2016

The Nudge Factor: Paternalism, Autonomy, and the Fight Against Obesity

Sophia Gibert
sgibert@wellesley.edu

The Nudge Factor: Paternalism, Autonomy, and the Fight Against Obesity

Sophia Holland Gibert

Submitted in Partial Fulfillment
of the
Prerequisite for Honors
in the Individual Major, Biology and Society

Wellesley College
April 2016

Acknowledgments

I would first like to thank my Thesis and Major advisor, Professor Jonathan Imber of the Sociology Department, for his guidance and support throughout this process. I cannot thank you enough for suggesting that I read *Nudge* on that day in late October and for encouraging me to follow my curiosity in choosing a topic.

Also, thank you to Professor Emily Buchholtz of the Biological Sciences Department for advising me in my individual major and being a caring and inspiring teacher.

I would also like to thank the members of my thesis committee—Professors Imber and Buchholtz, as well as Professor Helena de Bres of the Philosophy Department, Professor Corrine Taylor, Director of the Quantitative Reasoning Program, and Professor Michèle Respaut of the French Department.

Thank you to the Curriculum Committee at Wellesley for allowing me to pursue an individual major in Biology and Society, as well as to the many professors who have made it possible for me to do so. I am exceedingly grateful for this opportunity and have come to appreciate it even more through this project.

Lastly, thank you my friends and family for your love and support and for coming on this long journey with me. To my Dad, thank you for keeping me sane and for supplying me with an endless stream of nudge-related puns. This would not have been possible without you.

Table of Contents

Introduction

Dr. Brian Wansink, behavioral science expert and Director of the Cornell University Food and Brand Lab, begins his book called *Mindless Eating: Why We Eat More Than We Think* with a full page dedicated to a single sentence: "The best diet is the one you don't know you're on" (Wansink 13). Wansink repeats this sentiment many more times as he suggests ways in which individuals and private companies can "reengineer" environments such that they promote effortless healthy eating and activity. In an environment designed for mindless dieting, dishware is far smaller (Wansink 68); nutrition labels are displayed prominently on all foods (194); snack food packages are designed with physical separations within them to give consumers "pause points," and the manufacturers of Pringles make every fourteenth chip the color red in order to remind us of the serving size (200). Companies quietly reduce the caloric density of their recipes without changing their volumes (201), and healthy foods, fruits and vegetables are labeled with descriptive titles on school lunch menus, referring to their geographic origins, tastes, and associated positive emotions (240).

Although Wansink does not discuss the role of the government in making this environment a reality, he does suggest that policymakers can help, and his ideas are not far removed from those that have been proposed, passed and implemented in the past fifteen years, with the intention of reducing food insecurity and lowering obesity rates. For instance, the U.S. Food and Drug Administration (FDA) requires a Nutrition Facts panel to appear on the back of most packaged foods, and it has recently proposed modifications that would increase the font size of calorie and serving size information and add a line highlighting added sugar (FDA "Proposed Changes"). Researchers in the UK have even advocated for a "traffic light" system,

which would display fat, saturated fat, sugar, and salt contents in red, yellow, or green circles depending on how high they are (UK National Health Service). Food banks in the Feeding America network have adopted strategies for promoting the purchase of healthy foods, including presenting produce first in their stores, placing simple recipe cards near less popular nutritious items, and displaying recommended daily amounts next to fruits, vegetables, and low-fat dairy products (Feeding America). A pilot study in New Mexico showed that placing yellow tape and signs in supermarket shopping carts to designate a section for placing produce increased purchases of these products significantly, without lowering profits for the stores (Behavioural Insights Team 14). Further supermarket-based interventions have been suggested, including re-ordering shelves to place certain items at eye-level, advertising social norm messages on shopping carts, and even having receipts display the total calories, fat, or added sugar purchased in a singe shopping trip, along with average and recommended values and prompts where healthier options are available (Halpern 171-172).

Other policies employ financial incentives, such as the 2014 farm bill's dedication of $100 million over the next five years to an incentive program that doubles the value of SNAP benefits (the new food stamps) when used to purchase fruits and vegetables at local farmers markets (Charles) and the tax breaks given to American food companies for changing the balance of nutrients in their products. This is not to mention more restrictive policies, such as the federal government's nutritional standards for the content of school meals (United States Department of Agriculture), New York City Mayor Michael Bloomberg's proposed law prohibiting the sale of sugar-sweetened beverages in sizes larger than 16oz., and even lawsuits launched against food manufacturers for advertising unwarranted health claims on their

packaging (Kersh). This is only a subset of the numerous practices that have been suggested and enacted for the purpose of improving eating environments.

Certainly, not everyone shares Wansink's good opinion of a world in which healthy eating and exercise are made easy for us: Mayor Bloomberg's ban on large-sized sodas has been called "arbitrary and capricious" (Weiner) "draconian" (Martel), "offensive" and worse, and he has even been nicknamed "The Nanny" (Blodget). Each of the practices described above has faced fierce opposition from those who advocate for laissez-faire capitalism in the name of consumer sovereignty. What motivates these objections is that these policies and others like them are instances of *paternalism*, meaning that they (are intended to) change people's behaviors or decisions in order to promote their wellbeing, and paternalism has traditionally been viewed as a violation of our autonomy and an objectionable practice. Government paternalism tends to evoke especially negative reactions, including the feeling that one is being infantilized, or "nannied" unnecessarily in some area of his life that he should be able to control. Many who defend strong anti-paternalism hold that the government is only justified in intervening with our lives in order to prevent harm to others, and never to prevent harm to ourselves. They often draw support from the views of utilitarian philosopher John Stuart Mill, who famously articulated how the government should and should not act toward its citizens in his influential political writing, *On Liberty,* in 1859. In this work, Mill proposes "one very simple principle, as entitled to govern absolutely the dealings of society with the individual" (Mill 9), which has become known as the "harm principle." It reads, "the sole end for which mankind are warranted, individually or collectively, in interfering with the liberty of action of any of their number is self-protection. […] His own good, either physical or moral, is not a sufficient warrant" (9).

In light of this view, how can the burden of disease associated with obesity be addressed? It has become increasingly clear that something must be done. The rates of overweight and obesity are rising throughout the world and are especially high in the United States: more than two thirds of adult Americans are overweight or obese, with more than one third being obese and more than six percent having what is classified as "extreme obesity" (National Institutes of Health). Approximately 17 percent of children ages two to 19 are obese, though this rate has remained fairly stable for the past decade and has even decreased significantly among children aged two to five. Obesity, which was officially recognized as a disease in itself in 2013 by the American Medical Association (Stoner and Cornwall), increases the risk of heart disease, diabetes, and other chronic conditions, and is estimated to cause fifteen percent of annual deaths and $150 billion of annual health care costs in the US (Downs et al.). Furthermore, obese adults are more likely to experience discrimination in academic and work settings, are more likely to exhibit bulimia and other eating disorders (Wadden and Stunkard), and are more likely to be diagnosed with psychological problems than non-obese adults (Wardle and Cooke).

Strict anti-paternalists might suggest that this problem be addressed through personal behavioral changes, aided in some instances by medical professionals. This approach depends on an understanding of obesity as being caused by individuals' choices rather than by environmental factors. At its most basic level, of course, obesity does result from an individual taking in more energy than is expended. Many people live in so-called "obesogenic" environments and yet do not become obese, indicating that these factors are not *direct* causes of the disease. Still, there are many diseases whose causes do not produce illness in every person they come into contact with, and we do not take this to be a reason not to address or prevent those causes. It is increasingly clear that aspects of people's lives, such as limited access to healthy foods, poor exercise

landscapes, abundance of energy-dense, nutrient-poor foods at restaurants, and work environments that discourage physical activity play an important role in causing obesity, and these environmental factors should not be ignored.

Given that strict anti-paternalism does not allow for a robust response to the burden of disease resulting from obesity, some level of government paternalism should be permitted in this area. The question becomes, then, to what extent is the government justified in intervening with our choices in order to promote our long-term health? There are at least some instances of government paternalism that most people take to be justified. A classic example is seatbelt legislation, which interferes with our freedom to drive or ride in cars without a seatbelt for the purpose of protecting us from injury. Another is the requirement that we obtain prescriptions for certain drugs and medications, which protects us from the consequences of misdiagnosing our own illnesses. A large body of literature has been dedicated to drawing distinctions between various types of paternalism with the hope of determining what makes these instances less objectionable than others and how these features can be controlled for so that some line can be drawn between acceptable and unacceptable government policies. This pursuit has grown substantially in the past ten years in response to a variety of paternalism proposed by behavioral economist Richard Thaler and law professor Cass Sunstein, called "libertarian paternalism" (Thaler and Sunstein "Libertarian…").

Libertarian paternalistic policies, affectionately known as "nudges," aim to change people's actions to promote their wellbeing, without restricting their freedom of choice. In contrast with taxes, incentives, and bans, which cut off their targets' available options in order to encourage certain behaviors, nudges employ insights from cognitive psychology to structure the choosing environment such that people are simply more likely to act in desired ways. The tactics

that Wansink proposes for mindless dieting, for instance, were they to be mandated by the government, would qualify as nudges, as would many of the obesity prevention policies that are currently in place. Any practice that increases the attractiveness, prominence, or social appeal of choosing healthy foods has the potential to nudge people toward better eating patterns without actually forcing them to change their behavior. In this way, libertarian paternalism is intended to avoid the central objection that faces other forms of paternalism, namely that they violate our autonomy, or power to direct our lives. As is indicated by its name, it is meant to be acceptable to libertarians, who are the traditional opponents of government paternalism.

Unfortunately, nudges do not avoid all ethical concerns, as their proponents hoped they would. For one thing, many argue that they still violate the autonomy of their targets, through manipulation instead of coercion. Furthermore, some hold that they are *more* threatening than traditional paternalistic practices because they inevitably lack transparency. In this project, I examine how libertarian paternalism is supposed to avoid the autonomy objection, and I evaluate its success in doing so. Because my aim is not only to uncover what is at stake in defending nudges, but also to find the most promising direction, ethically speaking, for creating and supporting obesity prevention policies, I focus on three specific strategies throughout this discussion: menu labeling regulations, which require calories to be posted on menu boards in top chain restaurants; optimal defaults, which create environments in which healthy eating and exercise are easy and automatic; and sugary beverage excise taxes, which raise the prices of sugar-sweetened drinks in the hopes of reducing their consumption. In the section that follows, I provide some background on these regulations and strategies. I have chosen them in part because they cover a range of the categories of paternalism that will be discussed, and in part because I have had some involvement in the public health work that supports each of them. As a result of

this focus, my project has become as much an opportunity for reflection on my own past and

future role in the field of health policy as it has been an academic pursuit.

Menu Labeling Regulations

The Patient Protection and Affordable Care Act (ACA), passed in 2010, includes a menu

labeling provision that requires all chain restaurants, retail food establishments, and vending

machines with twenty or more locations in the United States to display the calorie contents of

their foods and beverages next to titles and prices on in-store menu boards. The law covers all

restaurant-type foods that are typically intended for one person to eat, including take-out food,

meals purchased at drive-through windows, self-service foods, snacks purchased at movie

theatres, and certain alcoholic beverages. In addition, it requires the inclusion of an "anchoring

statement," which informs consumers of the recommended daily caloric intake. These

regulations were finalized by the Food and Drug Administration (FDA) in November of 2014,

originally to be implemented by December, 2015 (Food and Drug Administration). However,

due to industry opposition, the compliance date was extended by one year (Bleich). Further

delays were recently proposed in an amendment called the Common Sense Nutrition Disclosure

Act of 2015, which has passed in the U.S. House of Representatives and was received by the

Senate in February of this year. If enacted, this amendment would not only extend the

compliance deadlines by a minimum of two years, but would also permit restaurants to

determine appropriate serving sizes for which to post calorie information rather than requiring

them to provide information about full items, would enlarge the allowable margin of error for

calorie counts, and would lower the requirements for proving that measurements are legitimate.

More importantly, it would eliminate certification requirements and would greatly reduce the

likelihood of punishment for non-compliance, as restaurants would no longer be liable in civil action, and those found to be in violation of the regulations would be given a 90-day window within which they could implement changes to avoid all penalties (Library of Congress). Not surprisingly, public health and nutrition experts have issued statements that this act would make nutrition labels more confusing for consumers and would significantly weaken the original legislation (Wootan 5).

As an amendment to the Nutrition Labeling and Education act of 1990, the menu-labeling provision is intended to help consumers make more informed food choices by removing the barriers that currently stand between them and nutrition information. More broadly, policymakers hope that the regulations will empower consumers to seek out nutrition facts in other contexts, will raise awareness of calories consumed outside the home, and will contribute toward lowering obesity rates. They certainly have the potential to do so, since American families now dedicate approximately half of their food-related spending to food eaten outside the home, and each day in the United States, 41 percent of adolescents and 36 percent of adults eat at a fast food restaurant (Bleich). Fast food is known to be more energy-dense than food consumed at home, and children eat almost twice as many calories when they eat out than when they eat the same meal at home (Wootan 1). Thus, if calorie-labeling can lead people to select healthier options at top chain restaurants, then it will have a dramatic impact on adult and childhood obesity.

Because several cities, counties, and private companies have implemented their own guidelines in advance of the federal regulations going into effect, researchers have been able to conduct preliminary studies of their impact. New York City was the first site in the United States to pass and implement calorie-labeling laws in 2008, followed by King County, Washington,

Philadelphia, San Francisco, California State as a whole, and Oregon (Downs et al.). More than

eighteen counties, cities, and states have similar regulations, though most of them have not been

implemented (The State of Obesity). In addition, McDonald's, Starbucks, Subway, and several

other companies have applied their own calorie-labeling rules in response to consumer demand

for transparency and in preparation for nation-wide regulations (CSPI "Menus & Menu

Boards"). Unfortunately, the current body of evidence shows that calorie-labeling does not have

a significant effect on consumers' purchasing decisions (Long et al.). Only about 33 percent of

customers even notice calorie labels, and those who do often fail to modify their behavior. In

some instances, particularly for males, the number of calories purchased actually *increases* when

menus are labeled (Bleich). Although calorie labels have not yet had the impact that they were

intended to have on consumer behavior, they have begun affecting the meals that are offered at

chain restaurants. In 2014, Bleich and her colleagues found that calories in newly introduced

menu items at 66 of the top 100 U.S. chain restaurants decreased significantly by approximately

60 calories after menu labeling regulations were passed, and this trend continued in the following

year (Bleich). It is possible that once all producers are required to post the calorie contents of

their foods prominently, this will lead to a further reduction in calories available, which could

affect obesity rates equally, without placing extra burden on individuals.

In addition to industry opposition, which remains strong, there have been objections to

menu labeling regulations from the general public. Some who oppose the policy maintain that

consumers do not want or need calorie information, since they are able to seek it out online or on

brochures that are available upon request in stores. However, National polls show that up to 83

percent of Americans support calorie-labeling requirements, with several county and city polls

finding even more overwhelming support (CSPI "Summary of Polls…"). Other opponents argue

that the regulations will not be effective enough to make an impact on the obesity problem and

are thus not worth the costs that they impose on restaurants. Although it is worrisome that the

effects seen thus far have been minimal, there is great potential for the impact to grow when

national regulations are implemented. Advocates of menu labeling recognize that the laws are

not enough to reverse the obesity epidemic on their own, but they maintain that they are one

important component of a complex solution. Furthermore, the costs imposed on restaurants are

minimal. The laws are aligned with requirements that the government already places on business

chains in other sectors, and complying with them is both practically and financially feasible for

restaurants, especially since they only apply to top chains, which tend to have highly

standardized menu items (Wootan 3). Still others object that calorie postings violate our freedom

to choose how we eat and that it is not the government's place to control our decisions regarding

what to order in restaurants. This, of course, will be explored in the chapters that follow.

Optimal Defaults

Optimal defaults structure decisions such that, in the event that no choice is selected, the

result is beneficial for the chooser. In certain decision-making contexts, the default option can be

neutral, where the chooser is required to make some choice in order to move on with a process;

however, this "forced choice" structure is typically only possible for very formal decisions,

where every person is presumed to encounter them. Most decisions must be structured with an

opt-in or opt-out format, or with some choice being designated as the most reasonable one. Using

optimal defaults is one way to respond to this necessity, as opposed to selecting defaults at

random or to serve some end other than the target's interests.

Some take the U.S. organ donation system to be an example of optimal defaults at work. It functions on an opt-in choice structure, where people can declare themselves organ donors when they obtain a driver's license. In the event that someone has no license, is missing their license upon arrival at a hospital, or has family members who are unaware of her organ donation status and strongly disagree with what is indicated on the license, it is assumed that she is *not* an organ donor (List). Supposing that it is preferable to most people not to be an organ donor, then the current system for declaring organ donor status involves an *optimal default*. As I will explain in Chapter 2, a thoughtfully designed default option can be a powerful tool for leading people toward a certain behavior. In some countries that operate on an opt-out structure for organ donation, nearly all people accept the default and declare organ donor status, while in the United States, this number is closer to 45 percent and is far lower in some states. Similarly, research shows that when enrollment into a pension plan is made to be the default option for employees, 85 percent of employees accept the enrollment status, whereas when employees must actively opt in to enrollment plans, only 26 to 43 percent do so (Choi et al.).

In the context of obesity prevention, the concept of optimal defaults is used to refer to all ways in which the environment can be reengineered to promote health, or at least to be non-obesogenic. With this understanding, optimal defaults refer to nearly all obesity prevention policies, including nudges, taxes, and restrictions, as long as they restructure aspects of people's daily lives such that it is easier to maintain healthy behaviors. However, for the purposes of this discussion, I use the term optimal defaults to refer to strategies that target ordering decisions made in restaurant settings. These policies might require that the drinks and sides that come with restaurant meals to be the healthiest available choices or that default portion sizes be smaller. Although optimal defaults at restaurants have not been studied extensively as a strategy for

obesity prevention, they have been observed in combination with other interventions. Researchers at ChildObesity180, a non-profit organization associated with Tufts University, examined the changes resulting from the implementation of a healthier children's menu at the Silver Diner restaurant chain, and their results demonstrate the potential for restaurant defaults to promote healthy ordering patterns. In this study, the authors observed that the implementation of a kids' menu featuring grilled chicken breast, fish tacos, spaghetti, and fresh strawberries resulted in significantly higher orders of meals meeting nutritional requirements and significantly lower orders of French fries and soda. Less healthy options remained on the menu, and switching from default options to less healthy options imposed no extra cost, and yet defaults had a considerable effect on the results. As the authors note, parents generally accepted defaults for their children's meals, and the total calories ordered decreased significantly after implementation *only* for those parents who did (Anzman-Frasca et al.).

Optimal defaults for use in restaurant settings are promising not only for the reasons described above with respect to menu labeling requirements, but also for two additional reasons. First of all, they place less of a burden on individuals to change their behaviors. While calorie-labeling laws require uptake by their targets, optimal defaults sway consumers at a less conscious level. Second, they are essentially free to implement. Thus, if the government were to require restaurants to use optimal defaults to promote healthy ordering patterns, there would be less concern regarding cost to producers.

Sugary Beverage Taxes

In recent years, the placement of an excise tax on sugary beverages has gained momentum as a strategy for obesity prevention, both in the United States and internationally. In

2013, the Mexican government passed the world's first country-wide excise tax on sugar-sweetened beverages as part of its federal budget, which was fully passed onto consumers and resulted in a price increase of approximately 10 percent, or one peso per liter (Colchero et al.). Thus far, the results in Mexico have been very promising. After only one year, sugary drink purchases dropped by 12 percent, decreasing at an increasing rate throughout the year (Colchero et al.). Even those who were previously skeptical about the potential for Mexico's tax to reduce calorie consumption have become convinced that excise taxes on sugar-sweetened drinks should be a widespread component of the response to obesity. Policymakers in the U.K. followed suit in March of this year with the announcement of an excise tax on sugar-sweetened beverages, the funds from which will be directed toward primary school education (Young). In the United States, Berkeley, California became the first city to pass an excise tax on soda and other sugar-sweetened beverages in 2014, with support 76 percent of voters (Barclay).

Although 34 U.S. states and the District of Columbia have applied sales taxes to sugar-sweetened soda sold in stores, and even more have applied sales taxes to soda sold in vending machines (Bridging the Gap), no state has passed an excise tax yet. Excise taxes are collected from manufacturers and are largely passed on to consumers through price increases, in contrast to sales taxes, which appear at the bottom of shopping receipts along with other information. From the perspective of public health, it is important that policies targeting obesity-promoting beverages be in the form of excise taxes because they are far more likely to be noticed by consumers and to have an effect. Research shows that, while sales taxes on soda are largely ineffective (Sturm), excise taxes that are high enough to result in a 20 percent price increase will decrease calorie consumption enough to result in significant weight loss (Novak, Finkelstein, Dharmasena).

Many researchers are hopeful that these policies will grow in popularity in the coming years, and this hope is not unfounded. Like the policies discussed above, soda taxes are supported by the overwhelming need for some action to prevent obesity. In Mexico, rising disease rates were influential in supporting the tax—the years preceding its implementation saw dramatic increases in the rate of type 2 diabetes, which is directly linked to excess sugar consumption. Furthermore, soda consumption is already falling in the United States and has declined by 25 percent in the past 20 years, which demonstrates that people may be ready to cut down on its consumption even more (Main). Unlike menu labeling regulations and optimal defaults, soda taxes target a single item that is known to have no nutritional value whatsoever. Added sugar has been increasingly recognized as a leading contributor to obesity, and the most recent Dietary Guidelines for Americans call for people to limit their intake of added sugars and shift away from consuming sugary beverages (Health.gov). There are 15 to 18 teaspoons of added sugar and over 240 calories in a regular 20-ounce bottle of soda, with large sweetened fountain drinks containing up to 700 calories. Half of Americans consume sugary beverages on any given day, and those who drink one to two cans of soda a day or more are at a greater risk of developing type 2 diabetes than those who drink soda rarely. In addition to this, sugary drinks are the greatest source of calories for American adolescents, and children get almost 11 percent of their daily calories from them.

