “whether the person being interfered with is acting voluntarily and knowledgably” (Dworkin, 2005). The classic case used to delineate soft paternalism from hard paternalism is Mill’s bad bridge example in which a person is about to walk across a damaged bridge. A soft paternalist would condone using force to stop the person from crossing the bridge to find out whether they know the condition of the bridge. Beyond that, a soft paternalist may not stop the person from crossing, for instance, if he knows the condition yet willingly chooses to cross and risk his life. Conversely, a hard paternalist would condone stopping the person against their will even after they are informed about the condition of the bridge. As for nudges, they

are classified soft paternalism because they do not eliminate the option for an individual to make a determined choice even if it is not in their own self-interest.

The second distinction is between broad and narrow paternalism. Broad paternalism is concerned with paternalistic acts by any individual, group, or institution (Dworkin, 2005). Narrow paternalism is only concerned “with the question of coercion by the state” (Dworkin, 2005). Even though Thaler and Sunstein maintain that nudges are desirable and advantageous in the private sector, their objective is to justify nudges in the public sector. For the purposes of this thesis, nudges will be defined as narrow because the issues being debated only deal with governmental policy implications. However, in other applications, nudges could be considered broad because they are intended to have usefulness in the private sector as well. The third distinction is between weak and strong paternalism. Weak paternalism states that “it is legitimate to interfere with the means that agents choose to achieve their ends, if those means are likely to defeat those ends” (Dworkin, 2005). Strong paternalism states that it is legitimate to include instances of interference when the ends themselves are determined to be mistaken or irrational (Dworkin, 2005). Nudges could be considered either weak or strong paternalism. For example, if people truly prefer to eat *well* then it is legitimate, on the weak view, to rearrange food in a cafeteria setting in such a way that the healthier options are more frequently chosen. In this scenario, a nudges are weak because they only interfere with the means that people use to make choices while preserving their end value of eating healthy. On the other hand, nudges could be considered strong paternalism because they attempt to modify other people’s mistaken preferences by way of

persuasively offering different alternatives. As in the same cafeteria example, rearranging food options for people that truly prefer to eat *poorly* would be an attempt to interfere with their end value of eating unhealthy by encouraging the exact opposite. In this sense, nudges are strong because they seek to change values for what are considered mistaken ends (i.e. eating unhealthy).

The fourth distinction is between pure and impure paternalism. Pure paternalism would be imposing interventions on the same group that is being protected or helped (Dworkin, 2005). Impure paternalism would be imposing interventions on a group for the sake of protecting or helping an entirely different group (Dworkin, 2005). Nudges are pure because they target the same group that is being protected or helped. For instance, the Save More Tomorrow program is directed towards individuals that want to save more money for themselves in retirement. At first glance, many nudges, particularly ones with default-rules, may seem as if they are targeting larger audiences than they are protecting because nudges are typically wide ranging and have large-scale application such as those that utilize clever ad campaigns (i.e. “Don’t mess with Texas”). But, nudges are non-mandatory so anyone that doesn’t want to be subjected to intervention can decline to participate or opt-out. So, nudges are not impure in that there are no required interventions imposed on certain people for the sake of helping entirely different people. The fifth distinction is between moral and welfare paternalism. Moral paternalism is the view that the purpose of intervention is to protect people against moral corruption (Dworkin, 2005). Welfare paternalism is the view that the purpose of intervention is to make people’s lives go better (Dworkin, 2005). There is no agenda to promote the

“moral wellbeing” of society in nudge policies; thus, they are welfare-based. In summary, libertarian paternalism falls into some of the comparatively benign distinctions of paternalism. Now, we can shift our discussion to concern about when we should be justified in implementing nudges.

Autonomy and beneficence

The basic philosophical tension surrounding paternalism taken as a whole hinges on concerns about when limitations on freedom or autonomy are justified. There have been numerous discussions attempting to set apart exactly “when a paternalistically motivated social policy or individual action interfering with persons for their own good and without their consent is justified” (Brock, 1988). More directly, paternalism must confront the two conflicting aims of preserving individual autonomy while at the same time promoting beneficence. One widely accepted view of paternalism is that autonomy must always trump beneficence when there is a conflict between the two (Feinberg, 1986; VanDeVeer, 1986; Glover, 1977; Arneson, 1980). This stance rules out hard paternalism and is taken by those who want to “avoid an unwarranted endorsement of excessive paternalism” (Brock, 1988). The opposite view in which beneficence should trump autonomy when there is a conflict between the two could also be argued. This stance is taken by those who believe that paternalistic intervention is needed in order to help people make decisions that are in their best interest. It permits hard paternalism which would allow interventions against people’s will by using deception, coercion, and violations of autonomy (Scoccia, 2008). Yet another view is a middle option between the opposing views and it is called the balancing view (Brock, 1983; Brock, 1988; Dworkin,

1971). The balancing view puts forth the commonsense notion that there need not be an overarching commitment to autonomy or beneficence in all cases whenever there is conflict. Rather, these values must be weighted and chosen on a case-by-case basis. Sometimes, we may determine that a person's wellbeing takes precedence over their autonomy; yet, at other times, we may determine that their autonomy takes precedence over their wellbeing.

In this thesis, it will be argued that paternalism is legitimate in some, although not all, cases. In particular, I find the balancing view based on commonsense to be compelling because in the public policy process, proposals are evaluated on a case-by-case basis. However, the practical difficulty is whether we can distinguish the exact conditions in which soft paternalism is justified. This thesis will explore two test cases classified as libertarian paternalistic policies and determine whether they promote beneficence while at the same time preserving autonomy. I will argue that a legitimate nudge needs to pass evaluation for an additional condition beyond what is included in Thaler and Sunstein's criteria. In many nudge proposals, beneficence considerations justifiably outweigh small, minor infringements of autonomy. The condition needed to rule out some more problematic test cases is that autonomy trumps beneficence to reject policies if *crucial* autonomy considerations are violated, where crucial is understood as not excessively curtailing self-direction, upholding negative obligations to respect individuals, and including interventions that individuals would choose if and only if they were fully rational and well-informed.

Types of error in implementation

Unfortunately, public policy is not perfect and there may be cause for concern that the government may make mistakes when deliberating about possible proposals. As for libertarian paternalism proposals, a good policy would be one that both helps people and at the same time is not too paternalistic- it achieves the right balance between beneficence and autonomy. A bad policy would be one that neglects to help people or that is too paternalistic- it does not attain the right balance of beneficence and autonomy. In an ideal world, we want to accept only the good policies and reject only the bad ones. Realistically, though, it's helpful to explore what happens when we are faced with error in policy implementation. The following table shows the possible types of outcomes of policy implementation.

	<i>REJECT POLICY</i>	<i>ACCEPT POLICY</i>
<i>Helps people and not too paternalistic</i>	BAD (FALSE POSITIVE ERROR)	GOOD
<i>Helps people but too paternalistic</i>	GOOD	BAD (FALSE NEGATIVE ERROR)

Implementation of libertarian paternalism can result in two different types of bad, or erroneous, outcomes. The first type of error is classified as a false positive in which we reject implementing policies even though they would have turned out to be good. The second type of error is classified as a false negative in which we implement policies even though they turn out to be bad. Surely, while both types of errors are to be avoided when creating and deliberating about specific proposals, it is a task that is easier said than done. The supposed goals of public policy are multi-faceted and can include various competing values such as promoting welfare, upholding autonomy, maintaining justice, maximizing

capabilities, etc. and it is challenging to uphold one at a time much less many of them simultaneously. In clarifying this issue, it helps to narrow down the two main goals for libertarian paternalist policies in question: promoting beneficence and upholding autonomy.