One significant challenge of getting soda taxes passed and maintained in the United States is that the laws will be implemented by city and county and thus will not yield as dramatic results as have been observed in Mexico. If the taxes only apply in certain locations, consumers will be able to travel short distances to purchase soda elsewhere at a lower price. This will reduce the effect of the taxes not only because consumers will be able to purchase the same amount of

soda as they would have purchased previously, but also because it will decrease the extent to which manufacturers are willing to pass excise taxes on to consumers in the form of price increases. This has already been observed in Berkeley, where less than half of the excise tax was passed on to consumers. This problem of people being led elsewhere to purchase the same products is what ultimately led the Danish government to lift their "fat taxes," which applied to all foods with a saturated fat content above 2.3 percent. When the tax went into effect, the government saw increasing numbers of Danes crossing the border mainly to Germany, where cheese prices were up to 30 percent lower (Kliff).

The background information provided here is not only relevant to current events (and upcoming voting ballots), but is also central to the discussion that follows. I return to these three strategies at all stages of exploring the ethical implications of using libertarian paternalism for the prevention of obesity and evaluate them in full in my concluding remarks. I begin by discussing the concept of paternalism itself, which is known to have a complex and evolving definition. Next, I consider the critical distinction between "means-related" and "ends-related" paternalism and review the psychological evidence that makes this distinction possible, especially in the context of food and consumption. Then, I provide a deeper understanding of libertarian paternalism and the related concept of "asymmetric paternalism," explaining how these practices interact with means and ends paternalism. Lastly, I examine the concept of autonomy and the autonomy objection in more depth and discuss two final and influential strategies for defending paternalism against it.

My preliminary conclusion is that the policies in question here can be justified as permissible instances of hard paternalism and, if the proper evidence is accumulated, can be

qualified as instances of soft paternalism. However, their permissibility is in virtue of their status as means-related and their potential for yielding substantial benefits for their targets, rather than in virtue of their status as nudges. Libertarian paternalism, I argue, is not as philosophically relevant as it is often portrayed to be. Furthermore, it faces challenges that do not apply to more restrictive varieties. Still, there are practical reasons to prefer the use of nudges—in general, people are insistent that the freedom to choose must be preserved, even when they do not hope or plan to choose the alternatives that they insist must remain. This reaction may be especially strong in the context of food choice, which holds moral weight and lies close to our identities. In light of this, we may do best to defend obesity prevention policies on the basis of the harm principle after all, making the case that firms cause harm to the public when they employ manipulative strategies to increase profits. All this being said, any good project prompts more questions than it answers, and this has been no exception.

Chapter 1
Defining Paternalism and its Varieties

In order to examine the different kinds of paternalism that have been proposed and the ways in which they are intended to avoid the autonomy objection, an understanding of paternalism itself is essential. Defining paternalism is not a straightforward task, and its definition continues to evolve within the philosophical literature. Some even argue that it will remain elusive unless other more fundamental philosophical disagreements are settled first, such as those regarding the definition of autonomy, the nature of rights, and the purpose of the government. In their introduction to a collection of works about paternalism, philosophers Christian Coons and Michael Weber defend their field of study against this worry, writing, "Careful reflection on paternalism's moral status may help illuminate or adjudicate debates about these deeper issues" (3). Through this project, I have come to agree with them, and it is clear that there is much agreement about this among others in the discipline, as the body of work addressing paternalism and its varieties has grown substantially since the 1970s.

Before moving toward a working definition of paternalism for this discussion, it is necessary to establish that obesity prevention policies in general and the three policies in question here—menu labeling regulations, optimal defaults, and sugary beverage taxes—are truly paternalistic. Paternalism tends to evoke powerful negative reactions from the public, perhaps because it is perceived to make a judgment about the target as incompetent, irrational, or a child, or perhaps because there is something morally objectionable about it. Whatever the reason, even those who support certain instances of paternalism are well aware that promoting any policy as being "for people's own good" is likely to be disastrous. In fact, policymakers regularly go to great lengths to justify their recommendations on non-paternalistic grounds. Even

the classic paternalistic law that requires motorcyclists to wear safety helmets was upheld by the

Rhode Island Supreme Court in the 1970s on grounds that align more with Mill's harm principle

than with paternalism. The state claimed that it could reasonably prohibit a person from doing

things that lead to his being put in jail because the cost of his incarceration requires public

support (Dworkin *The Monist* sec. I). Although this justification seems far-fetched, it is not

surprising that it was employed instead of one referring to the safety of individuals. It is

commonly recognized that policies against smoking and other tobacco use became most

successful once they were justified with evidence about the detriments of second-hand smoke

rather than with reference to the health risks taken on by smokers (Brandt ch. 9), and almost

every single public health research paper that supports an obesity intervention appeals to non-

paternalistic justifications within the first few paragraphs of the introduction by citing not just the

number of people who are overweight and obese, but also the medical costs that the problem

imposes on the public.

Can obesity prevention policies be defended in similarly uncontroversial ways? Julian Le

Grand, Professor of Social Policy at the London School of Economics, and Bill New, an

independent policy analyst working in the U.K., jointly argue that harm-based accounts rarely

justify seemingly paternalistic policies in *full* (Le Grand and New 53-59). For example, if the

government bans excessively dangerous behavior, which we might think is paternalistic, this is

often defended on the basis that it reduces medical expenses for society or even protects people

from the undue psychic harm that they would experience if forced to witness a violent accident

or watch a loved one suffer from a preventable chronic disease. However, if the true motivation

behind requiring motorcyclists to wear helmets or requiring swimmers to be in the presence of a

lifeguard were to reduce costs to society, then these would not be the policies in place, as these

goals could be accomplished in other ways. For example, requiring people to pay a fee toward the costs of their medical treatment or rescue beforehand could offset medical costs (62). Additionally, if the true motivation behind these policies were to diminish psychic costs, then this would not be enough to have them passed, since limiting all activities that result in psychic costs to someone would be unreasonably restrictive (63).

The most promising non-paternalistic way in which to justify obesity prevention policies is by way of extending the harm principle to include market failures and inefficiencies as harms to the public. For example, we might think the public is harmed by the pollution that a firm produces, so in taxing that firm to account for the negative externality of pollution, the government might be said to protect the public from harm. Similarly, the government's subsidies of underfunded public goods might be justified on non-paternalistic grounds, if this is said to be done in order to prevent members of the public from harming one another through their free-riding tendencies. One might think that menu regulations can be justified in this same way, as a method of correcting the market failure of imperfect information. However, as critics have emphasized, menu labeling does more than simply correct market failure—it emphasizes certain facts over others that people are then encouraged to preferentially attend to for their own good (Le Grand and New 68). Menu labeling imposes on people a certain value—the value that calorie content is as important as price when ordering in restaurants.

The case is equally shaky for optimal defaults and sugary beverage taxes. Optimal defaults might be justified as non-paternalistic in cases where selecting some default is inevitable. As Thaler and Sunstein point out, it is often impossible to avoid designating something as the default, and in these cases, it is better to select the healthful choice than to select something at random or to select something that maximizes a company's profits (*Nudge* 1-

3). However, even if these were the only possible ways in which to select a default, which it is not clear they are, selecting a default based on maximizing consumers' wellbeing would still qualify as paternalistic. Lastly, one might hold that sugary beverage taxes are non-paternalistic because they offset the negative health-related externalities imposed by companies that sell soft drinks. However, the taxes collected amount to far more than would be needed to offset the medical costs imposed by drinking soda. Although this calculation is difficult to make, it is at least true that the taxes are not directly channeled toward reducing insurance costs for society. The revenues from the taxes are being redirected in some cases toward other public health efforts to prevent obesity and in other cases toward different areas entirely, such as primary school education.

Defining Paternalism

In examining paternalism and the question of its justification, authors have used a wide variety of definitions in a wide variety of contexts. Gerald Dworkin, professor of moral, political, and legal philosophy, has written on paternalism numerous times over the past forty years and has revised his own definition throughout this time, claiming in a recent entry in the *Stanford Encyclopedia of Philosophy* that scholars should carefully adopt different definitions of paternalism in different contexts based on the particular circumstance being examined (3). Le Grand and New explain that defining the concept is controversial primarily because changes in its definition determine how easily and in what cases it can be justified (7). The first issue at stake is whether the definition should be descriptive or normative. Ideally, any definition will be normatively neutral unless there is considerable reason for it not to be (Dworkin "Defining..." 24-25). Many claim that there is nothing fundamentally problematic about paternalism and even

nothing offensive about its use in certain contexts. However, it is common in everyday practice to call something paternalistic as a way of condemning it, and if there is something objectionable about the concept in itself, then we may do best to include a moral evaluation in its definition, similarly to how we might define murder as "the *wrongful* taking of a human life" (25). Choosing between these approaches certainly affects where the burden of proof lies: morally neutral definitions require anti-paternalists to explain what is worrisome about the concept, while more traditional definitions require the paternalists to justify and defend their policies (Dworkin *Stanford Encyclopedia* 5). Because I aim to explore whether paternalism is justifiable, it is most reasonable to begin with a merely descriptive definition and proceed from there.

In its simplest form, paternalism is considered to be the practice of interfering in the outcome of someone's life for that person's own good. We might interfere in someone's choices or actions, but we may also do something to him, such as when a doctor performs a treatment on a patient or when a mother gives her child a bath. Some allow that a failure to act or an omission of information can be paternalistic, such as when a professor does not recommend a student for an academic program she knows will be "out of his league" (Dworkin "Second Thoughts" 106) or when a doctor lies to his patient about a terminal illness (Dworkin "Defining Paternalism" 26). Some authors evaluate whether a practice is paternalistic based on the actual or hypothetical intentions of those who enact it, while others consider its most plausible explanations or justifications (26). One source of variation is that the term "paternalistic" can be applied to acts, to entities, to motives, or to reasons, and the actors can include individuals, private companies, or governments. This provides an enormous range of cases for one definition to explain, and indeed, new definitions tend to be motivated by the failure of a past definition to include intuitively paternalistic cases or exclude intuitively non-paternalistic ones.

The traditional definition of paternalism has three components: (1) the practice restricts or limits the target's liberty or freedom of choice, typically by imposing constraints; (2) it is justified only on the basis that it promotes the target's own good or furthers his interests; (3) it is imposed in the absence of the target's consent (Le Grand and New 8). Although the second and third aspects face some practical issues, authors have largely kept them constant while challenging the first component, that paternalism must involve some restriction of liberty. In a 1972 article, Dworkin defined paternalism as "the interference with a person's liberty of action justified by reasons referring exclusively to the welfare, good, happiness, needs, interests, or values of the person being coerced" (sec. I). The notion of coercion is included because, on a standard view, violations of freedom occur either through physical control or through coercion, which involves forcing someone to do something by removing all other reasonable options. Dworkin's description of paternalism in 1972 as a restriction of liberty was met with objections over the next ten years, as authors proposed intuitively paternalistic cases that involved neither force nor coercion. To be fair, Dworkin had noted in his original piece that there were cases of paternalism that fell outside the bounds of his definition but had nonetheless focused on coercion as the main component because it fit with the specific cases he was examining (sec. II). Despite this reasoning, in 1983, he recognized the complaints and expanded his definition of paternalism, meanwhile evaluating the attempts that had been made to revise it (105-106). Coming to similar conclusions, Le Grand and New review the range of definitions that have been defended since Dworkin's first article (ch. 2). From these accounts, we can understand the complexity of the concept and can begin to predict the ways in which different definitions will lead to different normative conclusions.

Suppose that an individual whose friend wants to play a game of tennis declines to play because he knows it upsets her that she always loses. Here, he refrains from playing mainly for her own good, and what he does seems paternalistic even though he does not force or coerce her and does not restrict her freedom. The traditional definition sorts this case incorrectly according to our intuition. To avoid this, we might require that a paternalistic intervention reduce the number of choices open to its target rather than restrict its target's liberty. Along these lines, David Archard proposes that "choice or opportunity to choose is denied or diminished" through paternalism (36-42), while Simon Clarke requires that the paternalist try "to close an option that would otherwise be open to" the target, in order to promote his or her good (Dworkin "Defining Paternalism" 30). Unfortunately, though this revision categorizes the tennis case well, it fails to capture other cases that we consider paternalistic. For instance, we might think that government policies supporting free admission to museums are paternalistic. Presumably, such policies are meant at least in part to encourage people to visit museums more, for their own cultural and intellectual benefit. Here, the practice does not restrict the target's freedom, and it actually serves to *increase* the number of choices available to those deciding how to spend their incomes.

Another way to revise the first component of the traditional definition—one that is defended by bioethicists Bernard Gert and Charles Culver—is to require that paternalism involve a violation of its target's rights or the violation of some moral rule. The doctor who refrains from telling his patient about a terminal illness certainly violates some right to information, while the tennis player surely disrespects his friend's wishes and is dishonest. Alas, this definition also fails to capture importantly paternalistic cases. Someone who hides his sleeping pills before a depressed friend visits his house in the hopes of preventing him from committing suicide acts paternalistically, though he does not violate any moral rule, nor does he violate his friend's

rights, since the pills are his own (Le Grand and New 9). Although he does not violate any moral

rules, he does treat his target in a way he does not wish to be treated. Dworkin suggests this as an

alternative to the traditional definition but immediately points out that it, too, would capture the

wrong cases, this time including extra cases that we would not consider paternalistic ("Second

Thoughts" 106).

A promising idea is to revise the first component of the definition so that paternalism

must interfere *either* with the target's liberty *or* his or her *autonomy*. In 1983, Dworkin writes,

"there must be a violation of a person's autonomy (which I conceive as a distinct notion from

that of liberty) for one to treat another paternalistically" (107). Autonomy is typically conceived

of as the condition of self-rule and will be discussed in much greater depth in Chapter 4.

However, it is important to note here that autonomy differs from liberty. This distinction is best

demonstrated by instances in which the two concepts come apart: for example, someone who is

severely paralyzed is restricted in her freedom to move and perform certain actions, but is not

restricted from making autonomous decisions about what happens to her; alternatively, a doctor

who tells her patient about the result of a medical test he has asked not to know about violates his

autonomy but does not restrict his freedom. Le Grand and New hold that Dworkin's autonomy-

based definition successfully captures the cases that we consider paternalistic. The tennis player

violates his friend's autonomy by undermining her decision to risk losing a game of tennis, and

the government violates people's autonomy by ignoring their judgments about the value of

museums. Each of these, of course, is still done without the target's consent and with the goal of

promoting his or her own good.

Interestingly, it is not clear that Dworkin himself considers an autonomy-based definition

to be satisfactory. In 2014, he cites the case of a wife hiding her sleeping pills from her suicidal

husband as an example of paternalism that limits neither freedom nor autonomy. He also

references an example in which a father gives money to one trustworthy child to be used in the

best interests of another, less trustworthy, child. Here, he writes, "there does not seem to be an

interference with the child's liberty nor on most conceptions the child's autonomy," and the

practice is paternalistic nonetheless (2). Thus, a definition of paternalism that includes a violation

of autonomy does not capture the correct cases. Furthermore, Le Grand and New argue that a

definition based on the violation of autonomy is too ambiguous. They argue that the terms

"interference" and "autonomy" are understood in a range of ways, leaving the definition of

paternalism vague (14). Of course, there is an additional compelling reason to avoid including

the violation of autonomy in the definition of paternalism: to do so is to concede that

paternalistic practices do indeed offend our autonomy. Each of the varieties of paternalism to be

introduced in the chapters that follow plays a role in supporting the view that some instances of

paternalism avoid the autonomy objection, and most do so by establishing a subset of cases that

are paternalistic and yet do not involve a violation of autonomy. Thus, to include a violation of

autonomy in the definition resolves the question to be addressed in this discussion before we

have begun.

Given this, the array of proposed definitions is unsatisfactory: paternalism is not by

definition a violation of liberty, nor is it a force that always decreases the number of available

choices, violates a moral rule, or offends autonomy. How, then, is it to be characterized?

Dworkin made another significant claim in his 1983 paper, immediately after asserting the

necessity of a violation of autonomy: "there must be a usurpation of decision making, either by

preventing people from doing what they have decided or by interfering with the way in which

they arrive at their decisions" (107). This requirement that the paternalist substitute his or her

judgment for the one made by the target is central to the revision that Le Grand and New settle on as well. They argue that what makes a policy paternalistic is that the government considers the individual's judgment to be insufficient and replaces it in order to further his or her own good (14). It is also a main component of the view put forth by professor of philosophy Seana Shiffrin (Dworkin "Defining Paternalism" 31).

This revised definition does well in sorting situations according to our intuitions about paternalism. In the case of the tennis player, he judges his friend's decision to play tennis as insufficient for making her happy. When the government subsidizes museum entrance fees, it overrides the citizens' judgments about the value of visiting museums. The doctor who hides a terminal illness from his patient judges that his patient's wish to know about his own condition is inadequate, and the wife who hides her sleeping pills from her husband assumes that his decision to end his life would be deficient. It also fulfills the requirement of being morally neutral, at least on its surface, and of withholding judgment on the question of whether paternalism violates autonomy. If we accept this as the third component of the working definition, then what else must be included?

The original component requiring that paternalism violate a target's freedom of choice or liberty was accompanied by two other premises: that the practice be justified only on the basis that it promotes the target's best interests, and that it be enacted without the target's consent. The first of these is essential, though it should be noted that it makes no reference to whether people's best interests are to be understood in an objective or subjective way. This will be the source of the first distinction, between ends-related and means-related paternalism, which is the subject of Chapter 2. The final component of the definition, which refers to the target's consent, is might be superfluous in combination with certain definitions. It is certainly superfluous for definitions that

include a violation of autonomy, as we generally understand consent to be given autonomously.

Even if we take on a definition of paternalism that requires that the actor substitute his own

judgment for the target's reasoning, we no longer need a clause referring to the lack of consent

because it does not make sense for someone to consent to his own judgment being overruled.

The only reason to consent to this would be if he realized the failure in his reasoning that

prompted the intervention, and if this were the case, he would adjust his behavior himself.

Thus far, I have considered the various definitions authors have defended for paternalism.

Although it may be tempting to view the debate as an evolution from a normative traditional

definition to a more morally neutral one, this would be overly simplistic. Many authors continue

to defend positions different from the one I have settled on, and these differences affect their

defenses and/or critiques of paternalistic policies. Going forward, I adopt the definition of

paternalism as a practice that addresses an insufficient judgment and does so to promote the

target's good, but it will become clear that other authors settle on different conceptions of

paternalism and that these differences have important implications on their views.

Further Distinctions

Beyond a working definition of paternalism, there is need for further helpful terminology.

Philosophers have categorized paternalism extensively, but the most useful distinctions for my

purposes will be means-related versus ends-related, soft versus hard, unmixed versus mixed, and

pure versus impure. First, the distinction between means and ends paternalism is of great

importance in many defenses of paternalistic government policies. In cases of means

paternalism, the individual's end goal is taken to be legitimate, and the intervention is concerned

only with promoting this goal. Alternatively, ends paternalism seeks to change the individual's

behavior in a way that leads to some goal that is considered to be objectively good for him. Thaler and Sunstein defend only means-related paternalism, as do all of the other authors whose views are considered here. This view is only coherent, though, if people do not always reach their goals on their own. Indeed, Dworkin implies in 1983 that *if* people *do* always reach their goals on their own, which is why the body of evidence showing that people consistently and predictably fail to reach their own ends is so relevant. These findings, which will be discussed in the next chapter, serve as the main justification for means-related paternalistic practices now, while the lack of evidence for the idea that people can be mistaken about their ends has become equally important in showing that ends-related paternalism is unjustifiable.

Unfortunately, the distinction between means and ends paternalism is rarely clear. Individuals might have multiple ends at once or over time that conflict, and a policy that enforces the importance of one of these over another and does not allow the individual to rank the two in a different manner might be considered ends-related. For instance, a policy that places quality and safety requirements on products prioritizes the end of safety over the end of low price and does not allow consumers to purchase riskier products at lower prices, even if these consumers have made a well-reasoned judgment that this decision would be better for them. This becomes even more complicated if the individual's contradicting ends overlap in time, such as when someone wants to quit smoking but also wants to live a long and healthy life (Le Grand and New 30). These short-term desires are referred to as first-order desires in the philosophical literature (e.g. the desire to smoke a cigarette), while long-term plans and preferences are called second-order desires (e.g. the desire not to desire to smoke a cigarette). Sarah Conly, professor of philosophy at Bowdoin College, even argues that there is no significant line to be drawn between means and ends paternalism. According to her, the difference between acting fully rationally and acting on

some mistakes in reasoning is a difference in degree, rather than in kind (Conly 6-7). This difficulty of characterizing people's desires as ends or as fleeting preferences recurs throughout this discussion.

The distinction between soft and hard paternalism is also central to discussions of paternalism in relation to the autonomy objection. In a general sense, soft paternalism is the view that some instances of paternalism avoid violating autonomy and that only these cases are morally permissible. Soft paternalists hold that acts of paternalism do not violate our autonomy as long as they interfere with actions that we are already performing non-autonomously. Thus, these scholars must explain which actions are done sufficiently non-autonomously such that they can be intervened with. Philosopher Joel Feinberg is the most famous proponent of this view, and he has created a framework for distinguishing these cases that relies on features of the action being performed in general, as well as of the individual performing it. In contrast, hard paternalism is the view that paternalism can violate our autonomy and still be permissible because autonomy is not important enough to override the importance of the wellbeing we stand to gain from some policies. Sarah Conly, who defends a strong account of hard paternalism, argues that certain instances of paternalism that do violate our autonomy are justifiable because they promote more benefits than costs for the targets and do not interfere with liberties that are central to our identities. I elaborate on his view in Chapter 4, along with Feinberg's theory of soft paternalism.