Error in implementation of policies can be likened to the discussion of trading off beneficence and autonomy. If the government adopts a view that autonomy should trump beneficence in all cases where they conflict, then it is analogous to cases where we reject policies that would have turned out to be good. The policy would have been helpful and it wouldn't have been too paternalistic if it hadn't been rejected. This autonomy-trumping stance is just the same as committing a false positive error. If the government adopts a view that beneficence should trump autonomy in all cases where they conflict, then it is analogous to cases where we accept policies that turn out to be bad. The policy would have been helpful but only at the expense of being too paternalistic. This beneficence-trumping stance is just the same as committing a false negative error. At the public policy level, the optimal solution would be to balance between the autonomy-trumping stance and the beneficence-trumping stance depending on the situation rather than only adhering to one or the other. To reduce the possibility for erroneous outcomes in implementation, it's helpful to more closely analyze the source of error. The difference between good and bad proposals is not whether we should help people— all libertarian paternalist proposals contain a beneficent element. The difference between good and bad proposals is in the distinction of whether it is “too paternalist.” This means the important task is to figure out which proposals begin to create unjustifiable

infringements on autonomy. On my view, there are limits on the extent to which infringements of autonomy should be allowed. Later in this thesis, I will return to the topic of balancing beneficence and autonomy after the problems of libertarian paternalism will be discussed. Before that, in the next section, the rationality findings related to libertarian paternalism will be presented. This will provide the underpinnings for why libertarian paternalist measures should be considered in the first place and why promoting beneficence is important.

Rationality Findings Related to Libertarian Paternalism

The primary motivation for supporters of libertarian paternalism comes from recent findings in behavioral economics on rationality. The findings in behavioral economics indicate that “people do not appear to do what is best for themselves” (Loewenstein & Ubel, 2008). Basically, individuals often fail to maximize their own long-term welfare. Although it was historically assumed that agents always act in accordance with rational self-interest, there has been an increasing tendency to assume that many psychological flaws in our human nature hinder us and prevent us from being rational. In particular, economic theory has traditionally assumed that agents have well-defined stable preferences, that they are adept at judging probabilities and can faultlessly integrate their preferences and beliefs by maximizing their utility. This traditional view has been replaced by the work of behavioral economists that has taken place over the past few decades.

A foundational perspective which can be traced back to early behavioral economics is called bounded rationality (Simon, 1955). Bounded rationality argues that

people face limitations on making optimal, rational choices and, as a result, often make poor choices. It is the starting point for many experimental examples exhibiting flaws in decision making that take place at the intersection of psychology and economics. With these findings, Thaler and Sunstein demonstrate situations where individuals fail to act rationally and by doing this they motivate the need for their policy proposals. They are not the first considering there is a move in economics to become “increasingly interventionist, stimulated in part by advances in behavioral economics” (Loewenstein & Ubel, 2008; Camerer, et al., 2003; Layard 2005; Frank, 1985; Goodin, 1991). It is important to note, however, that behavioral economics does not abandon the neoclassical approach to economics “based on utility maximization, equilibrium, and efficiency” (Camerer & Loewenstein, 2002). Rather, the appeal of behavioral economics is to improve economics by “generating theoretical insights, making better predictions of field phenomena, and suggesting better policy” (Camerer & Loewenstein, 2002). So, the same basic economic framework is being advanced by taking limits on rationality into account.

The prevalence of bounded rationality has been supported by a great deal research in behavioral economics. Some of the earlier experimental examples in the field generated “prospect theory” which was proposed as an alternative to expected utility theory (Tversky & Kahneman, 1974; Kahneman & Tversky 1979). Expected utility theory assumes that individuals always make accurate choices when confronted with uncertain options concerning gains and losses simply by calculating probabilities and combining them with utility values based on stable context-independent preferences. Under expected utility theory, it is assumed that individuals will choose the option that

results in the highest utility, which is thought to be the option that is best for them. In opposition to this, prospect theory exhibits the fact that individuals, in reality, are not very accurate at the kinds of choices involved with uncertainty. Instead, individuals are highly sensitive to gains and losses when they are presented from different, even arbitrary, reference points. Under prospect theory, individuals do not choose the best option because they have been unduly influenced by other things besides mere utility and probability such as situational factors. An overwhelming amount of experimentation in various areas suggests that erroneous decision-making is the result of ill-conceived judgments, ill-defined choices, as well as other considerations involving risk, such as reference dependence. The research in each of these areas will be elaborated in turn.

Ill-conceived judgments

In behavioral economics, it is not assumed that individuals always accurately adjust judgment in light of changes in the likeliness of an outcome. In fact, there are many examples to the contrary depicting instances of poor probability judgment. The first examples fall under the classification of different “heuristic mechanisms” which lead people to make judgments that do not conform to true probabilities (Kahneman & Frederick, 2002). The most well known is the “availability heuristic” which explains that people make predictions based on how easy it is to recall past events or imagine future events (Kahneman & Frederick, 2002). Heuristics contribute to other biases such as the “hindsight bias” which is overestimating probabilities that were attached to events in the past and the “curse of knowledge” which is the tendency to “find it hard to imagine how little others know” (Tversky & Kahneman, 1974; Camerer & Loewenstein, 2002). In

addition, heuristics play a role in employing mental shortcuts as in “representativeness” wherein we overestimate occurrences of stereotypes within a particular class type (Kahneman, 2002; Camerer & Loewenstein, 2002). In other words, we tend to overestimate occurrences of patterns. Ill-conceived judgments also come in the form of “confirmation bias” which is the tendency to correctly attribute evidence in support of a belief while at the same time mistakenly attributing counter-evidence to that same belief (Rabin & Schrag, 1999). Other instances of self-serving motivations are evident, such as “wishful thinking,” in which we subjectively assign a higher likelihood to outcomes that are favorable to ourselves (Babcock & Loewenstein, 1997).

Ill-defined choices

In behavioral economics, it is not assumed that individuals have perfectly defined preferences. There are numerous examples showing that the choices individuals make are the result of ill-defined preferences. One common example is “framing effects” in which actual choices are influenced by the way choice options are presented (Tversky & Kahneman, 1981). Another example is “anchoring” in which we make quantity choices based on adjustment to another quantity that we confidently know (Tversky & Kahneman, 1974). The problem stems from the fact that the anchored quantity frequently results in an inadequate adjustment when the anchor is arbitrarily low or high. Depending on the method used to present choices, individuals can even experience “preference reversals” manifested when the way choices are presented evokes opposite choices when given a comparison between two items (Grether & Plott, 1979). Moreover, when we have to make a choice out of a group of options, our preference is influenced by

the other options included the group; this is called “context effect” (Huber, Payne & Puto, 1982).

Other risk situations

Examples of bounded rationality have been discovered within several other spheres relating to our current situation, past consumption, and tolerance for risk.

“Reference dependence” shows that people have exaggerated preferences for objects that have been given to them or that are relative to other mechanisms that influence choice (Kahneman & Tversky, 1979; Knetsch, 1992; Kahneman, 2002). Reference dependence is exhibited when people are faced with a choice between two or more options in which their differences are drastically drawn out. When presented with the drastically differing options, people will often make choices based on what appears more attractive only in relation to the other option(s). Closely related to this is the “endowment effect” in which people assign a higher value to objects that were recently given to them, in terms of willingness to sell, than people who do not have the same endowment merely because of their ownership (Kahneman et al, 1990). Both reference dependence and the endowment effect are manifestations of “loss aversion” which is essentially the notion that people would rather avoid losses than obtain gains (Kahneman, 2002). Other examples include “mental accounting” which occurs when people assign separate mental accounts to divide money for various purposes even though money does not technically have to be controlled in this way (Thaler, 1980; Thaler, 1999). An “immediacy effect” is demonstrated when individuals who are normally willing to forego larger, future financial rewards for immediate rewards that result in immediate consumption switch their

preferences to the future financial rewards when both rewards are delayed (Prelec & Loewenstein, 1991). This shows that individuals have an irrational penchant for immediate consumption. Next, “magnitude effects” describe situations where people dislike delays in consumption more than they like speeding up consumption (Loewenstein & Prelec, 1992; Loewenstein, 1988). In addition, even something as simple as emotions, depending on the situation, can inadvertently influence decisions by making us focus on short-term rather than long-term benefits (Romer, 2000).

Bounded rationality as a reason to support nudges

Sunstein & Thaler provide many of the aforementioned examples of bounded rationality as reasons for why libertarian paternalist government intervention is justified. In their book, they phrase the employment of clever designs as “choice architecture” that take bounded rationality into account and that nudge us towards decisions that will make us better off. Some of the examples from behavioral economics discussed above that are outlined by Sunstein and Thaler include the availability heuristic, representativeness, framing, anchoring, loss aversion, and mental accounting. They also offer some other examples not yet discussed such as instances where we use faulty rules of thumb or mental shortcuts instead of accurate analysis, when we have overconfidence and fail to plan for likely risks, and when we permit status quo biases in which current inertia or laziness prevent us from taking action. Sunstein and Thaler also put forth that we have many temptations that lead us towards mindless choosing and a basic inability to exercise self-control. Moreover, we have a tendency to just “follow the herd” and act in accordance with what others around us are doing. We worry too much about what people

think about us, especially in potentially embarrassing circumstances which is the “spotlight effect,” and we are susceptible to stimulus that “primes” us towards certain behavior by persuasively gearing us up and dismantling any preventative barriers. Because of these factors, they claim, we are in a position where nudges are desirable for us because they significantly improve our wellbeing in the areas of health, wealth, and happiness where our society is besieged. (Nudge, Chapters 1-3)

The findings from behavioral economics show many instances in which we are not making decisions that are in our own long-term best interest and in which intervention can counterbalance our natural tendencies and improve our ability to make good decisions. However, the prevalence of bounded rationality should not automatically suggest support for unwarranted paternalism; we should still exercise caution and be concerned with providing proper justification for intervention. After all, there may be cases in which mass education programs are better justified than nudges. There is still more work to be done in determining when interventionist methods are appropriate, selecting the degree at which paternalism becomes too coercive, and finding the proper balance for various goals of public policy. Thaler and Sunstein would agree about the need for careful consideration in implementation but would eagerly defend their policies by maintaining that the libertarian element is sufficient to withstand the threat over-intervention. They claim the libertarian element would preserve freedom of choice and allow individuals to pursue and satisfy their own desires. Although, it is debatable whether the criteria they provide is satisfactory for actual policy application. In the next

section, I will introduce some of the general types of objections to libertarian paternalism and differentiate which is of most importance.

General Problems of Libertarian Paternalism

Although the insights from behavioral economics seem warranted and quite persuasive, it is important to investigate the problems of the interventionist position. There are three main criticisms of libertarian paternalism. First, some critics object by claiming that libertarian paternalism is not paternalistic enough. Second, some critics object by claiming that libertarian paternalism is not libertarian enough. Third, some critics argue that libertarian paternalism has weaknesses in its application and the problematic cases that it might allow. This problem also looks at the extent to which unjustified libertarian paternalism needs to be guarded against. Now, we will explore the three criticisms in depth.

Too paternalist

The first possible problem with libertarian paternalism is that it is too paternalist. This manifests itself in several different forms but the most general position is held by extreme libertarians. Basically, the position maintains that interventionism at any cost is too costly and any coerciveness is and too coercive. Stated another way, individuals ought to be left alone and allowed to make decisions that make them worse off and that the government should not encourage or intervene one way or another. Moreover, as Thaler and Sunstein point out, libertarian paternalism can slide down a slippery slope to harder forms of paternalism. The slippery slope argument implies that the government will implement policies that are increasingly more and more paternalist as time goes on

even if we begin with modest policies. The concern is that there will be no stopping the government from overstepping and sliding all the way down the slope to aggressive policies. (Nudge, pg 239-241; Glaeser, 2006) They acknowledge and respond to this objection by making three refutations. First, they say that the slippery slope argument “ducks the question of whether ...proposals have merit in and of themselves” (Nudge, pg 240). Second, they say that their construction of the definition of the libertarian paternalism has a built in response to the slippery slope argument. Namely, the libertarian element of it is enough to guard against the possibility that policies will become too paternalistic. Third, they say that sometimes it is necessary for the government to make a choice one way or another as part of choice architecture for new programs and, all things considered, nudges that steer people towards making the decisions that will make them better off is preferable to the alternative of doing the opposite or nothing at all. However, Thaler and Sunstein themselves admit hesitation in that “deciding where to stop, and when to call a nudge a shove..., is tricky” (Nudge, pg 254).

Another possible formulation of this general objection is made by Edward Glaeser. Glaeser takes the slippery slope argument further and makes an interesting case against libertarian paternalism by saying that it is more dangerous than hard paternalism (Glaeser, 2006). Glaeser classifies the aforementioned behavioral economics research as situationalism in which erroneous decision-making characterized by our limited ability to reason, our reliance on short-term wellbeing, and context dependence ultimately results in long-term wellbeing being compromised (Glaeser, 2003). He calls these instances of

situationalism because decisions are made based on isolated or localized situations without regard for the long-term. He brings up many other real-world examples that exhibit the fact that individuals are extremely subject to social influence and tend to make erroneous decisions. (Glaeser, 2006) For instance, some specific examples described in detail by Glaeser include false advertising, belief manipulation in the form of indoctrination in schools in certain countries, and media techniques used solely for belief formation. He uses these examples to show that while traditional welfare economics is in need of reform this should not necessarily be taken as an argument for more government intervention. Glaeser builds his case on grounds that are counter to what we might intuitively think about soft paternalism which is that the government is potentially the solution to the problem of bounded rationality at the individual level. He claims that both individuals and governments are bad at making decisions and exhibit bounded rationality but that governments are even worse at decision-making than individuals. Furthermore, once governments start implementing soft paternalism, they have a strong incentive to conceal and then abuse their power.

I reject the extreme libertarian objection because many policies hold great promise to increase societal wellbeing. The benefits should not be dismissed merely from the threat that the government will poorly execute policies. In actuality, there are many successful examples of soft paternalist policy implementation that are widely accepted as both desirable and prudent. I also accept the characterization against strict anti-paternalism that it ignores distributive dimensions of fairness (Arneson, 1989). Restricting all paternalism “in effect gives to the haves and takes from the have-nots”

because “more able agents are more likely to do better for themselves choosing among an unrestricted range of options whereas less able agents are more likely to opt for a bad option that would have been removed from the choice set” (Arneson, 1989). In terms of one of the various goals of government, some acts of paternalism are warranted in principled support of fairness for the less advantaged.