Two final distinctions are of particular importance in the context of obesity prevention: that between pure and impure paternalism, and that between mixed and unmixed cases. As Dworkin writes, "in pure paternalism the class being protected is identical with the class being interfered with, e.g. preventing swimmers from swimming when lifeguards are not present." In

contrast, impure paternalism interferes with one entity or group of people in order to promote the interests of a separate, smaller group (Dworkin *The Monist* sec. III). Usury laws are a classic example of impure paternalism, as they forbid people who loan money from loaning money at very high interest rates with the purpose of protecting people who would tend to enter into such risky contracts (Le Grand and New 17). Most of the obesity prevention strategies mentioned previously place restrictions on industries, fast food restaurants, and schools, rather than directly on people, which qualifies them as impure.

This is a source of tension, as most objections to these policies come from these third parties. Although the practices are libertarian-paternalistic toward people, they are restrictive toward the private sector. Is this problematic? It is possible that impure paternalism requires stronger justification than pure paternalism, demanding even more conclusive evidence that the policy will have a significant benefit to individuals' welfare. I will return to this question at the conclusion of my project. Lastly, while unmixed interventions can be justified only and entirely on paternalistic grounds, mixed policies have some paternalistic and some non-paternalistic justifications. As explained earlier on in this chapter, the cases in question here most plausibly qualify as mixed paternalism, as do most paternalistic government policies that exist today.

These distinctions will become clearer in the chapters that follow. As I will show, each is integral in supporting the case that some acts of paternalism avoid the autonomy objection, one way or another, which is the central objection to the three obesity prevention policies in question. In the next chapter, I elaborate on the distinction between means-related and ends-related paternalism, which requires evidence that people systematically fail to reach their own ends. Since all of the authors whose work I consider throughout this discussion are means-paternalists,

they all cite the work of psychologists and behavioral economists in order to demonstrate

people's regular failures to get what they want. Although they organize the various types of

evidence into different categories, the points are fairly consistent across all. In the next chapter, I

will describe the most important of these findings and examine which reasoning failures are in

play when we act in ways that lead to overweight and obesity.

Chapter 2
Behavioral Economics and the Case for Means Paternalism

According to traditional economic theory, individuals are rational agents who make

decisions by considering their preferences and options and choosing whatever will maximize

their utility, or happiness. These fully rational beings, whom we now understand to be more

theoretical than realistic, are what scholars call "homo economicus," economic man, or "Econs,"

for short. Assuming they have all the information relevant to a situation, so-called Econs always

do what is best for them within their constraints, with the exception of some random errors. As

the authors Thaler and Sunstein write, "to qualify as Econs, people are not required to make

perfect forecasts (that would require omniscience), but they are required to make unbiased

forecasts. That is, the forecasts can be wrong, but they can't be systematically wrong in a

predictable direction" (Thaler and Sunstein *Nudge* 7). Econs' ability to make reasoned decisions

and predictions about their own preferences, then, are not diminished by the framing of

information, by the timing of an action's consequences, or by the decisions made by others

around them (Le Grand and New 80).

This was not always the dominant model for describing human decision-making. The

scholars whose ideas underpin the discipline of economics were truly interested in explaining

people's actions and did not ignore the idiosyncrasies and features of our reasoning that will be

presented in this chapter. For example, Adam Smith, known as the father of economics, reported

on our tendency toward loss aversion. It was not until the turn of the twentieth century that

economists became increasingly concerned with establishing their field as one of rigor and

principle, and as a result, they attempted to separate it from the field of psychology, which was

seen at the time as unscientific relative to the natural sciences. Due to this shift away from

psychology, both economic theory and the policies that stemmed from it stuck more rigidly to the assumption of perfect rationality (Camerer, Loewenstein, and Rabin 5-6). Later on, some economists began to reincorporate psychology into their economic models—for instance, Herbert Simon, one of the most influential social scientists of the century, argued in 1947 that an accurate predictive model of human behavior could not be created without combined insights from economics and psychology. Contrary to the traditional model, Simon held that people have limited cognitive abilities that lead them to make sub-optimal decisions, even in the presence of perfect information. Rather than abandon the idea that humans are rational beings, he proposed a new model of behavior based on the notion of "bounded rationality." On this view, people use reasoning tools in very limited ways because of their limited technical abilities, biases toward certain groups, and positions in social hierarchies (Le Grand and New 83).

While economists such as Simon were challenging the field's assumptions from within, others developed more and more detailed predictive models that rested on the foundation of perfect utility-maximization. These models, however, stood in stark contrast with the work of psychologists of the time, who began framing their findings as direct evidence of our irrationality. Most famously, psychologists Amos Tversky and Daniel Kahneman published studies in 1979, revealing numerous factors that influence our decisions in an illogical way (Samson 1). Evidence of this kind amassed, and a conference was held on the subject in 1986, allowing researchers to share their results. This became a milestone in the history of what is now known as behavioral economics—a field with the goal of developing models that utilize psychological findings to make accurate predictions of our behavior (Camerer, Loewenstein, and Rabin 7).

Referring to cognitive biases as *irrational,* or as reasoning *failures,* is not uncontroversial in part because it seems to condemn them. The heuristics presented in this chapter are said to exist because they either were or are helpful in making life-sustaining decisions very quickly, so it is not as if we hope to do away with them entirely in making government policies. In the 1990s, Daniel Kahneman established his "dual-system theoretical framework," which explains our thought process in terms of two systems and emphasizes the importance of both. According to his model, System 1 processes information in an intuitive, automatic way, while System 2 processes it in a more reflective, analytical way (Samson 2). When we use System 1 to make choices, we are able to make them very quickly and successfully, such as when we dodge a ball that has been thrown at us or when we run from something that we quickly judge to be dangerous. We are prone to using these same processes in the context of more complicated decisions, which leads us to act in consistently suboptimal ways, but the systems themselves are advantageous (Thaler and Sunstein *Nudge* 19-20).

It is unclear whether to judge this misapplication of heuristics as rational or irrational, which makes it difficult to assess our choices. It is especially difficult to judge decisions that are influenced heavily by biases, such as those that benefit others, those that serve present-oriented desires over future-oriented needs, and those that involve behaving in a socially acceptable manner at the expense of other aspects of welfare. Because a central goal of economics is to accurately predict behaviors on a large scale, findings that are considered rigorous and consistent enough are often adopted from behavioral economics into mathematical economic models. Some then consider them to be rational ways of acting, while others continue to view them as reasoning failures. As this occurs, it becomes even less clear what we consider to be rational. Despite this uncertainty, many of the tendencies discussed in this chapter are clearly irrational,

allowing them to serve the purpose that they do in supporting a variety of paternalism called means paternalism.

Under the assumption that we are utility-maximizing agents, all paternalism imposes ends on its targets. If we are fully rational, then we will obtain the maximal level of welfare as we define it, as long as we have enough information. So, if the government has any role to play in ensuring that people can maximize their own welfare, then the only way for it to do so is to guarantee that there is perfect information available to all. If the government interferes to *change* our behavior to promote our wellbeing, it will be clear that this actually promotes some idea of wellbeing that we do not endorse. In this case, the government's intervention will impose values on us, or will be ends-paternalistic. Ends paternalism poses a threat to our autonomy regardless of whether one takes autonomy to be a right or a component of welfare. If autonomy is an inalienable right, then violating it to promote any good is unjustified. If autonomy is a component of wellbeing, then any justifiable act of ends paternalism will bring enough good to the individual to offset his lost autonomy. Since the result of an ends-paternalistic policy is by definition not something that the individual considers to be good for him in the long run, it must be good for him in an objective sense. In other words, the ends paternalist commits to an objective standard of welfare—a position that, although not indefensible, is problematic.

This view takes the form of either moral perfectionism or welfare perfectionism, where moral perfectionists maintain that a good life requires living without sinful behavior, and welfare perfectionists hold that pleasure, knowledge, or some other goal or set of goals is necessary and sufficient to live well. Clearly, the moral perfectionist faces the challenge of explaining what behavior is sinful and why living a moral life maximizes a person's wellbeing during life. Meanwhile, the welfare perfectionist must determine exactly what is objectively good for all

people and establish why it is so (Conly 106). This is further complicated if the standard of welfare involves multiple good things, as these must be ranked in a coherent way. These views are difficult to defend because it is strange to think that someone's life may be going well without her feeling that it is going well. While some may have a subjective standard of welfare that matches whatever objective standard is imposed on them, many will not, especially if that standard becomes deranged from how we generally think of wellbeing (104). A related problem is that allowing for a perfectionist view provides no set limit on what a ruler can declare to be objectively good. Although no perfectionist author argues that the government has the unlimited right to coerce its subjects to act in any way that it deems best for them, it is easy to imagine dangerous scenarios in which a ruler could oppress his people simply because he claims to have their best interests in mind. In light of these worries, most scholars of paternalism do not defend ends paternalism. They are explicit in endorsing *only* means-related paternalism—a variety that is made possible by the behavioral economics findings to be discussed in this chapter (Conly 88-89, Le Grand and New 3, Sunstein 19-20). This is why, in their popular defense of libertarian paternalism, Thaler and Sunstein repeatedly stress that they defend paternalism that aims to make its targets better off, "*as judged by themselves*" (*Nudge* 5, their emphasis). This qualification is their way of stating their deliberate support of only means-related policies.

As I have said, the accumulation of evidence against the traditional conception of human decision-making has created an opening for a different type of paternalistic intervention—one that recognizes systematic mistakes in the means we use to reach our own ends and then interferes to change those behaviors. This variety is fittingly referred to as means paternalism and is often viewed as being less problematic than ends paternalism because it leaves targeted individuals to act in accordance with their own conceptions of what is good for them, rather than

with those of the paternalist. Why exactly is means paternalism less threatening to our ability to self-govern? The answer is that the action it leads us to perform appears very similar to other actions that we perform autonomously.

According to a prominent understanding of autonomy—a set of views labeled "coherentist," an agent acts autonomously when the motive that guides her behavior coheres with her true values. For example, if I decide to perform an action because of motive X, then I am autonomous with respect to my decision if and only if X is aligned with one or more of my higher-order desires regarding the action, my long-term plans or goals, my set of stable preferences, or my overall character traits, depending on which version of the coherentist conception one ascribes to (Buss sec. 2). This view fits well with some of our intuitions about autonomous agents: we can say that a recovering drug addict is not autonomous with respect to her decision to take drugs if she does not want to desire the drugs but that someone who risks his life by riding a motorcycle is self-governing because he has higher-order desires or life plans not only to live safely but also to enjoy the thrills of life to an extent. On this view, means paternalism does not seem to violate autonomy at all. When a target performs a particular action because a means-paternalistic policy leads her to, her motivation for acting must align with her ends, in virtue of the policy being means-related. Thus, the target's policy-influenced action is done autonomously, so means paternalism avoids the autonomy objection on its own.

If this is taken to be true, then both libertarian and restrictive varieties of means paternalism avoid autonomy in virtue of being means-related. Indeed, scholars have not only asserted that means paternalism avoids the traditional autonomy objection, but have also claimed that it is not paternalism at all. This response often comes from anti-paternalists, who hold that the permissibility of means paternalism should not be cited as evidence that paternalism is not

fundamentally problematic. One philosopher who makes this case is Richard Arneson, who

defends John Stuart Mill's anti-paternalism. He cites the following example of means-related

paternalism and argues that it is permissible: under the condition that "every person in a society

prefers most of all not to be confronted with dueling situations," then "a legal ban against

dueling would be nonpaternalistic, since nobody's freedom is being restricted against his will"

(471). Furthermore, even if not *every* person in society has this desire, "if it is this pattern of

desires that generates reasons for forbidding dueling, then the antidueling law is non-

paternalistic" (471-472). However, he claims that this example should not be used as a

counterexample to Mill's anti-paternalistic view because it is not a case of paternalism at all.

Arneson's dismissal of this policy as non-paternalistic relies on his definition of paternalism,

which includes the qualification that the action must be "carried out either against his present

will or against his prior commitment" (471). Recall that I have taken on a broader definition,

which requires only that the individual's judgment is considered insufficient and is replaced by

that of the paternalist. On this definition of paternalism, Arneson's example is paternalistic; thus,

if it is not problematic, then there is at least one instance of paternalism that is not problematic.

Philosopher Sarah Conly makes reference to these arguments, asking toward the

beginning of her book if her (means-paternalistic) view is even controversial (Conly 42-43).

Though she welcomes the idea that her argument is uncontroversial, she cautions that it is not as

simple as this. The reason for her hesitation is that it is not simple to discover what people hold

as long-term ends or how they would act were they not suffering from reasoning failures. She

writes, "What the discussion of cognitive bias shows us is that the difference between an

informed and an uninformed person is complex," pointing to cases of smokers who are

somewhat aware of the risks of smoking but may not fully grasp how it will be to become

diseased as a result (44). This difficulty of distinguishing between people who are truly acting

irrationally and those who are simply acting on their short-term desires, which was discussed in

the previous chapter, is what leaves means paternalism open to normative objection. I take it that

Conly is correct in remaining skeptical about the claim that all cases of means paternalism avoid

the autonomy objection. One might reasonably argue that even if a policy brings someone to act

in a way that is similar to how they act in other settings, it may still violate her autonomy. This is

a position that I will develop further in Chapter 4 in discussing Feinberg's theory of soft

paternalism and Conly's hard-paternalistic view.

 The primary goal of this chapter is to describe the psychological findings that challenge

conventional utility theory, presenting evidence of our systematic reasoning failures and

emphasizing the aspects of this evidence that are most relevant to our decisions about food and

consumption. After doing so, I examine how proof of these irrational tendencies is supposed to

support the position that there can be truly means-related paternalism and how the specific

findings I have presented identify the three policies of focus in this project—menu labeling

regulations, optimal defaults, and sugary beverage taxes—as means-paternalistic. Ultimately, I

present two concerns about the success of this strategy for justifying paternalism. The first

targets the empirical possibility of making a group-level policy truly means-paternalistic toward

all of its targets, and the second addresses whether means paternalism alone can avoid the

autonomy objection. Responses to these concerns will be the subjects of Chapter 3 and Chapter

4, respectively.

<div align="center">---</div>

 There are many ways in which our decisions are systematically influenced by factors

other than our preferences and the options we are presented with. Here, I discuss the effects of

technical limitations, framing, relativity, and priming, limitations of the imagination, weakness of will, and social norms. Each bias that I consider plays a role in our decisions about food and consumption and prompts policymakers to employ the strategies of menu labeling regulations, optimal defaults, and sugary beverage taxes. Each has also been well established and has been replicated in numerous settings. These findings are highly interconnected and constantly evolving, as the fields they originate in are relatively new. In many cases, I cite classic studies that were the first to report these tendencies. However, I also incorporate findings from the most prominent behavioral scientists of today, with a special focus on those that are generated in food psychology labs.

Technical Limitations

First of all, we experience limitations in our ability to understand information, especially when it involves probability and logic. It is conceivable that these difficulties lead us to make choices that are not in our best interests. For instance, we exhibit "gambler's fallacy," which is the tendency to think that some independent event is more likely to happen at the next opportunity if it has just occurred several times in a row (Le Grand and New 85). We also exhibit "base-rate neglect," which is the habit of ignoring underlying data that affect the likelihood of certain events. For example, if study participants are given the information that a particular sample of people is made up of 30 percent lawyers and 70 percent engineers and are then given a neutral description of an individual from the sample who is "successful and well-liked in his career," participants will estimate that the probability of him being an engineer is equal to that of him being a lawyer—50 percent (87). Thus, they ignore the underlying statistic that indicates a higher likelihood of him being an engineer. These issues with probability are most worrisome

when they manifest themselves as inaccurate predictions of risk and consequences, especially with respect to chronic disease.

Even in situations that we would normally evaluate correctly, we are prone to using the "availability heuristic," or the mental process by which we evaluate choices or concepts based on the information we can most easily call to mind, rather than the relevant facts. For example, many people who understand the rule of logic that governs conjunctions—that the likelihood of A *and* B occurring will always be lower than the likelihood of A occurring alone or of B occurring alone—fail to apply it correctly when the logical answer does not match their stereotypes. To demonstrate this difficulty in a study setting, Tversky and Kahneman gave participants this description: "Linda is 31 years old, single, outspoken, and very bright. She majored in philosophy. As a student, she was deeply concerned with issues of discrimination and social justice, and also participated in anti-nuclear demonstrations" (Le Grand and New 85). When participants were asked whether it was more likely that Linda is now "active in the feminist movement," "a bank teller," or "a bank teller and active in the feminist movement," they consistently ranked the third statement (the conjunction) as more likely than either of the first options (its components). Groups of varied education level made the same error, even if they showed that they were able to follow the conjunction rule perfectly well when it was presented with letters or numbers (Le Grand and New 86). Again, this heuristic may be helpful when we are required to make a quick judgment, but it can also interfere with our ability to reach our desired results.

Another factor that can override our understanding of statistical evidence is overconfidence or over-optimism. We exhibit this tendency in many contexts: we overestimate the salary we will receive in the future, underestimate the amount of time an assignment will

take, and underestimate the chances that we will get a divorce (Thaler and Sunstein 32).

Overconfidence can have serious effects in many cases involving risky behavior and is a primary

reason for which we do not give risk its full weight in cost-benefit analyses. Even if we

understand probability, we may still underestimate the chances of getting a chronic disease as a

result of smoking cigarettes, visiting tanning beds, or overeating.

Salience and Framing

In addition to being influenced by technical limitations, we are swayed by the

prominence of choices or information, as well as by their framing. Salience, or the level to which

something is noticeable, affects our likelihood of reacting to or selecting it. We can increase the

salience of information by making it larger, brighter, or more colorful, and we can increase the

salience of objects by ordering them first in a sequence, placing them at eye-level, or locating

them in unavoidable spaces. We will consume more of a food if it is more salient, regardless of

how hungry we are or how much we enjoy eating it. For instance, the longer a meal lasts, the

more we will eat, and this is how factors such as the pleasant environment in a restaurant make

their impact on how much we consume. One study performed by Dr. Brian Wansink and the

Cornell University Food and Brand Lab demonstrated that people attending a conference filled

68 percent of their plates with the first three items presented in a breakfast buffet line, regardless

of whether these items were fruits or rich dishes (Wansink *Mindless Eating* 106). Another of his

studies established what he calls the "'See-Food' Trap," or the tendency to eat more of

something when it is visible. By giving two groups of secretaries containers of Hershey's Kisses

for their desks that were either clear or solid white and then observing how much they consumed

over the course of several weeks, he demonstrated that those who constantly saw the chocolates

reached into their dishes 71 percent more often than those who saw solid white containers (78-79). Wansink also showed that salience can work in our favor: when we are visually reminded of how much we have eaten, for instance with a pile of bones left over from our meal of chicken wings, we will consume less than if the evidence is out of sight (37-40).

A second important factor in our failure to remain objective when considering our options is how they are framed. One way in which choices can be presented is with strategically chosen words and phrases. For instance, we are more likely to undergo a medical procedure when we are told that it results in 90 percent of patients remaining *alive* for five years than when we are told that it results in 10 percent of patients *dying* within five years (Le Grand and New 89). We also overreact to offers that include the word "free." Dan Ariely, Professor of Psychology and Behavioral Economics at Duke University, has performed numerous studies on this "irrationality of zero cost." For instance, he and his colleagues set up a table in a large public space to sell chocolates, offering one chocolate per customer and a choice between a Lindt truffle for 15 cents and a Hershey's Kiss for one cent. As anyone who is familiar with these chocolates might predict, upon approaching the table and seeing the prices, most people purchased Lindt Truffles: about 73 percent chose truffles, and about 27 percent chose Kisses. Next, the researchers changed the prices, offering truffles for 14 cents each and Kisses for free. Now, 31 percent chose truffles, and 69 percent chose Kisses. Even though the price difference between the choices remained constant, people overwhelmingly took the free option, neglecting to perform the cost-benefit analysis that they were evidently completing in the prior situation. Additional experiments confirmed this finding, showing that discounting the prices from 27 cents for truffles and two cents for Kisses to 26 cents for truffles and one cent for Kisses did not

result in *any* change in the proportion of buyers purchasing Kisses, while reducing the price just

one step further to 25 cents for truffles and zero cents for Kisses did (Ariely 59-60).

We overreact not only to the absence of monetary costs but also to the absence of time

costs. We may spend a huge amount of time waiting in line to receive something that we do not

value very much, simply because it is free (Ariely 67). The power of zero cost is common

knowledge to marketing experts, who add free promotional items to otherwise unpopular

products to increase sales, as well as to food manufacturers, who create zero-calorie, fat-free, and

gluten-free products to capitalize on the effect of zero-guilt (68). As Ariely suggests, one reason

for which we overreact to the word "free" is that we are loss-averse, and zero cost eliminates loss

from the decision-making process. Indeed, our loss aversion and resulting tendency to overvalue

items we own is so great that we are consistently willing to pay far less to obtain something than

we are willing to accept to give it up (Thaler and Sunstein *Nudge* 33-34).

This "endowment effect" applies not only to objects but also to opinions, states of being,

and opportunities (Halpern 140). For instance, it might contribute to us valuing something we

have, such as freedom or pleasure, over something we could gain from making a choice, such as

the safety that results from wearing a seatbelt (Le Grand and New 90). Similarly, we are willing

to explicitly give up utility in order to avoid losing the freedom to change our minds, even when

we know with certainty that changing our minds will never yield more utility. To illustrate this,

Ariely and his colleagues showed study participants a screen with three doors on it and allowed

them to click a mouse up to 100 times within the course of a game. Clicking on a door led the

participant to a virtual room with a designated monetary range (e.g. 10 to 20 cents), and each

time she clicked inside the room, she won an amount of money within that range. In the first

version of this study, the participants could switch rooms as many times as they wanted to, and

all rooms remained active no matter what. All participants succeeded in maximizing their winnings by switching rooms until they found the room with the highest monetary range and then staying there to spend their remaining clicks (Ariely 188-189). In the next version of the game, however, any room that remained unclicked for 12 clicks in a row closed permanently. Although it was still in the participants' best interests to find and stay in the room with the highest range, they chose to sacrifice winnings in order to keep all the doors open. On average, they earned less money than those who had participated in the first version. Even when participants were given hundreds of practice rounds or were explicitly told the monetary amounts that awaited them within the virtual rooms, they gave up money in order to keep doors open (190-191). Not surprisingly, this type of extreme loss aversion, as well as the resulting tendency to overvalue what we have, has been manipulated by the private sector. As Ariely points out, promotions such as free trials and 30-day money-back guarantees lead us to take on ownership of products before we agree to purchase them, effortlessly raising the level to which we value them without any rational basis (176-177).

A last and often-overlooked feature of framing is the degree to which decisions are made fun, easy, attractive, and personal. In designing policies that incorporate evidence from cognitive psychology, behavioral scientist David Halpern and his colleagues at the UK government's Behavioral Insights Team (or "Nudge Unit") emphasize the importance of making desired behaviors Easy, Attractive, Social, and Timely (EAST). The attractiveness component of this strategy involves projects such as adding names to letters, which by itself can increase response rate by up to 20 percent (Halpern 103). Similarly researchers in the Netherlands showed that cutting whole-grain bread into fun shapes doubled its consumption by children (101). Using the same knowledge in the opposite way, many organizations have harnessed the power of shock

and disgust in various public service campaigns to dissuade people from smoking, encourage

them to wash their hands, and persuade them to wear seatbelts (98-99). Again, none of these

strategies would exert influence if we only made decisions on the basis of objective reasoning

and utility-maximization.

Relativity and Anchoring

Another factor that greatly affects how we evaluate our options is relativity. The classic

example of the influence of relativity on perception is demonstrated by the picture below, in

which the same solid black circle appears larger on the left, where it is surrounded by smaller

circles, than it does on the right, where it is surrounded by larger circles (Ariely 7).
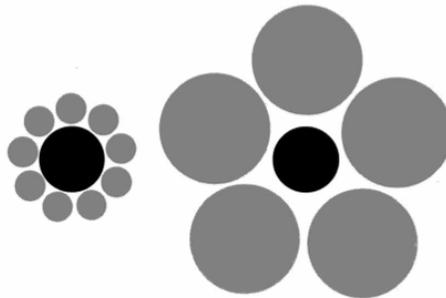
The same effect is seen with respect to food and drink—we make judgments about how

much we are consuming based on background objects such as plates, bowls, glasses, and

packages. People will predict that the same hamburger contains 18 percent more calories when it

is served on a small plate than when it is served on a large plate (Wansink 68). In one

experiment, Wansink and his colleagues held an ice cream social for Ph.D. students where they

allowed guests to serve themselves and secretly varied the sizes of the available bowls and ice

cream scoops. The findings were dramatic: those who were given 34-ounce bowls and large

scoops served themselves 57 percent more ice cream by weight than those given 17-ounce bowls

and small scoops (67). It has also been shown that we struggle to judge the volume of liquid in drinking glasses. On average, we unknowingly drink 25 to 30 percent more when given a short, wide glass than a tall, skinny glass, and this "horizontal-vertical illusion" affects even those with the most experienced pouring liquid—bartenders (64).

When it comes to packages, we eat 20 to 25 percent more from larger packages of breakfast, lunch, and dinner foods than we do from smaller packages. The effect is even more severe in the case of snack food: in one experiment, in which adults who were watching a film at a PTA meeting were given either a one-pound or half-pound bag of M&M's to snack on, those with the larger bag ate almost twice as many candies—or about 264 calories more (Wansink 59). The influence of relativity is so strong that even those who are explicitly taught about its effects exhibit the same tendencies demonstrated here just weeks after their lesson (68). In each of these randomized studies, participants who end up consuming more do not do so because of reasoned decisions about how the foods and drinks will contribute to their utility. Instead, their consumption choices are skewed in a predictable way by comparisons with other objects. The power of relativity in the context of consumption is particularly important when we consider that it has been predicted that 72 percent of the calories we consume come from foods we eat out of plates, bowls, and glasses (61), and we tend to consume about 92 percent of what we serve ourselves (59).

Knowledge of our tendency to assign value to things based on the features of whatever surrounds them rather than on their objective characteristics informs the pricing of goods. Although the standard economic model tells us that we will choose to purchase something based on its cost and the amount of utility it will bring us, we actually value a single item with a single price differently based on additions to the set of options that surrounds it. Restaurant-owners,

knowing that we are not likely to choose an item if it is the most expensive one available, often maximize their profit by making the second-most-expensive option on the menu whatever they hope to sell the most of (Ariely 4). In fact, marketers know that simply having *something* for customers to compare a product to increases sales, especially when we are unfamiliar with the items. In a successful demonstration of this phenomenon, Williams-Sonoma once introduced a larger, more expensive bread-making machine for the sole purpose of increasing sales—not of the new product, but of the original, smaller one (14-15).

This strategy also works when we already have multiple options. Marketers can shift our demand toward a product by adding an extra option that is very similar to one of the existing ones but is clearly less desirable. For example, study participants choosing between two journal subscriptions to the *Economist*—an internet-only subscription for $59 and a print-and-internet subscription for $125—chose according to their preferences: 68 people chose the first option, and 32 chose the second. However, when an additional option of a print-only subscription for $125 was added to the list, only 16 students chose the internet-only option, while the remaining 84 chose the print-and-internet subscription. By adding a choice that would clearly attract no demand but that was easily comparable to an existing option and clearly less desirable than it, the researchers shifted demand toward the more expensive product. Ariely and his colleagues show that this "decoy effect" has ramifications for dating, as well: when students are shown headshots of two attractive members of the opposite sex, their responses about who they would prefer to date vary. However, when an additional photo is added—a computer-generated version of one of the people that is made to be far less attractive than its original—results are skewed toward the original photo of that person. In other words, the addition of a decoy that no one is expected to select shifts choices toward the option that can easily be compared to its inferior version.

Finally, the way in which we assign value to items is not only influenced by surrounding options but is also affected by arbitrary numbers. This phenomenon is called "anchoring" (Thaler and Sunstein *Nudge* 23) or "anchor and adjustment bias" (Le Grand and New 88). For example, when people are asked to predict whether an apple contains more or fewer than 50 calories, they answer that it contains more and estimate the value at about 66 calories. However, when they are asked to predict whether an apple contains more or fewer than 150 calories, they estimate the value at about 114 calories (Wansink 24). Participants in a study who were asked to write down the last two digits of their social security numbers and then to write whether or not they would pay that amount of money in dollars for various items, the amount they were willing to pay for the items varied widely. Furthermore, once they had agreed to these prices as acceptable values for the items, different anchors in additional experiments did not override the effects of the first anchor—the social security number (Ariely 29-30). This tendency goes entirely against traditional utility theory, which indicates that we will decide whether to purchase something based on its objective qualities and the amount of utility it will bring us.

Priming and Expectation Assimilation

Priming refers to an effect of our memory in which one occurrence changes our behavior or perception with respect to a subsequent occurrence. Priming people with certain words can cause them to act in a way that corresponds with those concepts without them even noticing. In one experiment, researchers gave participants a task that involved unscrambling sentences. One group was given sentence-scrambles involving words that related to rudeness, such as "aggressive," "annoying," and "intrude," while a second group was given sentence-scrambles involving words that had to do with politeness, such as "honor," "considerate," and "sensitive."

When the participants were sent into another laboratory to wait to perform a second task, they encountered a staged scene in which a researcher attempted to explain the sentence-unscrambling task to a struggling participant. The group that had been primed with the concept of politeness waited 9.3 minutes on average before interrupting, and the group that had been primed with the concept of rudeness waited for only 5.5 minutes (Ariely 213-214). In a similar study, participants that were primed with words having to do with the elderly took much longer to walk down a hallway to a second experimental room than participants that had not been purposefully primed (214-215). The power of this type of priming is often used to encourage desired behaviors. For example, those at the Nudge Unit employ the cost-effective strategy of posing well-timed questions that prompt targets to plan out *when* and *how* they will perform a particular behavior such as voting (Halpern 144-145). In a study of 40,000 Americans, simply asking people whether they intended to buy a new car within the next six months increased purchase rates by 35 percent (Thaler and Sunstein *Nudge* 71). This phenomenon, called "implementation intention," has also been shown to occur with respect to health-related behaviors such as visiting the doctor, flossing, and eating fewer fatty foods (71-72).

A second, even more powerful effect of priming is that it leads us to perceive subsequent events in predictable ways. In other words, our perceptions conform to whatever we have been primed to think and feel. This "expectation assimilation" is embraced by restaurant owners, designers, and marketing experts alike. Wansink and his colleagues demonstrated its influence at a fine dining lab by serving a complimentary glass of wine to restaurant-goers from a labeled bottle that read either "NEW from California" or "NEW from North Dakota." Experiencing identical (cheap) wine, those who had been primed with the concept of a California vineyard rated the wine as more tasty, rated their meals as more tasty, ate 11 percent more of their food on

average, *and* stayed at their tables for 10 minutes longer than those who had been primed with

the thought of wine from North Dakota (Wansink 20-22). Simply expecting the wine to taste

better influenced their perceptions not only of the wine but also of the entirety of their

experience.

The way in which food appears also affects how we perceive it: in another experiment,

Wansink and his team had a cafeteria serve free brownies to its customers under the false

pretenses that they were testing out a new recipe. Those who were given their free brownie on a

napkin rated it as "okay, but nothing special" and claimed they would pay 53 cents for it; those

who were served their brownie on a paper plate rated it as "good" and offered to pay 76 cents;

those who received their brownie on a white piece of china rated it as "excellent" and were

willing to pay $1.27 (123). Even the way in which food is described primes customers to

perceive it differently. In another cafeteria experiment, researchers found that foods described

with more adjectives were purchased 27 percent more often, were consistently rated as being of

higher value, were ranked as tastier, and even led to more favorable attitudes about the cafeterias

in which they were sold (125-127).

Priming people so that they will perceive things in a particular way is clearly powerful,

but it must be done at the proper moment—before an experience occurs. If these tactics are

applied after we have formed our own preferences and opinions on the basis of what we truly

perceive, they have no effect (Ariely 206). Regardless of how our beliefs about certain

experiences and products arise, they tend to endure once they have been formed. This tendency,

called "cognitive inertia" or "confirmatory bias," can even lead us to reinterpret new information

in a ways that reinforce our existing beliefs (Halpern 269). Thus, it is no surprise that placing

attractive labels on items, associating products with attractive people and concepts, and creating

appealing store environments are such effective methods of increasing brand loyalty and maximizing sales.

Limitations of Imagination

Aside from having trouble processing information and forming objective opinions in the face of context and priming, we also have limited imaginations, which makes it challenging to predict what we will desire in the future and thus more difficult for us to act on future-oriented goals and to try new things. When we have a very low level of experience with a given decision and its consequences, or "thinness of experience," this results in various types of "projection biases," which lead us to favor choices that have easily-imagined outcomes over those that lead to less predictable consequences (Le Grand and New 91-92). This can manifest itself in two ways: procrastination and "status quo bias." Both varieties of projection bias are only made worse by our imperfect memory. As both Halpern and Wansink point out, we experience strong "primacy and recency effects," tending to have a stronger memory of the peak and end of an experience and a weaker memory of its low point and middle (Wansink 156-157, Halpern 230). We also tend to ignore the durations of the peak and end points, especially when we think about an event in hindsight (Le Grand and New 93). The less we are able to remember how an event makes us feel, the worse we are at predicting how a similar one will affect us in the future.

In the case of procrastination, we face a choice between accepting some negative consequence now for the sake of some positive result later, and accepting a positive result now to the detriment of our future selves. Since we are closer in time to the immediate consequences and thus can imagine them more easily (Le Grand and New 94-95), they hold more weight in our decision-making process (Thaler and Sunstein *Nudge* 40-42). Studies of time discounting, or the

rate at which things become less valuable as they occur further in the future, show that people prefer to receive a smaller amount of money now than to receive a larger amount in one month's time (Samson 3). If we were Econs, of course, we would value something at the same level now as in the future, adjusted only slightly for inflation. The most effective way to reduce the regrettable consequences of procrastination is to impose strict deadlines. To show this, Ariely used different deadline-setting strategies in three of his classes and examined the effect on the students' grades. One class had no deadlines but simply had to turn in all their papers before the end of the semester; another class could set deadlines individually but would all experience penalties if they turned in their papers past their self-set deadlines. The third class had traditional deadlines throughout the semester. As one might predict, those in the third class performed best on their assignments, while those in the second class performed somewhere in the middle, and students in the first class performed the worst (Ariely 144-145). Though informative, this result is not promising for those who oppose paternalistic intervention in all settings.

A second important variation of projection bias is what authors call "status quo bias" (Le Grand and New 94). This predisposition to sticking with the situation that we are already in or the choice that we have previously made is related to cognitive inertia and confirmatory bias, mentioned in the previous section, but applies to options and choices rather than to beliefs. Status quo bias is what keeps us sitting in the same seat during class week after week, forgetting to cancel our magazine subscriptions, and leaving the TV on the same channel during commercial breaks (Thaler and Sunstein *Nudge* 34-35). It is also a major reason for which most people accept the default retirement plan and never switch (Le Grand and New 97). Because we have trouble imagining how a different situation will affect us and because we are so attached to the things we already have, we tend to resist change. Status quo bias has great influence even when

we are not responsible for choosing the initial state of being, which is what allows others to manipulate our decisions so effectively just by carefully constructing default options (Thaler and Sunstein *Nudge* 35).

Weakness of Will

Weakness of will, also described as "temptation" by Thaler and Sunstein, exerts its influence when we know what we want in the long run but act instead on conflicting, short-term desires. It can be classified as a choice failure, rather than a reasoning failure, because it interferes after we have successfully determined what choice is best for us. This understanding of limited willpower has been central to descriptions of human behavior for thousands of years: Aristotle called it *akrasia*, or lack of self-control, and described it as the phenomenon of an individual knowing what the best judgment is and yet acting against that judgment (Le Grand and New 97). It is not difficult to identify instances of limited willpower, and Thaler and Sunstein illustrate this tendency with a lighthearted example in which guests at a dinner party are offered a bowl of nuts. When asked at a separate time how many pre-dinner nuts they would prefer to consume in such a circumstance, most prefer to eat just a few and then stop. However, when the bowl is left on the table all night, guests continued to consume nuts, giving in to temptation and also to mindlessness, or Wansink's "See-Food Trap" (Thaler and Sunstein *Nudge* 41-42).

A psychological explanation for limited willpower put forth by Loewenstein is that we experience a "hot-cold empathy gap," meaning that when we are outside of a situation, we underestimate how strong our desires will be within that situation (Loewenstein). For the most part, psychological evidence pertaining to weakness of will is somewhat deficient in comparison

to that which establishes other irrational tendencies. Le Grand and New attribute this in part to an ongoing debate about whether actions resulting from limited willpower should truly be classified as irrational. Some maintain that favoring our first-order preferences over our second-order preferences, as we do when we exhibit failures of willpower, is actually rational. In addition, although Aristotle said that people who exhibit weakness of will do not reject the value of virtue and are thus not "evil," this particular choice failure evokes a sense of moral evaluation (Le Grand and New 98). We tend to think of strong willpower as a trait that implies good character. While the reasoning failures, and to some extent the social choice failure introduced in the next section, tend to be linked with innocent mindlessness or condition of being busy, limited willpower holds more moral weight. It is for these reasons that I do not dwell on weakness of will as a primary example of our irrationality, as I do with regard to other biases.

Social Norms

A last irrational tendency that we exhibit is to be heavily influenced by social norms. Social norms come in two varieties: descriptive norms indicate how the majority of people in fact *do* act, while injunctive norms indicate what behavior is socially acceptable, or how we *should* act. The classic example of the power of descriptive social norms is that we give incorrect answers to tests of perception when people around us answer incorrectly, even though we easily perform well on the same tests when we are alone (Thaler and Sunstein *Nudge* 56-57). For instance, when participants in a study were asked to match lines drawn on paper with lines of the same length on a screen, they almost never erred when performing the task alone but erred over one-third of the time when asked to complete it in a group that was giving incorrect answers (56). The more difficult a task becomes, the more likely we are to conform to whatever answer is

given around us (57-58), and discussed earlier, incorrect beliefs tend to endure despite new evidence. While Econs' decisions would never shift for such an irrational reason as the fact that "everyone else is doing it," humans are very susceptible to this factor (54).

Injunctive norms also influence our behavior in powerful ways. One reason for which we care so much about what we *should* be doing is that we experience "spotlight effect"—the tendency to think that others notice more about our behavior than they actually do (Thaler and Sunstein *Nudge* 60-61). For example, students in one study predicted that 46 percent of people they encountered in a room would remember what was pictured on their t-shirt, while in reality only 21 percent had noticed (61). As a result of this bias, we are far more likely to donate to a cause or participate in a campaign if our actions are displayed to the public in some way, whether this is through a sticker or pin advertising our contribution or through the publication of participants' names.

When injunctive and descriptive norms come into conflict, descriptive norms tend to dictate our behavior (Halpern 108). This is the main reasoning underlying the shift in strategy that has been seen for campaigns aiming to reduce alcohol abuse on college campuses. When these campaigns focused on illuminating the dangers of excessive alcohol consumption and underage drinking, they were criticized for contributing to rising rates of alcohol abuse by overemphasizing the number of people engaging in risky drinking and thus normalizing the behavior. In more recent years, campaigns have focused on raising awareness of how rare alcohol abuse is and highlighting how many students refrain from drinking entirely before they are of legal age (Blanton et al.). With similar reasoning, Robert Cialdini, author of *Influence: The Psychology of Persuasion*, calls attention to what he names the "big mistake," or the

overemphasis that policymakers put on how many people are doing something wrong, rather than how many are doing it right or what they should do.

It is rarely this clear which type of social norms are exerting an effect on us. Together, descriptive and injunctive norms are powerful in the context of consumption. As researchers have shown, people who normally eat lightly tend to eat more in large groups, while people who are generally heavy eaters eat less around others. On average, we eat about 35 percent more when we eat with one other person than when we eat alone. We eat about 75 percent more when we eat in a group of four, and 96 percent more in a group of eight or more (DeCastro). This contributes to what some call the contagiousness of obesity and the finding that our weights are significantly correlated with those of our closest family members and friends.

The findings described in this chapter establish that we are not perfectly rational beings and do not always choose to do what is best for ourselves. Thus, as explained earlier, they provide an opening for a less objectionable variety of paternalism that interferes with our behavior to help us reach our *own* goals. The case has been made that menu labeling regulations, optimal defaults, sugary beverage taxes, and countless other policies like them are means-paternalistic in this way. Indeed, it is plausible that when we order high-calorie, low-nutrient foods at quick-service restaurants and purchase sugary beverages, we do so not *only* because they taste great and have little economic cost, but also because we suffer from many of the reasoning failures presented here. For one thing, unhealthy food is prominently listed and pictured on menu boards and wall advertisements, and soda is displayed at cash registers and is placed at eye level on shelves to increase convenience and salience. Advertisements for these products frame them with appealing words and even offer free sides, free toys, free entries into

raffles, and discounted combo-meal options to capitalize on the power of zero cost. Attractive associations with famous celebrities, child-friendly characters, and beautiful actors and actresses are common in all forms of marketing, and the Coca-Cola company goes so far as to personalize their bottles and cans with common first names (Matheson).

Behavioral insights are even adopted into pricing strategies. Instead of being priced per unit, items are typically priced so that the middle option in the set—the one most people tend toward—is extra expensive, pushing consumers toward the largest available size. After all, it only costs a few cents more! By utilizing anchoring and adjustment bias, taking into account the decoy effect, considering priming effects and expectation assimilation, and taking advantage of status quo bias to encourage habit formation, companies can raise profits. Meanwhile, anyone who has ever vowed to stop drinking soda, cut down on fast food, or consume fewer calories, has likely found themselves going along with the food choices of the people around him or has suffered weakness of will during the moment of truth. Insofar as menu labeling regulations, optimal defaults, and sugary beverage taxes take into account these reasoning failures and aim to change our behaviors so that we will make the choices we would have made if we had been acting rationally, they may be means-paternalistic.

Although means-paternalistic policies are an attractive option for promoting people's wellbeing without infringing on their autonomy, there are two concerns with this strategy. First of all, it is impossible to know for sure that a population-level policy is means-paternalistic with respect to each of its targets. To show that any practice is means paternalistic requires more than just evidence of our reasoning failures. On an individual level, we must know three things: (1) I make systematic mistakes in my reasoning; (2) these mistakes often cause me to fail to reach my own ends, and (3) you can interfere to help me make the decision that I would have formed if I

had not been suffering from reasoning failures. Another way of formulating (3) is that you can interfere to make me do something that better coheres with my overall life goals. The evidence presented in this chapter has supported premise (1), and premise (2) is supported by all the cases in which people regret their own decisions—whether or not they can point to the specific irrational tendencies that were the cause of their troubles. Premise (3) is more difficult to defend. It is plausible that this can be done between individuals: if you were to sit down with me and understand my true ends and the tendencies that get in my way of achieving them, you could intervene with my behavior to help me reach my goals.

To be successful, the process would be complicated. Not only would you have to weigh the benefits of your interventions against costs such as my diminished sense of privacy, loss of self-esteem, or feelings of being excessively punished, but you would also have to check in at intervals to ensure that my ends had not changed. The goal of sorting out my true ends from those preferences and desires that are influenced by my irrationality would be difficult, if not impossible, since we do not have a good understanding of the boundaries between rational and irrational, being fully informed and uninformed. If my ends can be identified, they must also be weighed, ordered, and qualified. How much do I value good health and long life, and how much do I enjoy indulgent food and long meals in the company of friends in a way that is independent of my tendencies to overeat and be influenced by social norms? This difficult and potentially impossible task may not be necessary if we think that it is enough to know vaguely what coheres with someone's values. However, many would argue that we simply cannot or do not know enough, which undermines even the claim that an individual-level act can be truly means-paternalistic. Even if this were possible, intensive individual-level paternalism is not feasible on a policy level and is certainly not what occurs when calorie labeling laws, optimal defaults, or

sugary beverage taxes are imposed. If we do help some people meet their ends through these policies, we do so with a degree of chance and cannot be sure that we are doing so at all. For a population, the government certainly does not know what every affected individual has as ends and thus whether an intervention coheres with his long-term plans, and without certainty about premise (3), no population-level regulation is guaranteed to be entirely means-paternalistic.