Too libertarian

The second purported problem with libertarian paternalism is that it is too libertarian. This main position is held by extreme paternalists. It relies on the assumption that individuals experience crippling bounded rationality and are unable to make optimal choices without assistance. Essentially, the goal is to protect people from making bad choices at any cost or with any coercive measure because freedom of choice is not a priority. While wholesale endorsement of strict paternalism is not common, there are some views that can be construed as having much more paternalist ambitions than libertarian paternalism. Some perfectionist political views that rely heavily on objective accounts of wellbeing potentially warrant higher levels of paternalistic interference (Arneson, 2000). There are also lines of reasoning which would give support for taxes, bans, and laws on behaviors deemed undesirable for society. Thaler and Sunstein briefly address this general objection to libertarian paternalism and dismiss it because taxes, bans, and laws imply high costs to society (Nudge, 252). They only support policies that fall within the threshold of “low costs.” This is also the threshold offered by asymmetric paternalists in which the actual costs and benefits of a policy are weighed. For the asymmetric paternalists, the benefits to the least sophisticated must be large and the costs

to the most sophisticated must be small (Camerer, et al., 2003). The least sophisticated group includes those who frequently make errors and the most sophisticated group includes those who are fully rational (Camerer, et al., 2003) For example, allowable impositions would include mandatory waiting periods before a couple may get divorced or “cooling off-periods” after making a purchase from a door-to-door salesman. Although, going further, neither asymmetric paternalists nor libertarian paternalists would give support to, for instance, bans even in the areas of safety, health, or consumer protection. In the paternalist line of reasoning, though, there are some who do give justification for these types of bans (Frank, 1985). Moreover, some strict paternalists would give justification for even more measures that restrict freedom of choice in cases where, it has been argued, beneficence trumps autonomy (Brock, 1988). For example, one stance that has been put forth is the imposition of a “corrective tax” on income to the point where people are deterred from seeking a higher relative income level; this tax is suggested in light of research showing that increases in relative income past a certain point do not contribute to happiness (Layard, 2005).

I reject the extreme paternalist objection to libertarian paternalism because government should be cautious about implementing policies that restrict autonomy. At the least, preservation of autonomy stands as a barrier to unchecked paternalism (Feinberg, 1986) or oppression. A basic conception of autonomy in which people are minimally responsible for their decisions and are able to speak for themselves (Christman, 2009) should be upheld as one of the varied goals of government. Furthermore, a basic conception of freedom of choice in which people have consumption

options should be upheld as one of the varied goals of an economy. As previously discussed, a widely held position in the paternalism literature is that autonomy should always trump beneficence when the two notions are in conflict. I am not defending that strong of a view; I am only defending the intuition behind it that sometimes autonomy must trump beneficence in order to achieve a proper balance between the two notions. In sum, strict paternalism should be dismissed to avoid the infringement of crucial autonomy conditions.

Criteria are not adequate

The third and a more compelling problem with libertarian paternalism is that Thaler and Sunstein attempt but fail to make an adequately clear criterion by which to judge exactly which proposals are desirable yet not too paternalistic. They offer the basic theoretical underpinnings for policy proposals but there is still much debate about what might and might not be allowed and considered as a nudge. In general, we want to respect people's freedom of choice unless we have reasons to believe otherwise that intervention is preferable but that does not tell us exactly where to draw the line. Drawing the precise distinction between which policies are justifiable and which are not justifiable would be helpful for determining the restrictions on the theory and also helpful, obviously, for the purpose of public policy implementation. There are other practical concerns that go along with this problem. First, government planners could do a bad job in the implementation acting as "evil nudgers" as Thaler and Sunstein suggest. They respond by saying that there should be monitoring of nudge-makers, that planners should be well-informed, and that people should still have freedom to choose alternatives.

Second, some critics contend that government should remain neutral on issues without swaying opinions one way or another. Thaler and Sunstein respond by saying that it is impossible for government to be neutral on some decisions. Even still, the problem of finding the exact limits of libertarian paternalism remains.

Overall, Thaler and Sunstein's criteria are inadequate. Portions of their book are devoted to how nudges can be implemented and to describing examples of nudges but not enough addresses the detailed conditions that make a nudge justifiable. Beyond meeting the requirements of being both libertarian and paternalist as they define those terms, Thaler and Sunstein give one "golden rule" for implementation of nudges which is the following: "offer nudges that are most likely to help and least likely to inflict harm" (Nudge, pg 74). They expand slightly by saying "people will need nudges for decisions that are difficult or rare, for which they do not get prompt feedback, and when they can easily understand" (Nudge, pg 74). So, although Thaler and Sunstein give a very intuitive, commonsensical rationale for deciding when to implement nudges, I disagree that it is specific enough to rule out problematic cases and error in implementation at the public policy level where decision-making must be methodical, well-defined, and designed to reduce error.

In sum, rather than focusing on the two extreme objections to libertarian paternalism, the more interesting and practical debate resides in the middle determining where limitations on autonomy are justified. While careful examination of policies should take place at the policy level anyway, I contend that it depends whether we prioritize beneficence or autonomy. Libertarian paternalism is legitimate in some cases

when the proper balance between autonomy and beneficence is achieved. The most practical and fruitful exercise is to explore whether we can distinguish what precisely makes libertarian paternalist proposals justifiable. There are crucial autonomy considerations that should trump beneficence in some certain cases. Numerous problem cases can be imagined but the next two sections will present two challenging cases that reside in the middle-ground of whether or not they fall within the criteria of justified libertarian paternalist policies.

The Problem of Subliminal Messages

One problematic case for libertarian paternalism that will be discussed is subliminal messages. It is problematic because it serves as a test for determining whether the criterion for judging libertarian paternalist policies is adequate. In this context, subliminal messages are a hypothetical example of an extremely discreet mechanism the government could use to influence people to make decisions that will improve their wellbeing. The problem stems from the fact that, as defined, libertarian paternalism allows for the use of subliminal messages. This is because libertarian paternalism allows for interferences in which people may not be aware of the mechanisms used to influence their decision-making, which is an essential feature of subliminal messages. For the purposes of this thesis, a subliminal message will be communication that is transmitted via means that are not discernable to the recipient and that influence behavior accordingly. Some simple examples of subliminal messages of the kind in question that challenge what should be considered libertarian paternalism include strategically placed

images on TV of happy people eating their vegetables or frequency wave announcements reminding people to brush their teeth (Bovens, 2008; Hausman & Welch, 2009).

Neither subliminal messages nor libertarian paternalism “rule out alternatives or make them more costly” (Hausman & Welch, 2009). Subliminal messages don’t rule out alternatives because they are among deceptions that only “make it appear to the individual that the choice set has changed” (Hausman & Welch, 2009). At worst, they only make some alternatives much more difficult to choose but not impossible to choose. To imagine, every time we see the strategically placed subliminal commercial depicting a family joyfully devouring their steamed asparagus, we may have the urge to do the same although we still may not make that choice for various reasons. We may have an iron-clad will of resistance, we may recognize and ponder the subliminal nature of the commercial and choose to overcome its message, or we may be too lazy to make the choice even though it feels pressing or urgent, and so on. Additionally, subliminal messages don’t make alternatives more costly because they are inexpensive yet effective. Even an amateur on a shoestring budget could film a 30 second commercial in an ordinary setting depicting a family with a voracious appetite for asparagus; plus, we could imagine that the television stations would “donate” allotted time to the government. Even still, on the face of it, subliminal message nudges could be objectionable because they “interfere with... preference formation” (Bovens, 2008) when the public is not aware of it. More importantly, subliminal message nudges are objectionable because we cannot opt-out of receiving the messages in instances when the government leads us towards a choice we wouldn’t have otherwise made (i.e. heartily eating asparagus).