A second concern is that even if true means paternalism can be accomplished at any scale, some instances of it may still violate autonomy. I demonstrated earlier that means paternalism is meant to avoid the autonomy objection because the motive that the target acts on under a means-related policy *coheres* with his long-term plans and ends. However, it does not seem that this is enough to establish that his autonomy has not been violated. What seems to be more important is his status with respect to this decision prior to, or without this intervention. Another way to frame this concern is that while he may endorse his motive for acting under the means-related policy, he is not actually in control of his immediate motive for action. If, in the absence of the policy, he would have had control over his motive, then the policy seems to interfere with his self-governance. Rather than condemn means paternalism as a violation of autonomy, this objection simply requires instances of means paternalism to be defended in some further way, be that through their qualification as cases of soft paternalism, or on the basis of a hard-paternalistic defense. I return to this challenge in Chapter 4, but first, I discuss the first concern raised here, regarding the capacity for means paternalism to be truly means-related toward all of its targets. As I will show in Chapter 3, libertarian paternalism and the related concept of asymmetric paternalism serve as ways of addressing this problem.

Chapter 3
Why Nudge? The Advantage of Libertarian Paternalism

In 2003, behavioral economist Richard Thaler and law professor Cass Sunstein published an article in the *American Economic Review* in which they coined the term "libertarian paternalism" and encouraged their audience to reconsider the traditional view that all paternalism involves coercion. They defended a type of policy that not only guides people's behavior to promote their wellbeing, but also preserves their freedom of choice. In this way, these policies accomplish many of the benefits of more restrictive measures, such as taxes and regulations, but may be acceptable to the classic opponents of paternalism—libertarians. Anticipating the reaction that their newly proposed term was incongruous, Thaler and Sunstein published a longer version of their argument in the *University of Chicago Law Review* later that year titled, "Libertarian Paternalism is Not an Oxymoron," which was followed by several articles on the same topic in the years that followed. The authors' work prompted an impressive debate that continues into the present, not only among those in their fields of study, but also in the disciplines of philosophy, bioethics, and public health, as well as in the public sphere and popular media. In 2008, they published a book-length defense called *Nudge: Improving Decisions about Health, Wealth, and Happiness*, which introduced a friendlier term for libertarian paternalistic policies that would only further their recognition and appeal: nudges.

As explained in the previous chapter, Thaler and Sunstein emphasize throughout their work that they defend means paternalism and not ends paternalism. However, they do not promote all forms of means paternalism—they only accept those practices that preserve liberty by not limiting people's choices (*Nudge* 5). In a follow-up book that elaborates on the ethical objections to nudges, Sunstein reiterates that the primary advantage of libertarian paternalism is

that it avoids coercion (Sunstein 17-18). Despite these assertions, I claimed at the end of Chapter 2 that the subject of this chapter would be a response to an empirical concern, that group-level means paternalism is not guaranteed to be means-paternalistic toward all of its targets. How are these two proposed advantages—that nudges avoid coercion, and that they ensure a population-wide policy will be means-paternalistic toward all people—linked? The purpose of this chapter is first to describe the concept of nudges and then to reconcile these two ideas and explain what the true advantage of libertarian paternalism is meant to be. In doing this, I draw on the work of business economist Colin Camerer and his colleagues regarding a similar idea called "asymmetric paternalism." Then, I explore objections to these accounts and examine how nudges interact with autonomy, considering the philosophical literature around the moral status of manipulation. Lastly, I discuss two further worries about libertarian paternalism and one practical benefit, each with tangential connections to the traditional autonomy objection.

What Are Nudges, and When Are They Useful?

In contrast to taxes, incentives, and bans, which cut off people's available options by making the consequences of some choices too costly, nudges structure the decision-making environment in a way that encourages people to exhibit certain behaviors. To illustrate this, Thaler and Sunstein introduce the concept of "choice architecture," which consists of all the factors that influence our behavior, both consciously and subconsciously, including all those discussed in the previous chapter—salience, wording, framing, default options, etc. (*Nudge* 11-13). A nudge works by manipulating aspects of choice architecture so that people will behave in ways that are good for them without even noticing. In a classic example, an employer might nudge her employees to save for retirement by creating a plan with an opt-out format rather than

an opt-in structure. Employees will be automatically enrolled in a retirement plan, and although they can opt out of it at any time for no charge, most will stick with the default plan. Some of those who stay in the plan would have signed up for it even if there had been an opt-in structure, but others would have stuck with any default, due to the effects of status quo bias, loss aversion, or procrastination (106-109). Thus, on an opt-in structure, these people would have lost out on retirement savings. Because of the intervention, they benefit in the long run and meanwhile remain free to choose either option.

Nudges by definition are not necessarily implemented intentionally or established with the purpose of helping their targets. Technically, 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" (Thaler and Sunstein *Nudge* 6). Choice architecture exists whether it is explicitly designed or accidental, and people and entities that structure choices are "choice architects*"* whether they do so unintentionally or purposefully, with the public in mind or with profit in mind. Some of the earliest examples of choice architects that Thaler and Sunstein introduce—people who design voting ballots, doctors who describe potential medication plans, and parents who describe college options to their children—are benevolent and relatable, no doubt because these examples serve to support the authors' forthcoming argument that nudging people is justifiable, common, and even inevitable (3). However, nudges can in theory be ends-paternalistic if they are implemented to improve the objective welfare of their targets in a way that does not align with those targets' own values. Even worse, they can be employed to benefit the agent rather than the target, or even to hurt the target. Despite their broad definition, Thaler and Sunstein refer to nudges with a more specific idea in mind. They consider the consequences of both intentional and unintentional nudges throughout their work but

ultimately support the implementation of intentional nudges only insofar as they are instances of

libertarian means-paternalism: they must aim to improve the wellbeing of their targets, as seen

by those who are affected. Throughout this chapter, I refer to nudges with this understanding in

mind.

As is evident from the nature of the choice architects that the authors refer to, nudges

need not be instituted by a government or other large entity. Individuals and small groups can

nudge others, and the ethical considerations of doing so are examined elsewhere in the bioethics

literature with respect to clinical research and doctor-patient interactions. Of course, Thaler and

Sunstein focus on the use of paternalism mainly by the government and private sector, defending

both. They also promote third-party, or impure, paternalistic measures in which firms are either

nudged or forced to nudge their customers. This is a constant source of tension in the debate over

libertarian paternalism because although a few of their proposed regulations involve nudging

private companies to nudge customers, *many* more involve imposing taxes or bans on private

entities to coerce them into nudging the public. All three of the policies of focus in this project

are cases of impure paternalism, as they require the private sector to adopt policies that then

nudge consumers to make healthier consumption choices. This concern has spread from

scholarly debates into the public sphere, with news sources expressing worry that although the

government may nudge people, it *shoves* industry in order to do so (Lobstein). I will return to

this subject later in this chapter and in Chapter 4, as it relates to the question of which costs

should be considered in evaluating paternalistic policies.

Thaler and Sunstein suggest that libertarian paternalism is most useful in situations that

are not conducive to trial-and-error learning and to solve problems that are not remedied by the

free market. When many of the reasoning failures discussed in Chapter 2 are in play at once,

such as when decision-makers face difficult technical choices, receive little or no immediate feedback about the consequences of their choices, encounter scenarios that they meet infrequently, or face decisions where the wrong choice leads to particularly severe consequences, nudges should be considered. For instance, decisions involving mortgages, health plans, or investment packages benefit more from nudges than those involving less technical knowledge, and decisions that result in immediate positive feedback but considerably delayed negative feedback, such as behaviors that lead to chronic disease or serious personal debt, are more appropriate for nudge interventions than those that provide immediate response (*Nudge* 75-75). Though the authors use the pervasiveness of difficult decision-making contexts specifically as evidence of the need for nudges, this argument is not exclusive to nudging—it can be used to support the implementation of any government regulation that targets behavior.

The Advantage of Nudging

Thaler and Sunstein promote libertarian means paternalism as a special variety of means paternalism that avoids coercion. Coercion, though a contested idea itself in the discipline of moral philosophy, is typically defined as the practice of compelling a person to do something that he is unwilling to do and is understood to involve a violation of autonomy (Anderson). Here, I refer to practices as coercive when they force their targets to behave in a certain way *and* infringe on their autonomy in doing so, while I refer to practices as restrictive when they do so without violating autonomy. Ignoring the second objection raised at the end of Chapter 2 and assuming for the purposes of this section that means paternalism avoids violating autonomy, then all forms of means paternalism—even taxes and bans—should be non-coercive. If this is the case, then what is the advantage of libertarian means paternalism?

To reiterate, a practice is means-paternalistic when it results in its target acting as he would have if he had not been suffering from reasoning failures, social influences, and weakness of will. I argued that on an individual level, it is possible to know that your intervention does not impose values on someone and thus affects him only means-paternalistically, but only through intensive conversation and policy revision. On a group level, it is possible to affect some or most people means-paternalistically if a policy promotes commonly held ends, but some people within the affected group may nonetheless be the targets of ends paternalism. Due to serious practical constraints, it cannot be known which or exactly how many members are affected by means-related versus ends-related paternalism when the policy is established. All that has been said thus far is true of both taxes and bans and libertarian forms of means-related and ends-related paternalism. Now, let us suppose that the group-level policy in question is a means-related tax or ban. In this case, all those under the scope of the policy will be threatened into changing their behavior. As a result, many will change their actions such that their own goals are met, so they will be the targets of means paternalism and will not be coerced. Others will have values forced on them, so they will be the targets of ends paternalism, which is always coercive. An infringement of autonomy only occurs with respect to the latter group, whose long-terms goals are not met.

Suppose instead that the policy in question is libertarian means-paternalistic. If it exerts effects on everyone, it too will lead those who have been suffering from reasoning failures and choice failures to get closer to their own ends and will lead those who have already been choosing optimally for themselves to fulfill ends other than their own. However, because the policy is not restrictive, those who do not have the ends that it promotes can simply opt out and behave in whatever way serves their goals, at no cost. Thaler and Sunstein write that "to count as

a mere nudge, an intervention must be cheap and easy to avoid" (6). Though they do not specify a particular threshold to evaluate ease or low cost, their criterion seems to be a one-click rule. In some cases this is literal: when you send an email with the word "attachment" in it and do not attach a file, Google nudges you with a pop-up box asking if you forgot your attachment. If you were not suffering from forgetfulness and did not intend to attach a file, you can avoid this nudge with one click of a box. Even when a nudge cannot be avoided with a literal click of the mouse, the cognitive and physical effort required to avoid it should be comparable (Thaler and Sunstein 252).

In theory, *all* those who would have been the targets of ends paternalism under a restrictive policy will avoid a nudge, leaving only targets of means paternalism to be affected. As long as the cost of resisting is low, all targets who react to the nudge will be helped to reach their own ends. This makes libertarian paternalism an exception to the worry that population-level means paternalism cannot guarantee to be means-related toward all targets and does not even require that policymakers identify which people fall into which category. In this way, libertarian paternalism minimizes or eliminates all ends paternalism, and in doing so, minimizes or eliminates all coercion from the regulatory process. This is, of course, still assuming that means-related paternalism is non-coercive.

Asymmetric Paternalism

In the same year that Thaler and Sunstein published their initial defense of libertarian paternalism, economist Colin Camerer and his colleagues introduced the concept of "asymmetric paternalism" in an article titled, "Regulation for Conservatives." According to the authors, "a regulation is asymmetrically paternalistic if it creates large benefits for those who make errors,

while imposing little or no harm on those who are fully rational" (Camerer et al. 1212). Their concept, though closely related to that of a nudge, is a framework for evaluating the acceptability of all paternalistic policies, rather than a single warranted variety of paternalism. They refer to it as "a new metric for evaluating the costs and benefits of regulatory options" (1251), meant to not only respond to concerns from "those (particularly economists) prone to rigid antipaternalism" (1212), but also to give enthusiastic paternalists of all kinds a way in which to evaluate the advantages and risks of their policy plans.

Camerer and his colleagues propose an equation for assessing the main costs and benefits of regulations (1219). If a policy *is* asymmetrically paternalistic, then:

$$(p * B) - [(1 - p) * C] - I + \Delta\pi \ > \ 0$$

Here, p represents the number of people who, prior to implementation, suffer from reasoning failures that distance them from their goals but who, after implementation, behave in a way that brings about their ends (Camerer et al. 1212). The benefit of any regulation is the wellbeing (B) gained by these people, or in my terminology, the wellbeing gained by those who are the targets of means paternalism under the policy. Meanwhile, the cost of the policy is the wellbeing (C) lost by those remaining people $(1 - p)$, who already act in their own best interest prior to implementation and are then forced to behave in a way that opposes their own ends afterward—those who are the targets of ends paternalism under the policy. In addition to these factors, Camerer and his colleagues take into account the government's implementation costs (I) and the change in profit experienced by any firm that is responsible for enforcing the regulation $(\Delta\pi)$. The authors are optimistic about the effect on firms, writing that an act of paternalism can result in benefits for a firm if the original reasoning failures in play lead consumers to undervalue its products. However, considering the remarkable use of marketing strategies across industries,

it is more likely that paternalism will impose costs on firms. It is controversial whether such costs should be considered in the development of nudges and other paternalistic regulations. I argue that it is acceptable to impose costs on a firm by implementing a nudge when that firm is benefitting largely because of reasoning failures and choice failures to begin with. Overall, the benefits minus these costs must be greater than zero for the regulation to be asymmetrical.

The asymmetric paternalism model can be applied to any regulation, from an act of ends paternalism to a restrictive but overall beneficial policy, to a cautiously implemented nudge. Ends paternalism will clearly yield a negative result in Camerer's equation because it results in high losses of wellbeing for a large number of people. If it promotes an objective standard of welfare that is so disconnected from our general concept of what it means to live a good life that it matches no person's subjective idea of wellbeing, then it can even result in zero benefits, in addition to high costs for everyone. At the other extreme, what the authors call a case of "pure asymmetric paternalism" yields benefits for some or many, while producing no costs for any targets, no costs for the government, and no costs for industry (Camerer et al. 1219). In order to avoid confusion with the pure and impure varieties of paternalism introduced in Chapter 1, I refer to this type of regulation as *truly* means-paternalistic, toward all of its targets. The authors write that the most purely asymmetric policy possible is a default regulation, though even this may involve some implementation costs. Because it allows anyone who would normally be the target of ends paternalism to opt out and is inexpensive to implement, a default regulation is purely asymmetric, or truly means-paternalistic in my language.

Most nudges fall *close* to this end of the spectrum, which is why libertarian paternalism and asymmetric paternalism are thought to be so similar. However, nudges are not the only policies that are asymmetrically paternalistic. One obvious example is seatbelt laws, which

threaten targets into doing something very small that reduces the risk of death and serious injury in accidents. Most, if not all targets value this benefit of lengthened, healthy life over the pleasure and convenience of driving without a seatbelt, and it is reasonable to assume that nearly everyone driving without one is suffering from one or more reasoning failures, social influences, or weakness of will. A tiny number of people may be making this choice rationally, in which case they will suffer a loss of wellbeing when they are threatened into buckling up, but this restrictive policy is asymmetrically paternalistic overall. Seatbelt regulation and other comparable examples highlight that although nudges and asymmetric paternalism rest on the same rationale, the latter is a framework to be used for justifying or condemning any paternalistic policy, while the former is a specific type of justifiable paternalism. Another difference between the two approaches is that Camerer and his colleagues are more explicit in evaluating the net benefit of a policy, while Thaler and Sunstein are more focused on promoting policies as long as costs are minimal to those who wish to opt out of the nudge. It seems to be a responsibility of the government, and sometimes of private companies, to implement beneficial nudges, regardless of exactly how much benefit they bring to people. This is not to say that Thaler and Sunstein would support implementing all possible nudges at the expense of the public's tax dollars but is rather to say that their approach comes across as more idealistic.

Because the government cannot identify which or how many of its citizens are benefitted by a policy, there is no way to observe whether it is asymmetrically paternalistic when it is restrictive. Those who are coerced by the regulation may become resentful or may protest, but this is unlikely to be a reliable representation of how many people are affected in this way. However, in the case of non-restrictive regulations, both authors imply that such a calculation can be estimated well (Sunstein 154, Camerer et al. 1254). Any target who finds that a non-

restrictive regulation does not promote his ends stands to lose wellbeing by going along with it and so is presumed to avoid it. Meanwhile, anyone who finds that the regulation does promote his ends stands to gain wellbeing by going along with it and so is presumed to allow himself to be nudged. With this understanding, the policymaker can estimate the total benefit yielded by her policy ($p * B$) and can learn that her regulation aligns with most of her targets' ends. Is this a sound method of observing a regulation's success?

Concerns for Libertarian Paternalism

Earlier in this chapter, I emphasized that the benefit of libertarian paternalism—that it ensures a group-level means-paternalistic regulation is means-related toward all of its targets and thus ensures the absence of coercion—is an advantage only *in theory*. I add this qualification because the advantage of nudges relies on the assumption that they target *all* and *only* people who currently act irrationally and can be helped to reach their own ends by the intervention. This assumption is unrealistic because nudges, by their nature, are not universal in their effect. This leads to two concerns. The first is that nudges do not capture *all* of the people who could benefit from them, were they more restrictive. While taxes and bans force all targets to change their behavior, nudges only nudge a subset of the people who stand to gain wellbeing. One can argue that because those who stand to gain are those who currently suffer from reasoning failures and choice failures, they are generally more vulnerable to the influence of choice architecture and will therefore more be more susceptible to nudges. However, some targets will remain uninfluenced and will continue to fail to reach their own ends. In fact, those who are not affected are precisely those who need an intervention the most, since they are so entrenched in their reasoning failures and choice failures that only harsher restrictions can help them (Conly 31-32).

Not only does this yield lower net benefits, but it also leads to a skewed view of the policy because the people who remain uninfluenced should be counted as part of the population p but are instead counted in $(1 - p)$. Although this does not lead to a higher calculated cost because the cost to avoiding a nudge (C) is zero, it does lead to a lower estimated population of people who benefit from the regulation. This may leave policymakers with the mistaken idea that the behaviors that their policy promotes do not align with people's subjective views of wellbeing.

The second concern is that nudges may not *only* target people who stand to gain wellbeing. The people whom libertarian paternalists aim to avoid targeting are also vulnerable to the influences of choice architecture, so it is unrealistic to expect that everyone who stands to lose wellbeing by going along with a nudge will notice this and resist it in time. One might respond that since these are the very people who already avoid the influences of choice architecture to make decisions in their best interests, they must be especially rational people, able to resist added choice architecture just as well. Alternatively, one might argue that it is the responsibility of these people to avoid nudges or that it does not matter if they actually avoid them, as long as they have the freedom to. Finance mathematician Riccardo Rebonato objects to this last point in his scholarly work on nudges, arguing that libertarian paternalism does not preserve liberty at all because it only promotes nominal freedom of choice, not effective freedom of choice. In other words, if people do not *actually* exercise the freedom to opt out of nudges, then their freedom to do so must be flawed or diminished in some way (Rebonato). Each of these defenses should be considered. However, when one uses a policy strategy precisely *because* it maneuvers around people's reasoning capacities to influence them without their awareness, it seems unfair to base the advantage of the policy on the assumption that these people will always overcome your meddling to do whatever is in their best interests.

The first of these concerns is only a problem insofar as we are comparing the total benefits yielded by libertarian paternalism versus restrictive policies. If there are substantial ethical reasons to prefer nudges, then it is not disastrous that they yield a lower increase in wellbeing than taxes and bans do. Sunstein makes this point late in his book, while responding to potential objections (152-153). However, the second concern undermines the advantage that libertarian paternalism has over other varieties of means paternalism. Both nudges and restrictions may lead some small group of people to behave in a way that does not cohere with their own ends, so neither variety can guarantee a policy to be truly means-paternalistic on the population level. The only difference is that while restrictive varieties ran the risk of inadvertently *coercing* these people by way of ends paternalism, libertarian varieties run the risk of accidentally *manipulating* them in the same way. The question becomes whether this type of manipulation is more, less, or equally problematic as the coercion that nudges avoid.

Means-Related and Ends-Related Manipulation

Manipulation lies on a continuum of strategies for changing others' behaviors, somewhere between reasoning and persuasion on one end and coercion on the other. Different varieties of manipulation serve to influence people's decision-making through different pathways. Bioethicists Amulya Mandava and Joseph Millum divide cases into those that alter perceptions and those that alter choices. When someone deliberately deceives another about the value of a good she is selling to him or plays on the guilt of another to extrapolate resources from her, the target's perceptions of her options are altered. These are not cases of persuasion because although they involve an interaction with the deceived person's reasoning capacities, causing her to weigh the costs and benefits of her decision differently, they use false information to do so.

Alternatively, manipulation can be used to change a target's actual options, if deceptive measures are used to get her into a situation in which her choices are limited (40-41). Philosopher J. S. Blumenthal-Barby divides cases of manipulation differently, on the basis of whether they bypass the target's awareness or work explicitly and whether they exploit facts about his psychological state, reasoning capacities, or emotions or manipulate facts. The physician who shows a habitual smoker a deeply frightening video about the risks of tobacco, the company that displays its product in commercials that play on the nostalgia of those watching, and the friend who reminds another friend of all her past wrongdoings to encourage her to do her a favor all play on the emotions or tendencies of their targets explicitly. In contrast, nudges almost always bypass their target's awareness.