Thaler and Sunstein attempt to confront the problem of subliminal messages by appealing to the publicity condition put forth by John Rawls which states that principles be known and understood by the public (Rawls, 1971). Thaler and Sunstein interpret this as meaning any policies must be publicly defensible by the government (Nudge, pg 247). They suggest that the publicity condition be imposed to rule out the possibility of subliminal messages because the government would be unwilling to publicly defend their use of them. But, their response is inadequate because “the government might be able and willing to defend the use of subliminal messages... to satisfy the publicity condition” (Hausman & Welch, 2009). One possible solution that has been suggested to protect against potential infringements of freedom is to make a requirement that the government *actually* informs the public of its messages, not whether it is merely able and willing to inform the public (Hausman & Welch, 2009). This could be done in a printed register of some sort where all subliminal messages are reported and disclosed. This raises another issue about subliminal messages wherein their effectiveness depends on secrecy and allowing the government to secretly broadcast subliminal messages leading to a slippery slope. In order to allay these concerns, it has been suggested that the public should be able to reasonably uncover every instance of a nudge even though that may make them less effective (Bovens, 2008).

Thaler and Sunstein admit that subliminal messages are objectionable on the grounds they are impossible for individuals to monitor (Nudge, 249). Even still, they cannot rule out subliminal messages because they do not represent a violation of their criteria for infringement on freedom which is only that policies must not rule out

alternatives or make them more costly (Hausman & Welch, 2009). This qualifies subliminal messages as nudges because there is nothing to rule them out on Thaler and Sunstein's view. Yet, the point stands that subliminal messages infringe on freedom in other ways by curtailing self-direction and by not showing respect for individuals. First, subliminal messages curtail self-direction because it is not clear how one could easily decline to accept subliminal messages in the same way that it is easy to decline enrolling a program such as Save More Tomorrow. Even if we were informed about the subliminal messages, we do not possess the means to turn off the continuous bombardment if we want to. Second, subliminal messages also do not show respect for individuals because people are inadequately informed about the reasons they are making their choices. This violates negative constraints on obligations to individuals to not "interfere with their efforts to make their own decisions and govern their own conduct" (Dillon, 2010). The more outright but similar example of a person being brainwashed is the classic case of a person who exhibits non-autonomous behavior. As is the case with brainwashing, a person's "rational faculties have played little... role in the formation of the preference" (Scoccia, 1990) with subliminal messages. Thaler and Sunstein's criteria for a nudge are inadequate because it allows subliminal messages and it is my view that they should be avoided because they represent crucial infringements on autonomy.

The Problem of the Cocaine Vaccine

Another problematic case for libertarian paternalism that will be discussed is the cocaine vaccine. Similar to subliminal messages, it serves as a test for whether the criterion for judging libertarian paternalist policies is adequate. Interestingly, there has

been recent pharmaceutical research developing and investigating the plausibility of such a drug (Fox, 1997; Cohen, 1997; Johnson et al, 2000; Kantak, 2003). So, it is probable that there may realistically be future policy discussion on this topic. Now, in this context, the cocaine vaccine will represent a hypothetical proposal that could be implemented by the government to help people make a choice that will improve their wellbeing. The cocaine vaccine would be “administered by injection... and remain in the bloodstream for an extended time” during which an anti-cocaine agent “renders cocaine both harmless and useless” in the body (Stevenson, 2005).

The cocaine vaccine could have several different applications. For instance, it could be administered to parolees, to welfare recipients, as a universal vaccination, or as a condition for certain types of employment (Stevenson, 2005). On the face it, requiring a cocaine vaccine for any purpose seems like a blatant violation of the libertarian commitment to not restrict freedom of choice. Upon closer inspection, we can easily imagine versions of the aforementioned cases that allow for freedom of choice to be maintained. First, the cocaine vaccine could be administered to parolees with drug charges as an option in exchange for reduced penalties or to welfare recipients as an option in exchange for increased benefits. Second, the cocaine vaccine could be administered to individuals who choose to elect themselves as recipients for preventive reasons. This might apply to individuals who know they have a tendency or past history to use cocaine but who wish to permanently overcome the temptation. Third, the cocaine vaccine could be administered universally by default with other childhood vaccinations unless an option to opt-out is exercised. Fourth, the cocaine vaccine could be required

for certain jobs where employees could potentially harm others as in airplane pilots, bus drivers, heavy machine operators, etc. In this case, individuals still have the option to pursue another job if they are opposed to the vaccination.

Thaler and Sunstein do not address the hypothetical case of the cocaine vaccine but it represents an interesting and useful application example for libertarian paternalism. The situation where individuals elect themselves for the cocaine vaccine, as in the second example outlined above, is similar to the real-life example of self-bans on gambling which Thaler and Sunstein do endorse. In the gambling self-ban, gambling addicts can choose to put themselves on a list that forbids them from entering a casino (Nudge, 235). They claim “self-bans are a great idea and suggest that research be done to explore ways to use this concept in other domains” (Nudge, 235). The central question is whether the cocaine vaccine violates the nudge criteria for infringement on freedom which is that policies must not rule out alternatives or make them more costly. Electing to choose a cocaine vaccine does not appear to rule out alternatives if one freely chooses to have the vaccination and apparently does not impose any costs (assuming that the development and implementation costs would be as non-prohibitive as other vaccines). But, even when people freely choose to rule out alternatives, they are in fact reducing their choice set. Strictly by their own definition, Thaler and Sunstein’s criteria do in fact disallow both the gambling self-ban and the cocaine vaccine if they are not to allow policies that “rule out alternatives.”

A similar example in the literature on paternalism is of Ulysses and the Sirens in which Ulysses ties himself to the mast to prevent himself from being lured by the songs

of the Sirens. In the case of Ulysses, gambling self-bans, and self-elected cocaine vaccines, we should want to say that the free choices to rule out potentially harmful alternatives are justified yet Sunstein and Thaler's criterion is inadequate in these scenarios because it is too strong. There may be even more reasons to believe that having more choices at our disposal may not necessarily lead to the maximization of wellbeing and may not be as all-important as Thaler and Sunstein assume. In fact, there is evidence pointing towards the view that we have a better chance of increasing wellbeing "from some limited scheme of constraints or burdens on choice to help us avoid common mistakes" (Haybron, 2007) than from having unlimited choices.

The situation where universal vaccination is implemented on a national level, as in the third example outlined above, also presents a challenge to Thaler and Sunstein's criteria. Again, the central question is whether the cocaine vaccine violates the nudge criteria for infringement on freedom which is that policies must not rule out alternatives or make them more costly. It could be argued that while the universal cocaine vaccine rules out one alternative (i.e. using cocaine) it actually increases other choices available to those who would otherwise have been impeded by drug use. New choices available to those newly reformed as drug-immune include more options for spending discretionary money, more ways to spend recreational time, the ability to hold down a job, staying out of jail, among many others. And, the universal cocaine vaccine does not make choices more costly because we could assume it would not impose any prohibitive costs beyond what is already incurred via the same administering process with other standard childhood vaccinations. Furthermore, the universal vaccine, like other childhood

vaccinations, would be highly encouraged yet not mandated; you maintain the option to opt-out.