Traditionally, manipulation has been viewed as *pro tanto* wrong (Blumenthal-Barby 123). Indeed, Mandava, Millum, and numerous other authors still maintain that all forms of manipulation violate the autonomy of their targets and are wrong to the extent that they do so (41). However, there has been a movement of scholars attempting to justify some types of manipulation on the basis that they do not violate autonomy. In this literature, manipulation is referred to as "non-argumentative influence" so as to avoid the negative connotations of the more common term (Blumenthal-Barby 123). While I suspend the assumption that it always involves a violation of autonomy, I continue to refer to the concept as manipulation. Authors who maintain that certain forms of manipulation do not violate autonomy typically appeal to the "hypothetical consent" of the targets. According to this view, any intervention respects autonomy as long as the target would consent to it, were he not suffering from reasoning failures (Le Grand and New 125). Blumenthal-Barby supports this position, evaluating the moral status of acts of manipulation based on whether they (1) do not block off any available options and (2) would be

acceptable to the targeted individual, were she fully informed and included in the development of the strategy (126).

Reliance on hypothetical consent is worrisome precisely because it is hypothetical. Many would argue that we cannot know such things as whether our targets would consent to an act of paternalism. More explicitly, since the consent does not exist in the actual world, it cannot justify actions that exist in the real world (Le Grand and New 125). Even if an argument based on hypothetical consent can justify some forms of manipulation against the objection that they violate autonomy, it will only justify means-related cases. It seems that a target will hypothetically consent to precisely those instances of manipulation that bring about his long-term goals. Ends-related manipulation, then, will violate autonomy, and since libertarian paternalism cannot guarantee to be means-paternalistic toward all its targets, it does not avoid the autonomy objection, despite doing away with coercion.

If there were no concern that nudges would lead some targets to behave against their best interests, then the conclusion would not be so dire. It is plausible that people are rational enough to avoid nudges that do not serve their ends, in which case libertarian paternalism would have a morally significant advantage over restrictive means paternalism by avoiding both ends-related coercion and ends-related manipulation. However, recall that everything up to this point has been discussed under the assumption that true means paternalism is successful in preserving autonomy. At the end of Chapter 2, I raised a second concern regarding the argument underlying this claim and suggested that even true means paternalism may offend our autonomy. If this is the case, then the moral status of means-related coercion and manipulation must be considered.

The features that determine whether manipulation is morally justifiable are the same features that identify the moral status of coercion (Mandava and Millum 42). As explained

earlier, authors have argued that the hypothetical consent of a target may indicate that an instance of manipulation does not violate autonomy. The same might be true for instances of coercion. However, these arguments rely on counterfactual claims about people's thoughts and actions in a "fully rational" state, which are difficult, if not impossible, to determine. Nonetheless, there are robust accounts of this sort, including one briefly introduced in Chapter 1, referred to as soft paternalism and defended most famously by political and legal philosopher Joel Feinberg. Both Sunstein and Camerer mention soft paternalism in their works, employing it as a further strategy for justifying libertarian and asymmetric paternalism (Sunstein 20, Camerer et al. 1213). Soft paternalism, addressed in Chapter 4, is limited in scope, and this has led some scholars to defend "hard paternalism" or "coercive paternalism" instead. These authors concede that paternalism offends autonomy but argue that the cost of this is not as severe as some claim it to be and is outweighed by the benefits that result from certain policies. This is the position held by philosopher Sarah Conly, whose work will also be discussed in the following chapter.

Transparency, Trust in Government, and Political Feasibility

The threat to autonomy is not the only concern about the government instating paternalistic regulations. Other worries and potential costs will be discussed in Chapter 4. However, two objections that are specific to libertarian paternalism should be considered here, as well as an important practical benefit of nudging that is often left out of ethical discussions. First of all, nudges are criticized for being less transparent than restrictive varieties of paternalism. This may be problematic either because it directly relates to the way in which they bypass people's awareness, which poses an extra threat to autonomy, or because it may make it more difficult for the public to protest and stop acts of paternalism that it finds unacceptable. The first

issue draws attention to the way in which nudges differ from both persuasion and coercion. Both of these strategies engage with the target as a rational agent—even if coercion is morally problematic, it at least engages with the target as someone who can evaluate options effectively. Libertarian paternalism instead *relies* on the irrationality of its targets in order to operate—a characteristic that some find troubling. It is not clear, though, that this feature of nudges has any bearing on whether a policy violates autonomy. British philosopher Luc Bovens suggests that bypassing reasoning reduces a target's control over his situation even further than coercion does, and Blumenthal-Barby also claims that measures that bypass reasoning seem more threatening to autonomy. Still, a successful formal account of this has not been made.

The second issue with transparency plays on the fear that people have of extreme government intrusion in the form of subliminal messaging. If the government is allowed to influence us in ways we are unaware of using nudges, then we may slide down the slippery slope to allow subconscious influence. Once we allow this, the government cannot be stopped from implementing intrusive means-paternalistic practices and even acts of ends paternalism because there will be no accountability. Although this type of totalitarian government system is troubling, the concern is exaggerated. Although libertarian paternalism lacks transparency in its interactions with targets, it is not as if the policies themselves are passed or implemented in secret. Furthermore, if this concern is justified, then even nudge policies can be made more transparent, simply at the cost of some level of effectiveness (Le Grand and New 142-143).

An additional objection to using nudges instead of restrictions to encourage behavior change is that they are only minimally effective and, as a result, may damage the perceived value of public policy and diminish the public's trust in the government's ability to govern successfully. Calorie-labeling regulations are a perfect example of this problem. Though they are

so non-invasive that many recognize them as an educational strategy rather than an act of paternalism, they have been accused of being overly intrusive and at the same time ridiculed for being ineffective. These results are rarely interpreted as a motivation to do more—to instate more nudges or impose harsher regulations so that obesity prevention policy can be more effective. Rather, they are seen as evidence of the government's inability to understand and help the public and a good reason to leave people to make their own choices. If libertarian paternalistic policies were to become the most popular and widespread form of regulation, then health policy could become known for being unproductive, further reducing the confidence that people have in public services. Furthermore, because nudges can be inexpensive to implement in comparison to restrictions (Thaler and Sunstein *Nudge* 13), public health may come to be even more severely underfunded as a result of the focus on nudge policy.

Obviously, there is another way to view these issues. Perhaps the most appealing feature of nudges is that they tend to be far more politically feasible to pass and implement than taxes and bans, for many reasons. It is common knowledge that people, especially Americans, are wary of paternalism, government overreach, and the "nanny state." Within the field of obesity prevention, researchers know not to use language that suggests any restriction or limitation of options. Parents are happy to have healthy meals added to menus in schools and restaurants but not to have unhealthy options removed. Similarly, restaurants are willing to introduce low-calorie foods, meals, and entire menus in response to calorie labeling regulations but refuse to reduce portion sizes and calorie contents of the least healthy items. People and firms are threatened by restrictions and annoyed by taxes, and libertarian paternalism is sensitive to this. As Thaler and Sunstein write, "if people want to smoke cigarettes, to eat a lot of candy, to choose an unsuitable health care plan, or to fail to save for retirement, libertarian paternalists will

not force them to do otherwise—or even make things hard for them" (5). Despite the concerns

they provoke, nudges may be the only way we can make a dent in the problems that face us

today.

Chapter 4
Overcoming the Autonomy Objection

The traditional argument against paternalism rests on two premises: (1) that paternalism violates autonomy, and (2) that the preservation of our autonomy is more important than any other reason for which paternalism would be imposed. In order to defend policies such as menu labeling, optimal defaults, and sugary beverage taxes, then, one must deny either premise (1) or premise (2). In Chapter 2, I discussed one strategy for defending paternalism against the autonomy objection, which involved rejecting premise (1) for a subset of paternalistic acts—those that are means-related. This attempt faces two challenges: first, there is the problem that means paternalism cannot guarantee to be means-paternalistic for every target in a population; second, there is the concern that even if it could, means paternalism would still violate its targets' autonomy. Chapter 3 introduced the concepts of libertarian and asymmetric paternalism in the effort to resolve the first of these two worries. In this chapter, I explore the second challenge. First, I consider the definition of autonomy and the frameworks that have been proposed for determining what undermines it. Then, I draw on Joel Feinberg's theory of "soft paternalism" in order to illuminate the problem for means paternalism and present his alternative strategy for avoiding the autonomy objection. Finally, I raise the concern that Feinberg's account does not extend to include the policies of interest in this project, and I introduce a last approach for defending paternalism that does this successfully—Sarah Conly's "hard paternalism." Her account differs from others in that it takes premise (1) of the autonomy objection to be true and denies premise (2), which requires a consideration of the value of autonomy. In this context, I briefly examine the utilitarian defense of anti-paternalism.

Autonomy and the Objection to Means Paternalism

Although most scholars who defend paternalism provide some characterization of autonomy in their works, this tends to be minimally specified, which leaves it unclear to the general reader how exactly their version of paternalism avoids violating it. A satisfactory definition of autonomy and, more importantly, a framework for determining what forces undermine it, is difficult to develop. The variety of autonomy relevant to this discussion is what philosophers call *personal autonomy*. This differs from *group autonomy*, which refers to the authority that a group of people has, or ought to have, to independently govern itself (Buss sec. 1). It also stands in contrast to *moral autonomy*, which is the capacity one has to impose moral law on oneself (Christman sec. 1). Personal autonomy is often characterized as self-rule or self-government and applies to all spheres of life. Generally speaking, an autonomous individual determines which actions she performs, decides what goals to work toward and what ideals to hold, and is in control of her life's course. For the purposes of this discussion, I maintain that most adults act autonomously most of the time, despite the fact that we exhibit bounded rationality and operate in a society full of other human beings, norms, and environmental factors. A reasonable account will describe some *basic* level of autonomy that all functioning adults have with respect to most of their decisions, rather than an *ideal level* that can only be obtained when one is isolated from all external influences. In this section, I explore four categories of frameworks for determining whether something violates our autonomy, highlighting their advantages and limitations. It is not my intention to resolve the ongoing debate among scholars of autonomy and free will regarding these views, but rather to provide a deeper understanding of what constitutes an invasion of autonomy so that this can be considered in relation to different defenses of paternalism.

The difficulty of determining what entities and conditions undermine the capacity for self-rule arises because of the relationship between personal autonomy and agency. Every individual has agency, in that he initiates all of his own actions, with the exception of life-sustaining bodily functions and reflexes. Thus, in one sense, an individual cannot lack authority over his behaviors because no one other than him initiates them. Despite this, we excuse people from responsibility for their actions when they are coerced, addicted, or manipulated in a particular way. In this sense, individuals *can* lack authority over themselves. It is generally agreed that an agent lacks autonomy when his motivation for acting derives its power from some external source rather than from within himself. For example, if he is enslaved, and he performs an action in order to avoid punishment, then he does not exhibit autonomy with respect to that action because his motivation clearly derives power from the ruler. A ruler's control is not the only force that we consider strong enough to override internal motives, but it is difficult to characterize which forces qualify as such. In order to do so, one must have a framework for evaluating motivations that consistently demonstrates which ones derive their power from a sufficiently external source such that we lack autonomy, while maintaining that most derive power from within us, even though we are constantly influenced by the environment.

According to philosopher Sarah Buss, scholars have proposed four categories of these frameworks. In Chapter 2, I briefly introduced the "coherentist" account, on which a person has autonomy with respect to her action only insofar as her motivation for performing it coheres with her higher-order desires, long-term goals, stable preferences, or overall character traits. On this view, it may seem that an act of means paternalism does not violate its targets' autonomy because, under the influence of the policy, the targets are moved to act by a motivation that coheres with their ends. However, one might worry about this interpretation because, although

the targets' post-policy motivations match their ends, this does not ensure that their pre-policy

motivations did not also match their ends. If someone is enslaved by a ruler who forces him not

to smoke cigarettes, knowing that he values his long-term health, is this an invasion of his

autonomy? One might think that the answer to this depends *not* on whether the motivation to

avoid punishment coheres with the person's end of long-term health, but rather on whether he

also acted autonomously with respect to smoking. Thus, we must come to a conclusion about

whether people whose behaviors are motivated by irrational tendencies and cognitive biases act

autonomously or non-autonomously with respect to their decisions. If they are understood to act

non-autonomously, then acts of means paternalism avoid the autonomy objection; however, if

they are understood to act autonomously, then means paternalism faces a challenge.

Regardless of whether means paternalism avoids this challenge, it seems that the

coherentist account focuses on only one part of the important information in determining what

undermines autonomy. This raises the question of whether this framework is the best mechanism

for sorting cases into those that involve a violation and those that do not. Indeed, the coherentist

account is known to face several challenges. First, it requires that we view agents as having

synchronic and diachronic selves—that is, points of view that are unified and persist over time.

Some object to viewing the self in this way (Buss sec. I). Secondly, coherentist views do not

require that people's values, preferences, or higher-order desires respond to facts about reality, or

that individuals be in control of their attitudes. Philosopher Susan Wolf proposes a thought

experiment in response to a closely related view of agency that illuminates this concern. She

proposes the case of JoJo, the son of an evil dictator, who is raised to have all of the same evil

views as his father and to hold his father as a role model. Although the coherentist would hold

JoJo responsible for any wrongdoing he does as a result, Wolf argues that he should not be

considered in control of his actions (Wolf 46). While this case may seem extreme, it is easy to see how this worry arises in the real world. For example, there are groups of women around the world who voluntarily undergo and support female genital operations (also referred to as female genital mutilation or FGM), fully aware of the medical risks and effects on the ability to experience sexual pleasure (Walley 406-410), and it is controversial whether they have autonomy with respect to these decisions or are influenced by their societies in a way that undermines their powers for self-rule. Is it enough that these women's decisions cohere with their overall beliefs and preferences about the operations? As a result of this worry, Wolf promotes a view of agency that not only requires the agent to perform actions that cohere with his deeper self, but also requires him to have a deep self that is "sane" (Wolf 46). In a similar way, one might advocate for a view of autonomy that requires a person's ends to interact with reality and respond to the agent's revisions in a certain way.

A second type of framework, referred to as the "reasons-responsiveness" view, responds to this challenge. Like coherentist accounts, the reasons-responsiveness view assesses the characteristics of an individual's motivations for acting. However, instead of requiring that these motivations cohere with the individual's ends, the view requires that they be the kind of motivations that the agent could reevaluate in response to a wide variety of facts about reality, were they presented. This framework is appealing because it identifies all those individuals who have been given their beliefs, rather than coming to them by way of experience. It also allows us to conclude that a person such as JoJo, who has been indoctrinated to believe that murder is morally permissible, does not act autonomously when he commits a murder, unless he has been given some legitimate opportunity to reevaluate these beliefs. A third set of accounts, which also relies on our deliberative capacities, is based on an individual's responsiveness to *reasoning*. On

"reasoning-responsiveness" views, someone acts autonomously if he is able to step back from external influences, social norms, and temporary emotional responses to assess whether his motivations for performing a particular action follow logically from his own beliefs and higher-order desires. Thus, what undermines his autonomy is anything that undermines his capacity for reasoning. Unlike the views described thus far, this conception focuses on the features of the agent rather than on those of the motivations. However, it is strikingly similar to coherentist accounts because it does not require that our beliefs and higher-order desires be based in reality. An agent can have false beliefs and dangerous higher-order desires, perhaps due to being brainwashed, and yet can still be responsive to reasoning. As long as all of his actions follow from these skewed beliefs and desires, he is autonomous with respect to them. A further worry for this view is that it is too demanding. Given the evidence that we only exhibit bounded rationality, it is likely that we fail to respond to reasoning with respect to many of our actions, including ones that we would like to consider autonomous.

A fourth and final explanation of what undermines our autonomy is "incompatibilism," which is the view that we do not in fact govern ourselves. Incompatibilism is a position regarding not only autonomy, but also agency. Those who defend incompatibilism take into account the fact that every event that occurs in the world is causally determined by something that happens before it and conclude that, since *we* do not cause the motivations that guide our actions, we are not in control of our actions. Incompatibilists get their name precisely because they argue that free will (the claim that we decide what actions to take) and determinism (the claim that every event has a causal explanation) are not compatible with one another. In other words, our actions are pre-determined, so we have no agency, let alone autonomy, with respect to them. Incompatibilism, though appealing for some philosophical reasons, will not be

considered seriously here because we generally take ourselves to have autonomy, and at least free will, with respect to our actions. Unfortunately, the reasons-responsiveness position also risks leading to an incompatibilist conclusion in the following way. On this account, a person must have ends, higher-order desires, or values that can be reevaluated and resisted on the basis of newly acquired facts about reality. Meanwhile, there are many aspects of our stable preferences, identities, and biological characteristics that we cannot be expected to reconsider or reject. The reasons-responsiveness view indicates that if someone acts on these traits or ingrained values, she cannot be held accountable for her action; however, we tend to think that we are responsible for actions of this kind, as most of our actions *are* influenced by our preferences and traits. In this way, the reasons-responsiveness view requires us to be in control of the circumstances of our identities in the same way that the incompatibilist view requires us to *cause* the factors that determine our actions in order for us to have free will.

There is an additional reason to favor coherentist and reasoning-responsiveness accounts over the others: it is not obvious what the connection is between being responsive to reality and being autonomous. In contrast, it is clearly important for our motivations to be coherent with our long-term goals because autonomy relies on us being able to direct our lives in the way we see fit, and it makes sense to view this judgment as relying on our ends and values. Similarly, it is clear why responsiveness to reasoning is importantly related to autonomy, since having the capacity to act in a way that follows from our own beliefs and desires will ensure that we are in fact more capable of directing our lives in a way that reaches our ends. Still, whether or not our ends are consistent with reality does not obviously play a role in our capacity to rule our lives in accordance with them. Buss attempts to save the reasons-responsiveness view by adding a premise: *if* we value, or higher-order-desire, ourselves to act in a way that is responsive to facts

and reasons, *then* it will be important to assess whether something blocks us from achieving this end because its doing so will undermine our ability to direct our own lives. However, her defense simply appeals to a combination of the coherentist view and a view on which all people hold responsiveness to reasons as an end for themselves. In this way, her defense of the framework makes it vulnerable to more objections than the coherentist view alone faces.

Although none of these four frameworks are entirely unproblematic, the coherentist view is promising, and it continues to be the most prominent view discussed in the literature on paternalism. Given this understanding, the challenge is to identify which goals, desires, and character traits belong in the conception of the individual's "true self," so that it can be determined whether particular motivations cohere with this. As it is construed by various coherentists, the true self does not incorporate desires that are formed under the influence of extreme emotions, short-term goals that conflict with an individual's ends, or decisions made based by faulty reasoning. This understanding is what allows us to declare that means paternalism avoids the autonomy objection. However, just as we do not require that adults be entirely isolated and calm in order to act autonomously, perhaps we should not hold this idealized self as the standard to which we compare all motivations for action. This view underlies Joel Feinberg's account of soft paternalism, which is the subject of the section that follows.

Soft (Anti-) Paternalism

In his exploration of the philosophical foundations of nudge theory, Cass Sunstein writes that nudges fall generally into the categories of means paternalism and soft paternalism (Sunstein 20). Similarly, Colin Camerer and his colleagues write that asymmetric paternalism is both

means-related and justified on soft-paternalistic grounds (Camerer et al. 1213). That these

authors classify their views as soft-paternalistic as well as means-paternalistic indicates that they

are prepared to face the objection that not all means paternalism avoids offending autonomy. In

this section, I examine Feinberg's account of soft paternalism and describe how it has been

stretched to include the types of practices that Thaler, Sunstein, and Camerer defend. Then, I

explore the difference between soft paternalism and means paternalism and conclude that,

although soft paternalism successfully justifies many government practices, it cannot be applied

to the three policies of focus in this project. This leads me to consider Sarah Conly's defense of

hard paternalism and its opposition in the section that follows.

Acts of soft paternalism do not infringe on the autonomy of their targets because they

intervene with behaviors that are already performed "nonvoluntarily" (Feinberg 12) or in

situations where there is reason to believe that an action is being performed nonvoluntarily, and a

temporary intervention can be used to determine whether this is the case (13). According to the

view, because the targeted individuals already *lack* autonomy with respect to their choices, any

further intervention does not violate their autonomy. It is important to note that Feinberg uses the

terms "voluntary" and "nonvoluntary" in his work where other authors substitute "autonomous"

and "nonautonomous," which again calls attention to the close link between agency

(voluntariness) and autonomy (Le Grand and New 111). The two sets of terms are considered

interchangeable in the literature and throughout this discussion. What level of autonomy must the

targeted individual lack in order for an intervention to be soft-paternalistic? Technically,

Feinberg's view justifies the prevention of behaviors that are *less* voluntary than a particular

threshold of voluntariness that is specific to the type of action. In order to evaluate the

permissibility of a paternalistic policy on his account, one must determine *both* the actual level of

voluntariness that an individual exhibits and the threshold of voluntariness that is required for her action in general. There are several factors to be considered in doing so.

First, one must ask who is performing the action, since the actual level of voluntariness is affected in part by whether the individual harms herself or consents to another person's harming her. On Feinberg's view, if an individual gives consent, then the actual level of voluntariness of her action necessarily matches or exceeds the voluntariness threshold, and soft paternalism is unwarranted (Feinberg 100). Of course, this applies only to instances in which an individual is in a position to give *legitimate* consent. A second important factor is whether the harm is *direct* or *indirect*. Cases of direct harm involve an individual either performing or consenting to an immediate injury, such as when he "swallows a lethal dose of arsenic" or "grasps an axe in his right hand and chops his left hand off," or when "I hire a surgeon to amputate my left hand" (101). In contrast, cases of indirect harm involve the individual increasing her risk of being injured, either by putting herself in dangerous situations (e.g. driving at excessive speeds) or by contributing to a disease state that increases the risk for injury (e.g. smoking). In general, direct harms require higher thresholds of voluntariness than indirect harms do. However, directness is only one of many features that affect the voluntariness threshold, so this is not to say that indirect harmful behavior can never be targeted by soft paternalism. The decision not to wear a seatbelt, for instance, has a high risk of causing severe injury and a small chance of promoting some unimportant benefit, so the associated voluntariness threshold is high, despite it being an indirect harm. Most people exhibit a level of autonomy below this threshold when they drive without seatbelts, so an intervention that forces them to wear seatbelts is soft-paternalistic and does not invade their autonomy.

The threshold of voluntariness also depends on the extent to which the individual or the individual's decision is *reasonable*. According to Feinberg, reasonableness differs from prudence: acting prudently requires a person to avoid risks, while acting reasonably simply requires him to balance potential risks with potential benefits. As long as there is much to gain from a risky action, it may be reasonable to do it (Feinberg 102-103). He also holds that reasonableness differs from rationality, which points to competence, sanity, and the general ability to deliberate. Reasoning failures, cognitive biases, social influences, and weakness of will, which I have thus far referred to as irrational biases, are instead qualified as *unreasonable* tendencies on his view. The only features that identify a person as irrational are temporary insanity and severe cognitive impairment (180). According to him, if an action is irrational, then it always qualifies as nonvoluntary, but if it is simply unreasonable, then this does not by itself determine that it is done non-autonomously (112). The voluntariness associated with unreasonable actions is determined by a combination of this feature and other features, such as the likelihood of the potential risks and benefits and the values of the ends that stand to be lost and gained (110-111). The result is that we can suffer from reasoning failures and choice failures that prevent us from reaching our long-term goals and behave rationally and autonomously in doing so. Thus, acts of means paternalism that interfere with these kinds of actions *do* violate our autonomy, unless there are other features that help to identify the actions as nonvoluntary. This constitutes Feinberg's most significant departure from scholars who claim that means paternalism avoids the autonomy objection and also marks the difference between his view and that of hard paternalism. As he explains, the practical difference between his view and hard paternalism is that "the hard paternalist would prohibit, and the soft paternalist would permit, voluntary unreasonable risk-taking" (106). Seen another way, the hard paternalist would permit

acts of paternalism that interfere with behaviors resulting from reasoning failures and choice failures, while the soft paternalist would not.

As I have said, Feinberg allows that we can act unreasonably in a way that leads us astray from our ends, long-term goals, or higher-order desires, and still have autonomy with respect to these actions. How, though, is this possible on a coherentist or reasoning-responsiveness view of autonomy? The motivation behind this behavior does not cohere with our ends, so the action should be ruled non-autonomous on a coherentist account, and we are unable to act in a way that follows from our beliefs and higher-order desires with respect to the behavior, so we should be considered non-autonomous on a reasoning-responsiveness view. Feinberg, however, objects to the narrow scope of the true self and the limited understanding of our ends that this employs. According to him, our true selves do not only include our perfectly rational ends, but also our lower-order desires, character traits, "tempermental proclivities, acquired loyalties, and spontaneous tastes." He writes, "instead of being ostracized as 'not rational,' these givens should become part of the test for the rationality of subsequent wants that must cohere with *them*" (Feinberg 111). If we expand our understanding of the true self to include these other components, then a choice that is made as a result of reasoning failures, social influences, or weakness of will may very well cohere with the true self and be ruled autonomous on a coherentist account. Similarly, a person who brings about these shorter-term desires does behave in a way that follows from his beliefs and desires, so he will be considered autonomous on a reasoning-responsiveness account. Feinberg is not the only scholar to defend this view of autonomy, and he employs support from them. For instance, he praises political philosopher Richard Arneson, who argues that we should guard people's rights to "choose and pursue life

plans that deviate from maximal rationality or that hamper future prospects of rational choice"
(474).

With this understanding of voluntariness, it becomes evident how soft paternalism and means paternalism differ. Means paternalism is said to avoid offending the targeted individual's autonomy simply because it gives him a new motivation to act that coheres with his ends. This makes no specification as to whether the target acts autonomously before the policy is instated. In contrast, acts of paternalism qualify as soft if they interfere to change actions that are done non-autonomously before the intervention. It is possible, though not likely, for a policy to be soft-paternalistic and ends-related, if it interferes with a nonvoluntary action and forces its target to behave in a different way that *also* does not cohere with his true self, broadly construed. More relevantly, policies can be both means-paternalistic and hard-paternalistic if they interfere with voluntary actions and force their targets to behave in a way that coheres better with their ends or coheres with a different set of their ends. As I have explained, Feinberg would take the three policies of interest in this project to fall into the latter category, ruling them impermissible. Sarah Conly, whose defense of hard paternalism is discussed in the section that follows, would also take them to fall into this category, but would rule them permissible for other reasons. Meanwhile, Thaler, Sunstein, and Camerer would plausibly consider such practices to be both means-related and soft, ruling them permissible. This difference results from their understanding of a larger scope of behaviors as nonvoluntary.

These observations illuminate another informative difference between soft paternalism and means paternalism: while means paternalism is typically seen as a strategy for saving the paternalist from the conclusion that all his policies are impermissible, soft paternalism was proposed as an anti-paternalistic view. Feinberg emphasizes that he would prefer to call his

account "soft anti-paternalism" in order to contrast it with "hard anti-paternalism," which is the view that absolutely no intervention to protect an individual from harming himself is warranted, but he decides that this would be too confusing (Feinberg 15). Nevertheless, he is explicit about his intention that soft paternalism be a view "closer to the liberalism of Mill than to the protectiveness of hard paternalism" (12).

Given this, should menu-labeling regulations, optimal defaults, and sugary beverage taxes be justified as instances of soft paternalism? As is evident by the kinds of policies that Thaler and Sunstein and Camerer et al. promote as soft-paternalistic, Feinberg's view has been stretched further than he would likely have approved of in two ways. First of all, many of these cases involve indirect harm. As mentioned earlier, Feinberg explicitly rejects smoking as the kind of action that should be intervened with, and he would likely come to the same conclusion for other indirect harms, such as failing to save for retirement, neglecting to exercise regularly, and consuming energy-dense, nutrient-poor foods. Secondly, the regulations that are promoted in the nudge literature all involve unreasonableness, rather than irrationality, which means that they should not automatically qualify as nonvoluntary, if one shares Feinberg's view about the distinction between unreasonableness and irrationality. Both of these characteristics have the potential to lower the voluntariness threshold of these actions, which makes it more likely that people will retain autonomy in performing them.

Still, it is not out of the question to classify the three policies of interest in this project as soft-paternalistic: given the appropriate empirical evidence there are two ways in which the actions involved could be classified as nonvoluntary. First, the case can be made that the consumption behaviors being targeted are actually direct harms-to-self, which would raise their voluntariness thresholds. There is an important difference between the two types of actions that

Feinberg calls indirect—there are those that put the individual at risk for being harmed, such as

driving at excessive speeds, and there are those that directly harm the individual in small,

imperceptible ways that together increase her risk of experiencing a larger, more obvious harm,

such as smoking, which can lead to heart disease, lung cancer, and other chronic issues. The

latter kind of "indirect" harm is actually direct because, although it goes unnoticed, it has a

harmful effect with each repetition. The excessive consumption of sugar, trans-fats, and other

components of food can be shown to fall into this category if there is reason to believe that these

components of our diets influence the body in negative ways at the cellular level, even in small

doses. Evidence that these behaviors harm the body *in advance* to the development of full-blown

heart disease, high blood pressure, diabetes, and other chronic conditions would raise the

voluntariness thresholds of the behaviors that calorie-labeling laws, optimal defaults, and soda

taxes aim to correct, such that these policies would be qualified as soft-paternalistic. A second

finding that would lead to this conclusion is evidence that the behaviors the policies target are

addictive. Addiction is arguably a variety of severe cognitive impairment, and not simply a type

of weakness of will, so if obesity-promoting foods were shown to be addictive, then their

consumption would count as nonvoluntary behavior and would warrant intervention.

Although I have established what type evidence would be required to make the case that

these policies are soft-paternalistic, I hold that this standard has not been met. Defining the

consumption of tasty and fattening food food as non-autonomous is worrisome because it is so

universal, quotidian, and often integral to our identities and cultures. I certainly do not deny that

there is consensus about which foods and beverages we should consume and which we should

avoid in order to lead a healthy life: it has long been known and promoted that a diet of the

appropriate energy density and consisting of mostly fruits, vegetables, whole grains, and lean

proteins is beneficial to health. However, in order to classify the choice to consume obesity-promoting foods and drinks as nonvoluntary, there must either be a *direct* connection established between a single food, nutrient, or food group and a quantifiable harm to the body, or proof that these products are addictive. In addition, this evidence must be strong—Sarah Conly argues that even smoking, for which there is far more evidence of direct harm and addiction, is not irrational enough to qualify as involuntary (Conly 179). This need to establish direct causation and severe risk of harm explains both the immense difficulty of defending obesity prevention policies against the autonomy objection and the intense focus that nutrition scientists seem to have on identifying individual nutrients and foods as causes of the obesity epidemic.

The Case for Hard, Coercive, Means Paternalism

In *Against Autonomy: Justifying Coercive Paternalism*, philosopher Sarah Conly defends acts of paternalism that violate their targets' autonomy insofar as they generate more benefits to their wellbeing than costs and only interfere with choices that are peripheral to their identities. As such, her view is hard-paternalistic, coercive, and means-related. First, it is hard-paternalistic simply because it justifies policies that *do* violate autonomy. Although Conly does not object to acts of soft paternalism, she aims to permit more practices than Feinberg allows—specifically, harms-to-self that arise as the result of reasoning failures and weakness of will, such as smoking (Conly 179). She argues that Feinberg's account cannot reasonably be stretched to include these policies, but that they are morally permissible, and even obligatory in some instances (3). Second, she promotes restrictive practices, rather than nudges, not because there is anything objectionable about nudges, but because they are no less offensive to our autonomy than taxes and bans, and at the same time yield far fewer benefits. As discussed in Chapter 3, libertarian and

asymmetric forms of paternalism manipulate their targets, which undermines their autonomy at least in those cases where restrictive varieties do. Conly does not do away with nudges entirely—since her view only allows paternalism that yields greater benefits than costs to their targets' wellbeing, and nudges impose very low costs, she promotes the use of nudges in circumstances where the potential benefits are only moderate (129). Third of all, Conly's view is means-paternalistic. Throughout her book, she clarifies that she supports only those policies that "assume agents' ends as given, and try to substitute external regulations for what is likely to be poor instrumental reasoning" (Conly 88). On her account, the moral status of paternalism depends on its capacity to promote *subjective* wellbeing for individuals. She writes, "I do not argue that there are objectively good ends" (43). Like Feinberg, she holds that means paternalism fails to avoid the autonomy objection in all applications, but unlike him, she takes some of the policies that it does not justify to be permissible on other grounds.

The case for hard paternalism rests heavily on the evidence of reasoning failures and choice failures that was presented in Chapter 2. Conly summarizes her own argument in the following way (Conly 2). She writes,

> In this book, I argue that the ground for valuing liberty is the claim that we are pre-eminently rational agents, each of us well suited to determining what goes on in our own life. There is ample evidence, however, from the fields of psychology and behavioral economics, that in many situations this is simply not true"

Her account of hard paternalism relies on this evidence in two ways. First, she argues that autonomy is not valuable enough such that its preservation justifies forgoing the substantial amounts of wellbeing that we could gain by enacting paternalistic regulation. The very capacity for paternalism to yield this great amount of wellbeing hinges on the fact that we are often unsuccessful in reaching our own highest ends. Second, her reason for thinking that autonomy is *not* this valuable relies on our irrational tendencies. She argues that we should only consider

autonomy to be more valuable than the wellbeing we stand to gain from paternalism if we are

perfectly rational beings. On the one hand, if autonomy is valuable instrumentally, as a way of

obtaining our ends, then it is clearly not perfectly valuable when we are not perfectly capable of

making the decisions to do this. On the other hand, if autonomy is valuable innately, then this

view likely rests on a conception of the human being as defined as our ability to self-govern,

which does not hold up when we are actually quite poor self-governors (2-3).

Who are Conly's opponents? As I have implied, one set of opponents may be

deontological thinkers, who judge the permissibility of actions on the basis of duties, obligations,

or rights. Deontologists come in many varieties, including Kantians, moral absolutists, divine

command theorists, and other more contemporary thinkers. If autonomy or freedom of choice

can be grounded as a right on any of these views, then no amount of wellbeing can justify

violating it, so paternalism is always unwarranted. For the purposes of this discussion, I set aside

this category of challenges, noting only that if any such account is meant to justify strict anti-

paternalism, then it must ground a right to autonomy in all realms of life, even those that do not

seem central to our identity or status as humans, and it must not involve a right or obligation that

arises due to our status as perfectly rational beings. Instead of this, I consider objections based in

consequentialism, which is the view that the moral status of any action depends on the

consequences that it brings about. Specifically, I examine a number of challenges raised by

utilitarian philosopher John Stuart Mill who, like other utilitarians, judges the permissibility of

actions on the basis of the amount of utility, or happiness, that they generate.

On the basis of utilitarianism, an act of paternalism is justified if and only if it generates a

net positive amount of utility. Ends paternalism, of course, yields no subjective benefits for the

individual, and at the very least leads to their losses of autonomy and liberty. However, once it

has been accepted that we are not perfectly rational agents, there is the potential for means

paternalism to yield very high benefits for the individual, by allowing her to reach her long-term

goals and desires. Meanwhile, paternalism has been said to cause the loss of utility in several

ways. First, insofar as liberty and autonomy are both components of wellbeing, most or all kinds

of paternalism will impose costs on the target. Additionally, John Stuart Mill argues that

paternalism limits the target's individuality (Mill 54) and causes him to be inauthentic and

lacking in affect (66). Furthermore, Joel Feinberg worries that paternalism infantilizes its targets,

which further diminishes their ability to learn from their mistakes (Feinberg "Legal Paternalism"

3). Others raise the concern that paternalism alienates the individual from his own goals and

plans or from the society that restricts him (Conly 78-79) and that it results in a loss of self-

esteem (87). Public health educator and philosopher Dan Wikler claims that paternalistic policies

impose undue psychological costs on people by making them feel excessively punished. Since

the behaviors that attract paternalistic regulation already cause harm to the individuals that

perform them, further punishment is likely to cause unwarranted suffering and humiliation

(Winkler 40-43). Lastly, authors are increasingly concerned that paternalistic policymaking

requires excessive collection of information, leading to a loss of "informational privacy" (Conly

132-137).

  That paternalism promotes these benefits and imposes these costs is controversial. First,

one can raise utility-based objections to the claim that means paternalism can benefit its targets,

which were identified in Chapters 2 and 3. Some point out that policymakers are vulnerable to

corruption, temptation, and the same reasoning failures that everyone else suffers from, and

others suggest that they are unable to judge the long-term goals of their targets from the outside

without intensive conversation (Le Grand and New 46-47). Mill himself argues that acts of

paternalism fail to increase wellbeing in part because policymakers are not in a good position to assess what is in the interests of the individual (Mill 81) and in part because individuals are more invested in promoting their own wellbeing than the state is (74). However, these concerns are not as serious as they may seem. Although policymakers are admittedly susceptible to corruption and poor decision-making, they are not any more inclined to act poorly as a result of this in the context of paternalistic policymaking than in the context of any other decision. In fact, they may be less inclined to bend to temptation in developing paternalistic policies because those that are passed will undoubtedly affect them, and not just others. Furthermore, technical limitations and framing effects plausibly affect policymakers less than they affect the public because policymakers have access to more expert advice and, most importantly, more decision-making time than people who have to make choices in the heat of the moment. Another way to express this is that policymakers are more likely to be processing these decisions through their System 2 reasoning than the targets of paternalism because they are further removed from the emotions of the decisions and make the decisions in a structured fashion as their entire job (Thaler and Sunstein *Nudge* 249-251).

Conversely, one might object to the claim that paternalism imposes any one of the costs mentioned above. For instance, although paternalism diminishes its targets' autonomy and liberty in the immediate sense, some instances clearly preserve long-term, or overall autonomy. Indeed, one of the significant exceptions that Mill makes to his anti-paternalistic "harm principle" is that paternalism is warranted when it prevents individuals from contracting themselves into slavery (Mill 101). He writes,

> In this and most other civilized countries, for example, an engagement by which a person should sell himself, or allow himself to be sold, as a slave would be null and void, neither enforced by law nor by opinion. […] By selling himself for a slave he abdicates his liberty; he forgoes any future use of it beyond that single act. He therefore defeats, in his

own case, the very purpose which is the justification of allowing him to dispose of himself.

Some authors have extended this rationale to defend all "autonomy-enhancing" forms of paternalism, even those that less-obviously preserve future liberty (Brink Section 3.7). On this view, any act of paternalism that extends someone's life maximizes their overall liberty because it literally extends the amount of time they have to act with liberty. However, this position is based on a technicality that most find to be unpersuasive, and although acts of paternalism may maximize overall liberty by extending life, they surely do not maximize overall autonomy by doing so. Furthermore, there is debate as to whether Mill should have allowed the prevention of slavery as an exception to his harm principle, or whether he was "led astray by a correct belief that prohibition of slavery is justifiable" and should have justified it on non-paternalistic grounds (Arneson 473).

One might also object to the view that paternalism diminishes its targets' individuality. Mill seems to understand individuality in a few ways—as uniqueness, as excellence, and as the full development of the self (Arneson 479). If individuality is the absence of conformity, then based on the findings presented in Chapter 2, means paternalism may actually increase individuality by guiding people away from the behavior that results from social norm influences. Conly writes that Mill's only support for thinking that paternalism would diminish individuality was that he "overestimated the degree to which humans would develop in varying, individual, and authentic ways if left without government controls" and "assumed that government intervention would always be a conservative influence, imposing the social conventions of the majority and repressing innovation" (55). Similarly, if individuality is understood as excellence, or a departure from mediocrity, then paternalism has the potential to promote it by helping people to reach their ultimate goals. Finally, on the understanding that individuality is the

development of the self, paternalism promotes it by allowing individuals to attain the ends and

higher-order desires that in part comprise their true selves (Conly 53-54). One can also challenge

the claim that paternalism decreases self-esteem. Although one might think it is damaging to our

self-esteem to be told that we require help to behave in ways that bring about our own ends,

Conly argues that this is only detrimental if we are indeed perfectly rational agents and are thus

being falsely accused. Since we are not fully rational, paternalism may actually allow us to gain

self-esteem, as it will lead us closer to a state of accepting our own abilities and limitations (98).

In building awareness of our limitations, we will be able to exhibit a form of humility that many

view as a virtue of character (189).

This conception of paternalism as beneficial *beyond* its capacity for leading people

toward their own ends, recurs throughout Conly's book. As noted earlier, she promotes some

acts of paternalism as obligatory and argues that it is actually more respectful to restrict people's

self-harming behaviors than it is to do nothing in order to respect their autonomy. She writes,

"those who say we should respect autonomy by letting people hurt themselves irreparably do not,

on my view, show as much respect for human value" (Conly 1-2). Conly also claims that

paternalistic regulations respond to people's abilities in an appropriate manner (9), free them

from "nagging concerns," and prevent them from having to spend time and energy self-

regulating their behavior when they could be doing things that they enjoy (90). The

understanding of paternalism as *more* respectful than the absence of intervention is also evident

in British psychologist David Halpern's book, *Inside the Nudge Unit,* which documents the

recent history of the U.K. government unit dedicated to developing effective nudge policies. He

writes, "years of nudging have taught me to be more respectful of the human condition,"

(Halpern 324) and "such interventions are about understanding who we are; about connecting

and communicating with each other better; and, frankly, about designing services for human beings, instead of 'econs'" (121).

Regardless of how one evaluates the costs and benefits of paternalism that have been presented here in light of the challenges that have been raised, two things remain clear. First of all, there are likely to be many cases of paternalism—even hard paternalism, that stop us from performing *such* detrimental behaviors that the net utility resulting from them is positive. Because of this, utilitarianism alone does not justify categorical anti-paternalism. Second, on Conly's view, what determines whether a policy is permissible is not the level of voluntariness of the target's action beforehand, but rather a careful consideration of (1) the level to which the target's post-intervention motive for acting coheres with her ends, (2) the likelihood of that policy bringing about a net gain of wellbeing for her, and (3) the status of the associated liberty (e.g. the liberty to smoke cigarettes) with regard to the individual's identity. This determination method is consistent with the strategy that she employs throughout her entire book, which involves evaluating the legitimacy of utility-based objections to paternalism and then internalizing the legitimate ones by accepting them as costs and weighing them against the benefits that are yielded. She extends this practical strategy to the autonomy objection, which allows her to overcome the mystique that often accompanies it, without ignoring its seriousness. As she explains, the intuition that violating autonomy for any purpose is inherently wrong is not itself a reason or an argument for rejecting policies that have the capacity to make our lives significantly better (194). Perhaps, given that we are not perfectly rational beings, what was underlying this strongly felt intuition has been weakened.

Conclusion

I began this discussion with three obesity prevention strategies, the objection that they violate our autonomy, and the promise that they could avoid this objection in virtue of being nudges. In what followed, I aimed to defend the policies against the autonomy objection by classifying them as particular varieties of paternalism. As I have demonstrated, the autonomy objection contains two premises—that paternalism violates our autonomy, and that this autonomy is more valuable than the wellbeing that we stand to gain from paternalism. I first explored ways in which the policies in question could avoid the first premise, aiming to distinguish them as means-paternalistic, libertarian-paternalistic, and soft-paternalistic. In Chapter 2, I explored the difference between means-related and ends-related paternalism and the findings from the fields of cognitive psychology and behavioral economics that allow us to consider the three policies in question to be instances of means paternalism. Having done this, I considered whether their status as means-related was enough to dismiss them from the autonomy objection, and I raised two concerns about this. First, population-wide acts of means paternalism are not necessarily means-related toward all of their targets; second, even if they were, they may still be problematic because the way in which they respect autonomy is subject to doubt.