So, does the universal vaccine pass Thaler and Sunstein's stipulated criteria for a nudge or is it too intrusive? Are crucial autonomy considerations being violated? On the face of it, many people might consider the idea of a universal cocaine vaccine to be an appalling violation of self-direction and the ability to live one's life according to the way one chooses despite the increased choices available to those who live drug-free. But, the objective itself does seem to hold the utmost respect for individuals in providing the opportunity to live addiction-free. On Thaler and Sunstein's view, the cocaine vaccine would not be allowed technically because, if chosen, it rules out an alternative, albeit non-mandated. But, let's reconsider whether the cocaine vaccine is actually a violation of autonomy because an important criterion has yet to be mentioned. In this case, one way to determine whether the cocaine vaccine is objectionable is to use an autonomy-based criterion and ask "if the people in question would consent to the interference if they were fully rational" (Scoccia, 1990) especially considering the fact that the vaccine would be administered during childhood and not to fully competent adults. Even if people irrationally believe that the vaccine represented an infringement on their self-direction, implementing the cocaine vaccine might be justified because they would consent if they were "fully rational and well informed" (Scoccia, 1990) about it. Surely, careful and rational contemplation concludes that living a life drug-free is more conducive to enhancing wellbeing than not. Granted that there may be some fully informed individuals who still choose to use cocaine occasionally or recreationally, the

main force of this criterion is to show that the vast majority of fully informed individuals would not choose to use cocaine given the high risks of addiction among other things. This is enough to show that Thaler and Sunstein's criteria for a nudge are inadequate because their reason for rejecting the universal cocaine vaccine (i.e. rules out alternatives) is not the right one. It is my view that Thaler and Sunstein's criteria should instead be supplemented by reasons that represent autonomy considerations because the reasons they provide for the instances of self-election and the universal cocaine vaccine are either conflicting or mistaken in their aims.

Additional Condition for Evaluating Justifiable Nudges

As stated above, the central debate surrounding libertarian paternalism hinges on concerns about when limitations on autonomy are justified for the sake of beneficence. In other words, we need to find the right balance between libertarian elements, which guard against coercion, and paternalism elements, which promote welfare. As Thaler and Sunstein's view on libertarian paternalism policies stands, it is inadequate because their criteria do not appropriately deal with the test cases of subliminal messages and cocaine vaccines. My view is that there is a middle option between the opposing elements, as Thaler and Sunstein also desire, but that it includes something else beyond what they stipulate to handle the aforementioned test cases. It is based on a balancing view which puts forth the commonsense notion that there need not be an overarching commitment to autonomy or beneficence in all cases whenever there is conflict. These values must be weighed on a case-by-case basis.

My view is that a legitimate nudge needs to pass evaluation of an additional condition beyond what is included in Thaler and Sunstein's proposal. Evaluating nudges involves weighing the opposing considerations of beneficence and autonomy and, depending on the situation, in some cases beneficence trumps autonomy while in other cases autonomy trumps beneficence. Many standard nudges, such as the cafeteria director scenario or the presumed consent organ donation scenario, are examples of when beneficence trumps autonomy and justifiably so. But, even though they meet the criteria put forth by Thaler and Sunstein, the subliminal messages nudges are unjustifiably allowed and the cocaine vaccine nudge are unjustifiably disallowed. This shows that their criteria are inadequate and are in need of an additional condition. The condition is that autonomy trumps beneficence to reject policies if three crucial autonomy considerations, described below, are violated. If crucial autonomy considerations are violated, this should serve as reason to deem a policy unjustified and unadoptable. The purpose of giving autonomy considerations special status is that caution should be exercised when evaluating autonomy considerations in order to avoid acceptance of a policy that turns out to be undesirable. Before describing the autonomy considerations, I will briefly review beneficence considerations.

Beneficence considerations

In most nudge cases, beneficence trumps autonomy in favor of acceptance of a proposal if beneficence criteria are overwhelming present along with only minor restrictions of autonomy. In the standard cases discussed in Thaler & Sunstein's book, some of the resulting specific beneficence considerations are that long-term happiness is

maximized, benefits accrue to individuals that are least sophisticated, and so on. If these improvements in wellbeing are the outcome of implementing nudge policies at the expense of only a minor violation in autonomy, such as a slight curtailment of self-direction (as in the cases of the cafeteria director example or presumed consent organ donation), then they are justified. To consider another example, body scanning machines at airports or other federal buildings are justified because they provide significant welfare benefits in the sense that they achieve more benefits than harm. They do so at the expense of only a slight violation of autonomy in terms of respect for privacy. However, it would be a mistake to think that there are not cases in which autonomy trumps beneficence.

Autonomy considerations

Autonomy is a notion that has many different meanings and applications. An overarching definition is that being autonomous is “to be directed by considerations, desires, conditions, and characteristics that are not simply imposed externally upon one, but are part of what can somehow be considered one’s authentic self” (Christman, 2009). Beyond this definition, though, there is much more complexity and little agreement among philosophers about choosing the specific conditions of autonomy. While autonomy is not a simple concept, there is a common assumption which is that autonomy captures a person’s ability “to act, reflect, and choose on the basis of factors that are somehow [his or] her own” (Christman, 2009). The three autonomy criteria I selected as necessary for nudge evaluation are chosen with that common standard in mind. While

my criteria draw upon different theories of autonomy, they are all needed to account for the great diversity of “nudgeable” applications.

Autonomy may trump beneficence in favor of rejection of a policy if crucial autonomy considerations are violated. The key issue is to delineate exactly which considerations to include in the criteria. Thaler and Sunstein define the libertarian element, or the autonomy considerations, as consisting of the following. First, nudges should not rule out alternatives. Second, nudges should not be costly. In order to rule out potentially coercive policies and allow the beneficial ones, I contend that at least the following three crucial autonomy considerations should be added. First, nudges should not curtail self-direction in a major way. As in the case of subliminal messages, self-direction is curtailed in a major way when the ability to refuse intervention is unavailable. Self-direction, or self-determination, is one main component of an autonomous life and should be upheld to a certain extent. Some limitation in self-direction is justified, as in the cafeteria director example, where certain food options are presented to you in a way that makes it more burdensome to choose what you really want- triple fudge brownies for lunch. However, it is not extremely difficult to overcome your desire if all that is required is looking past the steamed vegetables or eating at a different location. Self-direction, in the cafeteria director example, is still preserved and the intervention is justified. In the subliminal message cases, as well as others that can be imagined, the major infringement on autonomy is that we are severely captive to opposition to our self-direction.

Second, nudges should show respect for individuals, understood as protecting negative constraints owed individuals to not severely “interfere with efforts to make decisions and govern their own conduct” (Dillon, 2010). This is one understanding of respect out of a few of the contending standard interpretations of respect in the current literature. As in the case of subliminal messages, individuals are not adequately informed about the reasons they are making their choices unlike when an opt-out program for organ donation is adopted, for instance. With opt-out organ donation, individuals can inform themselves about the benefits of complying and still make a choice to decline. Their rational faculties can play a role in their decision-making. In contrast, with subliminal messages people do not actively and expressly have the option to consider alternatives and employ rational faculties to independently form a preference.

Third, nudges are justified if and only if the person would consent if they were fully rational and well-informed (Railton, 1986). This general conception of autonomy is motivated because people often make choices that are counterproductive to wellbeing if left to their own devices with insufficient information at their disposal. Given this tendency, people could benefit from interventions that lead them to the choices they would make if they were more rational and better informed about what will actually increase their wellbeing. This preserves the element for individuals to be able to self-select their desires with a guarantee that interventions would not go beyond what they themselves would do with full information if they had it. As we know from the findings in behavioral economics, people do face limitations on making rational choices and it makes sense to consider what decisions and nudges people would be open to if they were

fully rational. In both the self-elected and universal cocaine vaccine, people who are fully rational and well-informed would choose to allow the self-elected cocaine vaccine because of the beneficence considerations under full information. Conversely, individuals who irrationally believe that the Save More Tomorrow program is worthless would surely consent to enrollment in it if they were fully rational and well-informed and thus offering it is justified.

I have provided three specific autonomy considerations by which to judge nudge policies in addition to the elements provided by Thaler and Sunstein. Adding these additional autonomy considerations to Thaler and Sunstein's standard criteria for a nudge will overcome the problematic cases of the subliminal message and the cocaine vaccine. Subliminal messages infringe on autonomy by curtailing self-direction and by not showing enough respect for people. Despite contrary intuitions, the cocaine vaccine does not infringe on crucial autonomy considerations because it does not curtail self-direction, disrespect people, or violate what people would do if they were fully rational and well-informed. These considerations provide a reasonable list with which to check each proposal on a case-by-base basis. My additional autonomy condition signals how to properly balance autonomy and beneficence and when to reject nudge policies. It is preferable to the less specific criteria put forth by Thaler and Sunstein because it is thorough enough to appropriately deal with problematic cases.

Objections

In this section, I will confront the main three objections to my view along with responses to each. The first objection to my view charges that it is too complex. Thaler

and Sunstein's theory have fewer and simpler criteria to apply. No doubt, such an original and ingenious theory as theirs is surely to be commended for being so straightforward and uncomplicated; however, it falls short in its ability to deal with complex cases. I respond to this objection by pointing out that Thaler and Sunstein's criteria can be both too strong and also not strong enough in terms of protecting autonomy considerations. So, while Thaler and Sunstein's criteria are clever, that they are not quite detailed enough to navigate the variety of real-world scenarios and are in need of additional considerations to become complete. This brings about another related concern about my view which charges that my additional considerations are too ambiguous. Particularly, it can be difficult, for example, to determine exactly what "fully rational and well informed" people might do in specific situations considering that there might not be agreement about the facts or values in question. I respond that a small amount of ambiguity necessarily comes with the territory of adding considerations. And, despite the fact there may not be a convergence of opinions on the facts or values in question, imagining what most people would do under the assumption of full information is still a helpful and useful indicator for whether nudges are justified or not. So, even though it will be hard to judge, adding considerations is something that must be done given the inadequacies of Thaler and Sunstein's criteria that become obvious upon inspection of problematic test cases.

The second objection to my view is the concern about exactly how to balance autonomy and beneficence in each case. A challenge with employing my additional condition for new cases is determining when autonomy should trump beneficence rather

than the other way around. On the balancing view, it can be difficult to find the “appropriate balance between these two potential mistakes” of infringing autonomy versus neglecting welfare promotion (Brock, 1988). There is certainly cause for concern about “erring too far in either direction” on these conflicting aims (Brock, 1988).

Admittedly, there is no easy rule to determine the appropriate place to delineate the conflicting aims on the balancing view; however, my response is that I have provided three distinct considerations that will help identify when autonomy has been violated in a major way. For the purpose of policy implementation, these considerations show us at what point proposals ought to be rejected and become autonomy trumping. Nudges should pass all three criteria and even if one autonomy criterion is violated, then a nudge is under-justified. The ambiguity stems from the fact that all three criteria may not be applicable for every case and that all three criteria may not produce a definite verdict one way or the other. Even still, there should be sufficient autonomy considerations to ultimately pass judgment on the nudge in question.

The third objection comes from the behavioral economics standpoint and argues that my view is too liberal. The evidence from behavioral economics confronts classic liberal conceptions of autonomy because it suggests people don’t function in their own best interest when in freedom-rich environments. Basically, we stand to benefit from helpful interventions more than we previously thought. I accept the findings from behavioral economics as well founded but I respond to this objection by stating that my view offers the right mix of liberalism and beneficence. “The liberal ideal of option freedom involves unbounded and unburdened choice and self-determination” (Haybron,

2007) but that is far from what I am arguing for which is autonomy checks on beneficent policies. So, I am only for a narrow view of self-determination which concerns “basic liberties, rights to noninterference, democratic participation” and the like- not a broad view of it which concerns freedom “in the sense of having a wide range of options” (Haybron, 2007). So, I think my reliance on adding considerations of autonomy are warranted because they still allow helpful beneficence.

Post-evaluation

I propose that nudge policies should be held to standards that are practical enough to be applied in the real world. One final thought that is relevant to the evaluation nudges is the implementation of post-evaluation process. A nudge policy must include review after implementation to determine whether it was successful at achieving its purported aim. As with other considerations, this too must be chosen depending on the circumstances and the policy in deliberation. Thaler and Sunstein do not provide a description of how a nudge might be evaluated after the fact to ensure that implementation is carried out as intended. Policy implementation in reality is often heavily weighed down by bureaucracy and sometimes by concealment. Although they do not address it in their writings, I believe that they would agree with the practical need for a system of checks and balances that serves the function of a “nudge oversight committee.” To reiterate, avoidance and extra protection from implementation of bad policies that infringe on autonomy is the reason for precaution and oversight.

Conclusion

The purpose of this thesis was to evaluate libertarian paternalism. First, the concept of paternalism, the different types of paternalism, as well as the central concepts of autonomy and beneficence were explained in detail. Second, the rationality findings from behavioral economics related to libertarian paternalism were summarized. Third, the general objections to libertarian paternalism were discussed. Fourth, the specific case problems of subliminal messages and the cocaine vaccine were described in detail. Next, I presented the conditions which must be evaluated in order for a nudge to be justified which overcome the specific problems facing libertarian paternalism.

The debate surrounding nudges can be summarized in terms of respect. Acts of beneficence are a display of respect for others while autonomy is preservation of self-respect. In order to justify nudges, a proper balance between respect for others and self-respect must be negotiated. To use an analogy, nudges consist of domino-effects where the first domino to fall is preservation of self-respect and the subsequent domino to fall is respect for others. Showing respect for others by improving the wellbeing of people with bounded rationality should not come at the expense of self-respect but rather after it or with it. We have negative obligations to not interfere with other's efforts "to make their own decisions and govern their own conduct" (Dillon, 2010) and then positive obligations "to protect them from their own self-harming decisions" (Dillon, 2010). The criteria that Thaler and Sunstein provide as the justification for nudges are inadequate. In this thesis, I have attempted to add an additional condition based on autonomy that effectively deals with problematic test cases. Ideally, nudges should accomplish both goals of preserving self-respect while maintaining respect for others.

WORKS CITED

- Arneson, Richard. (1980). "Mill vs. Paternalism." *Ethics*. Vol. 90. pg 470-480.
- Arneson, Richard. (1989). "Paternalism, Utility, Fairness." *Revue Internationale de Philosophie*. Vol. 170. pg 409-423.
- Arneson, Richard. (2000). "Perfectionism and Politics." *Ethics*. Vol. 111, No. 1.
- Babcock, L. and Loewenstein, G. (1997). "Explaining Bargaining Impasse: the role of self-serving biases." *Journal of Economic Perspectives*. Vol. 11. pg 109-126.
- Bovens, Luc. (2008). "The Ethics of Nudge." *Preference Change: Approaches from Philosophy, Economics and Psychology*. Springer.
- Brock, Dan. (1983). "Paternalism and Promoting the Good." *Paternalism*. University of Minnesota Press.
- Brock, Dan. (1988). "Paternalism and Autonomy." *Ethics*. Vol. 98, No. 3. pg 550-565.
- Camerer, Colin and Loewenstein, George. (2002). *Behavioral Economics: Past, Present, Future*. Draft.
<http://econ2.econ.iastate.edu/tesfatsi/camererintrochapter.advancesbe.pdf>
- Camerer, C., Issacharoff, S., Loewenstein, G., O'Donoghue, T. and Rabin, M. (2003) "Regulation for Conservatives: Behavioral Economics and the Case for Asymmetric Paternalism." *University of Pennsylvania Law Review*. Vol. 151, No. 3. pg 1211-1254.
- Christman, John. (2009). "Autonomy in Moral and Political Philosophy." *Stanford Encyclopedia of Philosophy*.
- Cohen, Peter. (1997). "Immunization for Prevention and Treatment of Cocaine Abuse: legal and ethical implications." *Drug and Alcohol Dependence*. Vol. 48. pg 167-74.
- Dillon, Robin. (2010). "Respect." *Stanford Encyclopedia of Philosophy*.
- Dworkin, Gerald. (1971). "Paternalism." *Morality and the Law*. Wadsworth.
- Dworkin, Gerald. (2005). "Paternalism." *Stanford Encyclopedia of Philosophy*.
- Feinberg, Joel. (1986). *Harm to Self*. Oxford University Press.
- Frank, Robert. (1985). *Choosing the Right Pond*. Oxford University Press.

Fox, Barbara. (1997). "Development of a Therapeutic Vaccine for the Treatment of Cocaine Addiction." *Drug and Alcohol Dependence*. Vol 48. pg 153-158.

Glaeser, Edward. (2003). "Psychology and the Market." National Bureau of Economic Research. Working Paper. <http://www.nber.org/papers/w10203>

Glaeser, Edward. (2006). "Paternalism and Psychology." *University of Chicago Law Review*. Vol. 73, No. 1. pg 133-156.

Glover, Jonathan. (1977). *Causing Death and Saving Lives*. Penguin.

Goodin, Robert. (1991). "Permissible Paternalism: In defense of the nanny state." *The Responsive Community*. Vol. 42.

Grether, D. and Plott, C. (1979). "Economic theory of choice and the preference reversal phenomenon." *American Economic Review*. Vol. 69, No. 4. pg 623-638.

Hausman, Daniel and Welch, Brynn. (2009.) "Debate: To Nudge or Not to Nudge." *Journal of Political Philosophy*. Blackwell Publishing.

Haybron, Daniel. (2007). "The Pursuit of Unhappiness." Draft. Social Science Research Network. http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1071822

Huber, J., Payne, J., and Puto, C. (1982). Adding asymmetrically dominated alternatives: violations of regularity and the similarity hypothesis. *Journal of Consumer Research*. Vol. 9, No. 1. pg 90-98.

Johnson, M., et al. (2000). "Active Cocaine Immunization Attenuates the Discriminative Properties of cocaine." *Exp. Clin. Psychopharmacol*. Vol. 8, No. 2. pg 163-167.

Kantac, Kathleen. "Vaccines Against Drugs of Abuse: A Viable treatment Option." *Drugs*. Vol. 63. pg 344-345.

Kahneman, D. and Tversky, A. (1979). "Prospect Theory: an analysis of decision under risk." *Econometrica*. Vol. 47. pg 263-291.

Kahneman, D., Knetsch, J. , and Thaler, R. (1990). "Experimental tests of the endowment effect and the coase theorem." *Journal of Political Economy*. Vol. 98. pg 1325-1348.

Kahneman, Daniel. (2002). "Maps of Bounded Rationality: A perspective on intuitive judgment and choice." Nobel Prize Lecture.

Knetsch, J. (1992). "Preferences and nonreversibility of indifference curves." *Journal of Economic Behavior and Organization*. Vol. 17, No. 1. pg 131-139.

Layard, Richard. (2005) "Happiness and public policy: a challenge to the profession." <http://www.iew.uzh.ch/study/courses/downloads/happy.pdf>

Loewenstein, G. (1988). "Frames of mind in intertemporal choice." *Management Science*. Vol. 34. pg. 200-214.

Loewenstein, G., and Prelec, D. (1992). "Anomalies in intertemporal choice: Evidence and an interpretation." *Quarterly Journal of Economics*. May. pg 573-597.

Loewenstein, George and Ubel, Peter. (2008). "Hedonic adaptation and the role of decision and experience utility in public policy." *Journal of Public Economics*. Vol 92. pg 1795-1810.

Prelec, D. and Loewenstein, G. (1991). "Decision making over time and under uncertainty: a common approach." *Management science*. Vol. 37. pg 770-786.

Rabin, M. and Schrag, J. "First impressions matter: a model of confirmatory bias." *Quarterly Journal of Economics*. Vol. 114, No. 1. pg 37-82.

Railton, Peter, 1986. "Moral Realism," *Philosophical Review*. Vol. 95. pg 163–207.

Rawls, John. (1971). *A Theory of Justice*. Oxford.

Romer, P. (2000). Thinking and Feeling. *American Economic Review*. Vol. 90., No. 2. pg 439-443.

Scoccia, Danny. (1990). "Paternalism and Respect for Autonomy." *Ethics*. Vol. 100, No. 2. pg 318-334.

Scoccia, Danny. (2008). "In Defense of Hard Paternalism." *Law and Philosophy*. Springer.

Simon, Herbert. (1955). "A Behavioral Model of Rational Choice." *Quarterly Journal of Economics*. Vol. 69, No. 1. pg 99-118.

Stevenson, Dru. (2005). "Libertarian Paternalism: The cocaine vaccine as a test case for the Sunstein/Thaler model." *Rutgers University Journal of Law and Urban Policy*. Vol. 3.

Sunstein, Cass and Thaler, Richard. (2003). "Libertarian Paternalism Is Not An Oxymoron." Working Paper. Joint Center for Regulatory Studies.

Thaler, Richard. (1980). "Toward a Positive Theory of Consumer Choice." *Journal of Economic Behavior and Organization*. Vol. 1. pg 39-60.

Thaler, Richard. (1999). "Mental accounting matters." *Journal of Behavioral Decision Making*. Vol. 12. pg. 183-206.

Thaler, Richard and Sunstein, Cass. (2009). *Nudge*. Penguin Books.

Tversky, A. and Kahneman, D. (1974). "Judgment under uncertainty: Heuristics and biases." *Science*. Vol. 185. pg 1124-1131.

Tversky, A and Kahneman, D. (1981). "The framing of decisions and the psychology of choice." *Science*. Vol. 211. pg 453-458.

VanDeVeer, Donald. (1986). *Paternalistic Intervention*. Princeton University Press.