Chapter 3 presented the concept of libertarian paternalism in depth, along with the similar concept of asymmetric paternalism, both of which were shown to address the first of the two concerns raised about means paternalism in the previous chapter. I explained that the advantage of nudges over restrictive policies is that they are intended to minimize the amount of ends-related paternalism that occurs when a means-related policy targets an entire population. Unfortunately, there was reason to think that nudges are not entirely successful in accomplishing

this goal, both because they do not reach enough people and because they risk reaching people who do not stand to benefit from them. It was concerning that nudges do not reach all the people who could benefit from them, especially because those who are left out may be precisely those who stand to benefit the most from government paternalism. Furthermore, it was worrisome that nudges may be ends-paternalistic toward some of their targets in the same way that restrictive policies are because the ends-related manipulation that they employ is problematic in the same ways that ends-related coercion is.

At this point in the discussion, I set aside these concerns and turned toward the second worry raised at the end of Chapter 2, regarding means paternalism and its failure to guarantee that practices will respect the autonomy of their targets. In Chapter 4, I elaborated on the challenges of developing a definition of autonomy and a framework by which to assess which forces undermine it, and I further explained the autonomy objection, with respect to means paternalism. Then, I presented Joel Feinberg's account of soft paternalism, which evaluates the permissibility of paternalistic practices not on the basis of their alignment with the target's true self, but rather on the basis of how voluntary the target's theoretical prior action would have been. I showed that the scope of soft paternalism as he defines it is far more limited than that of means paternalism because he holds that our irrational tendencies and cognitive biases do not undermine our ability to behave reasonably and autonomously. Meanwhile, Thaler and Sunstein, Camerer, and other scholars have extended his view of soft paternalism by considering more of our actions to be nonvoluntary, which allows them to qualify far more practices as morally permissible in virtue of being soft-paternalistic. I argued that the policies in question in this project should not be classified as soft-paternalistic unless further evidence is accumulated that identifies fast food and soda consumption as a direct harm or reveals them as addictive.

In the final section, I considered whether these strategies and others like them could be defended on hard-paternalistic grounds, given that they should not be included as instances of soft paternalism. I presented Sarah Conly's view of hard paternalism, which holds acts of paternalism to be permissible even when they do violate our autonomy, as long as they yield more benefits than costs for their targets and do not interfere with liberties that are central to their identities. As such, Conly's view differs from the others in that she accepts the first premise of the autonomy objection and defends certain practices against the second premise. I considered utility-based objections to paternalism, which are taken into consideration as costs under Conly's view, and I reviewed her criteria for assessing the moral permissibility of paternalistic actions.

Here, I examine whether the three policies in question are permissible on her view and which of their features identify them as such. First, recall that Conly's three criteria are: (1) the extent to which a policy is means-paternalistic; (2) the potential for that policy to bring about a net gain of wellbeing for the target; and (3) the importance of the liberty that is denied by the policy. First, consider the case of menu labeling. It is plausible that this practice is means-related because, for the most part, people do hold good health and long life above the desire to order high-calorie options at fast food restaurants, especially when this choice is made outside of the decision-making environment. New Year's Resolutions would not be so commonly dedicated to dieting if this were not the case. We also know that restaurants capitalize on many of the reasoning failures discussed in Chapter 2 in order to raise their profits. While restaurants do not employ these strategies for the purpose of making people unhealthy, as Wansink emphasizes repeatedly in his book, the same practices that maximize their profits tend to lead consumers to purchase more energy-dense, nutrient-poor food, eat it more quickly, and feel hungrier sooner. Second, menu labeling does have the potential to bring about a net gain in wellbeing for the

individual. This is both because the regulations will likely have greater effect on consumers once they are implemented everywhere, thus leading to lower-calorie orders, and because restaurants are already reducing the calorie contents of their foods significantly in preparation for these regulations, as described previously. Once all restaurants are required to list nutrition facts on menus in stores, restaurants will not want to post incredibly high calorie counts. Although doing so may attract attention at first, once these practices are incorporated into daily life, it is reasonable to predict that the attraction of selling unreasonably high-calorie food will fade.

Even if these benefits turn out to be small, the costs imposed on consumers are minimal, as are the negative consequences introduced by Mill and other scholars, such as the risk of excessive punishment and loss of affect. As I alluded to throughout this discussion, some might worry that although the costs imposed on consumers are minimal, the costs imposed on firms are significant. Of course, because menu labeling and the other policies of focus here are not cost-restrictive, this is not of major concern. However, even if they were, I argue that this would not constitute a reason to judge them differently because firms currently earn an excess of profit that can be attributed to their use of manipulative strategies to encourage business. Insofar as the costs to firms do not exceed the benefits that they presently enjoy as a result of their more manipulative marketing practices, they should not be considered in evaluating these practices from an ethical standpoint. In sum, menu labeling yields net positive benefits, satisfying Conly's second criterion.

Lastly, we must consider whether menu labeling policy infringes on a liberty that is important to the identity of the individual. In virtue of being a nudge, it does not infringe on any liberties, except perhaps the liberty to order food in a restaurant without being aware of how much energy it contains. One might object to this evaluation by arguing that I have construed

this liberty too narrowly. Instead, menu labeling laws should be seen as violating our liberty to make choices that are free of government control, or excessive "nannying." However, it is unrealistic to think that we make any choices without influence from the environment and from the government. Not only are the decisions we make structured by the world around us, but they *must* be structured in some way. To hold that calorie-labeling violates our liberty to make choices free of government influence is to demand a standard of decisional freedom that is too high.

Optimal defaults and sugary beverage taxes can be evaluated in largely the same way as menu labeling regulations with respect to these three criteria. However, there are two important differences. First, optimal defaults impose even fewer costs and fewer benefits, though this does not affect their status as permissible on a hard-paternalistic account because the net benefit is still likely positive. Second, sugary beverage taxes have the potential to be qualified as soft-paternalistic because the evidence that I claimed was needed in order to qualify obesity policies as such is more promising with respect to soft drinks. It is more promising to make the case that sugary beverages harm the body directly, even in small doses, than it is to make this case for obesogenic food more generally. This is particularly easy to do if dental health is considered. In addition, there is evidence that sugary beverages may be addictive. Either of these features may raise the voluntariness threshold of soda consumption such that drinking large amounts of soda is classified as nonvoluntary, and this would in turn allow restrictions of soda to be classified as soft-paternalistic. Of course, even if the case cannot be made, sugary beverage taxes are still permissible on hard-paternalistic grounds, as they promote even more benefits than the other two policies and impose only slightly higher costs.

I have concluded that all three of the policies of interest in this study are morally permissible. However, are menu labeling regulations and optimal defaults permissible in virtue of being nudges, or in virtue of their other features? There are two ways in which their status as nudges may make them superior to restrictive practices. First of all, it reduces the costs that they impose on targets, which factors into the consideration of permissibility on Conly's view. However, libertarian paternalism also yields far fewer benefits and yields benefits for fewer people. Because many restrictive policies can yield high net benefits, it is not of great importance that policies be nudges for this reason. Second of all, the fact that these policies are nudges allows them to minimize the amount of unintended ends paternalism that results from their implementation. Putting aside the worry that I raised for this advantage in Chapter 3, it seems that this may be significant. Still, government policy is designed so that it will help the most people possible, and it is never expected to reach everyone in a population in the desired way. There are unintended consequences of any large-scale action. If an act of means paternalism is not beneficial to all of its targets, this does not undermine its status as means-related. Recall that even Arneson, who defends Mill's strict anti-paternalism, concedes that anti-dueling regulations are means-paternalistic (and permissible) despite the fact that some people in the population hold ends that are different from those of the majority. Because the anti-dueling legislation is enacted for the purpose of helping the majority of people reach their ends, it is allowable. We should view obesity prevention policies in the same way—even if a practice serves as ends-paternalism toward a small number of people unintentionally, it is still justifiable, and if this policy is able to yield very high benefits, as more restrictive policies can, then it should be implemented.

At the end of Chapter 3, I noted two additional challenges for nudges—that they lack transparency, and that they have the potential to undermine our trust in the power of public

health policy. However, I also introduced an additional benefit, which is that they are more acceptable to the public than restrictive policies at this time, especially in the United States. Through examining nudges in this project, this has revealed itself as their most significant advantage. Though they are not philosophically superior to restrictive policies, they are far more feasible to implement. In fact, during a conference call just last week with representatives from many of the most influential childhood obesity organizations in the country, the priority of creating policies and programs that do not limit people's available choices was emphasized. Ultimately, we do not want our options cut off, even if we do not intend to choose those options. Perhaps this is another example of our irrational tendency to avoid loss, which was demonstrated so well in Ariely's experiment, where subjects were given a finite number of clicks to use in virtual rooms. Whether this is irrational or not, though, it is a factor that must be dealt with in the fight against obesity.

Given this challenge, how should menu labeling regulations, optimal defaults, and sugary beverage taxes be justified to the public, and how can more restrictive policies be defended? Practically speaking, the most promising defense might refer to the harm being done by private firms to the public by way of manipulative marketing practices. Critics note that this defense is unpromising because this type of action does not count as harm under most conceptions of harm. In other words, one person cannot harm another through an action that she consents to, and people are taken to consent to the marketing practices used by firms because they continue to purchase and consume their products. However, each time this defense is raised, it appears even less convincing. Especially in the context of television advertising, we know that people's actual preferences and desires are influenced by marketing practices, a fact which has the potential to impact our ends, not just our means. If this is the case, then in what sense are we consenting to

the manipulative practices? Of course, one might object that it is unwise to view our daily interactions with the private sector as lacking consent, but this approach remains attractive.

In the introduction to this discussion, I wrote that there has been a philosophical divide on the subject of obesity prevention for some time, where some place the responsibility of addressing the issue on individuals, and others place it on the government. I should note that I have approached this issue from the perspective of anti-paternalists, who place more responsibility on the individual. The autonomy objection, though, is not the only objection that faces nudges and other minimally restrictive obesity prevention policies. There are also those who hold that these strategies place *too much* responsibility on the individual, which distracts from structural determinants of disease and contributes to the stigma that is attached to being overweight and obese. Instead, the government should target the food system more broadly, perhaps placing limits on the private sector regarding how and when it can create and sell obesogenic products, or aiming to reform farming subsidies and other national and global food policies. For the purposes of this discussion, I have set aside this view, though it is clear that it merits further study. It is at least true that on the spectrum of obesity prevention as individual responsibility or full government control, these policies are a step in the right direction.

Works Cited

Anderson, Scott. "Coercion." *The Stanford Encyclopedia of Philosophy.* Ed. Edward N. Zalta

(2015): n. p. Web. 6 March 2016.

Anzman-Frasca, Stephanie, Megan P. Mueller, Sarah Silwa, Peter R. Dolan, Linda Harelick,

Susan B. Roberts, Kyle Washburn, and Christina D. Economos. "Changes in Children's

Meal Orders Following Healthy Menu Modifications at Regional US Restaurant Chain."

Pediatric Obesity 23.5 (2015) : 1055-1062. Web. 2 April 2016.

"Applying Behavioural Insights to Health." *Gov.uk*. Cabinet Office: Behavioural Insights Team,

December 2010. Web. 17 April 2016.

Archard, David. "Paternalism Defined." *Analysis* 50.1 (1990) : 36-42. Web. 20 November 2015.

Ariely, Dan. *Predictably Irrational: The Hidden Forces That Shape Our Decisions*. New York:

Harper Perennial, 2008. Print.

Arneson, Richard J. "Mill versus Paternalism." *Ethics* 90.4 (1980) : 470-489. Web. 6 March

2016.

Barclay, Eliza. "Berkeley's Sugary Drinks Are Getting Pricier, Thanks to New Tax." *NPR*.

NPR's The Salt: What's On Your Plate, 2015. Web. 17 April 2016.

"Behavioral Economics: Past, Present, Future." *Advances in Behavioral Economics*. Ed.

Camerer, Colin F., George Loewenstein, and Matthew Rabin. Princeton: Princeton

University Press, 2004. Print.

Blanton H, Koblitz A, and McCaul KD. "Misperceptions about norm misperceptions:

descriptive, injunctive, and affective "social norming" efforts to change health

behaviors." *Social and Personality Psychology Compass* (2008) : 1379-1399. Web. 26

February 2016.

Bleich, Sara N., Julia A. Wolfson, and Marian P. Jarlenski. "Calorie Changes in Chain

　　Restaurant Menu Items: Implications for Obesity and Evaluations of Menu Labeling."

　　American Journal of Preventive Medicine 48.1 (2015) : 70-75. Web. 17 April 2016.

Blodget, Henry. "THE NANNY: Bloomberg's Soda Ban is So Offensive He's Getting Mocked

　　in The New York Times." *Business Insider*. Business Insider, 2012. Web. 10 December

　　2015.

Blumenthal-Barby, J.S. "A Framework for Assessing the Moral Status of 'Manipulation.'"

　　Manipulation: Theory and Practice. Ed. Christian Coons and Michael Weber. Oxford:

　　Oxford University Press, 2014. Print.

Bovens, Luc. "The Ethics of *Nudge.*" *Preference Change: Approaches from Philosophy,*

　　Economics and psychology. Ed. Grune-Yanoff, Till and Sven Ove Hansson. Berlin:

　　Springer, Theory and Decision Library A, 2009. Print.

Brandt, Allan M. *The Cigarette Century*. New York: Basic Books, 2007. Print.

Brink, David. "Mill's Moral and Political Philosophy." *The Stanford Encyclopedia of*

　　Philosophy. Ed. Edward N. Zalta (2014) n. p. Web. 2 April 2016.

Buss, Sarah. "Personal Autonomy." *The Stanford Encyclopedia of Philosophy*. Ed. Edward N.

　　Zalta (2014) n. p. Web. 2 April 2016.

Camerer Colin, George Loewenstein, and Matthew Rabin. *Advances in Behavioral Economics.*

　　Princeton: Princeton University Press, 2004. Print.

Camerer, Colin, Samuel Issacharoff, George Loewenstein, Ted O'Donoghue, and Matthew

　　Rabin. "Regulation for Conservatives: Behavioral Economics and the Case for

　　'Asymmetric Paternalism.'" *University of Pennsylvania Law Review* 151 (2003) : 1211-

　　1254. Web. 26 February 2016.

Charles, Dan. "How 'Double Bucks' For Food Stamps Conquered Captiol Hill." *Npr*. National

    Public Radio's The Salt: What's On Your Plate*,* 2014. Web. 20 November 2015.

Choi, J., D. Laibson, B. Madrian, A. Metrick. "Optimal defaults." *Am Econ Rev* 93 (2003) : 180-

    185. Web. 26 February 2016.

Christman, John. "Autonomy in Moral and Political Philosophy." *The Stanford Encyclopedia of*

    Philosophy. Ed. Edward N. Zalta (2015) n. p. Web. 2 April 2016.

Cialdini RB. "Crafting normative messages to protect the environment." *Current Directions in*

    Psychological Science 12.4 (2003) : 105-109. Web. 26 February 2016.

Colchero, M. Arantxa, Barry M. Popkin, Juan A. Rivera, and Shu Wen Ng. "Beverage Purchases

    from Stores in Mexico Under the Excise Tax on Sugar Sweetened Beverages:

    Observational Study." *British Medical Journal* 352 (2016) : 6704. Web. 17 April 2016.

Common Sense Nutrition Disclosure Act of 2015, H.R.2017. *Congress.gov*. Library of Congress.

    Web. 17 February 2016.

Conly, Sarah. *Against Autonomy: Justifying Coercive Paternalism.* Cambridge: Cambridge

    University Press, 2013. Print.

Coons, Christian and Michael Weber. *Paternalism: Theory and Practice.* Cambridge: Cambridge

    University Press, 2013. Print.

DeCastro JM. "Eating behavior: lessons from the real world of humans." *Ingestive Behavior and*

    Obesity 16 (2000) : 800-813. Web. 26 February 2016.

Dharmasena, S. and O. Capps Jr. "Intended and Unintended Consequences of a Proposed

    National Tax on Sugar-Sweetened Beverages to Combat the U.S. Obesity Problem."

    Health Economics 21.6 (2012) : 669-694. Web. 17 April 2016.

"Dietary Guidelines 2015-2020 Eight Edition: Executive Summary." *Health.gov.* Health.gov.

Web. 17 April 2016.

Dworkin, Gerald. "Defining Paternalism." *Paternalism: Theory and Practice*. Ed. Christian

Coons and Michael Weber. Cambridge: Cambridge University Press, 2013. Print.

Dworkin, Gerald. "Paternalism." *The Monist* 56.1 (1972) : 64-84. Print.

Dworkin, Gerald. "Paternalism." *The Stanford Encyclopedia of Philosophy.* Ed. Edward N. Zalta

(Summer 2014): n. p. Web. 20 November 2015.

Dworkin, Gerald. "Paternalism: Some Second Thoughts." *Paternalism*. Ed. Rolf Sartorius.

Minneapolis: University of Minnesota Press, 1983. Print.

Feinberg, Joel. "Legal Paternalism." *Paternalism.* Ed. Rolf Sartorius. Minneapolis: University of

Minnesota Press, 1983. Print.

Finkelstein, E.A., C. Zhen, M. Bilger, J. Nonnemaker, A.M. Farooqui, and J.E. Todd.

"Implications of a Sugar-Sweetened Beverage (SSB) Tax When Substitutions to Non-

Beverage Items are Considered." *Journal of Health Economics* 32.1 (2013) : 219-39.

Web. 17 April 2016.

"Food Labels." *Nhs.uk.* U.K. National Health Service, 13 July 2015. Web. 17 April 2016.

Gray, John N. "John Stuart Mill: Traditional and Revisionist Interpretations." *Library of

Economics and Liberty* (1979) n. p. Web. 20 April 2016.

"Have the Right to Know: Menu Labeling: Menus & Menu Boards with Nutrition Information."

Cspinet. Center for Science in the Public Interest, 2008. Web. 17 April 2016.

Halpern, David. *Inside the Nudge Unit: How Small Changes Can Make a Big Difference*.

London: WH Allen, 2015. Print.

"Information to Make Choices: Dietary Guidelines for Americans, Food and Menu Labeling,

and Marketing Standards." *Stateofobesity*. The State of Obesity, n. d. Web. 17 April

2016.

Kersh, Rogan. "Of Nannies and Nudges: the Current State of U.S. Obesity Policymaking." *The*

Royal Society for Public Health 129 (2015) : 1083-1091. Web. 12 December 2015.

Kliff, Sarah. "Denmark Scraps World's First Fat Tax." *WashingtonPost*. The Washington Post,

13 November 2012. Web. 18 April 2016.

Le Grand, Julian and Bill New. *Government Paternalism: Nanny State or Helpful Friend?*

Princeton: Princeton University Press, 2015. Print.

List, Justin. "To Donate or Not: Is That the Question?" *AMA Journal of Ethics* 7.9 (2005) : n. p.

Web. 19 April 2016.

Lobstein, Tim. "Sugar: a shove to industry rather than a nudge to consumers? *Clinical Key at*

New York Medical College 4 (2016) : 86-87. Web. 9 April 2016.

Loewenstein, G. "Out of control: visceral influences on behavior." *Organizational Behavior and*

Human Decision Processes 65 (1996) : 272-292. Web. 19 February 2016.

Long, Michael W., Deirdre K. Tobias, Angie L. Cradock, Holly Batchelder, and Steven L.

Gortmaker. "Systematic Review and Meta-Analysis of the Impact of Restaurant Menu

Calorie Labeling." *American Journal of Public Health* 105.5 (2015) : e11-e24. Web. 7

April 2016.

Main, Douglas. "Soda Tax in Mexico Causes Serious Sales Dip in Sugary Drinks." *Newsweek*.

Newsweek: Tech & Science, 6 January 2016. Web. 18 April 2016.

Martel, Frances. "Jon Stewart Rails Against Bloomberg's 'Draconian' Soda Ban With Piles of

Gross 'Legal' Food." *Mediaite*. Medialite, 2012. Web. 10 December 2015.

Matheson, Whitney. "Is your name on a Coke bottle? Find out now…" *USA Today*. USA Today,

    23 June 2014. Web. 27 February 2016.

Mill, John Stuart. *On Liberty*. Ed. Elizabeth Rapaport. Indianapolis: Hackett, 1978. Print.

Novak Nicole L. and Kelly D. Brownell. "Taxation as Prevention and as a Treatment for

    Obesity: The Case of Sugar-Sweetened Beverages." *Curr Pharm Des* 17.12 (2011) :

    1218-1222. Web. 14 April 2016.

"Nudges: Making the Healthy Choice the Easy Choice." *Healthyfoodbankhub*. Feeding America,

    n. d. Web. 17 April 2016.

 "Overview of FDA Labeling Requirements for Restaurants, Similar Retail Food Establishments

    and Vending Machines." *FDA*. U.S. Food and Drug Administration, 9 July 2015. Web.

    17 April 2016.

"Overweight & Obesity: Childhood Obesity Facts." *Cdc.gov*. Center for Disease Control and

    Prevention, 19 June 2015. Web. 17 April 2016.

 "Overweight and Obesity Statistics." *Niddk.nih.gov*. National Institutes of Diabetes and

    Digestive and Kidney Diseases, October 2012. Web. 17 April 2016.

Park, Alice. "The New York City Soda Ban, and a Brief History of Bloomberg's Nudges." *Time*

    Magazine. Time Magazine, 2012. Web. 21 March 2016.

"Proposed Changes to the Nutrition Facts Label." *FDA*. U.S. Food and Drug Administration, 23

    October 2015. Web. 17 April 2016.

Radnitz, Cynthia, Katharine Loeb, Julie DiMatteo, Kathleen Keller, Nancy Zucker, and Marlene

    Schwartz. "Optimal Defaults in the Prevention of Pediatric Obesity: From Platform to

    Practice." *J Food Nutr Disord* 2.5 (2013) : 1-16. Web. 14 April 2016.

Rebonato, Riccardo. 2012. *Taking Liberties: A Critical Examination of Libertarian Paternalism.*

Houndmills: Palgrave Macmillan. Print.

Samson, Alain. "An Introduction to Behavioral Economics." *The Behavioral Economics Guide*

2014. 1st ed. Ed. Alain Samson. 2014. Web. 14 April 2016.

"School Meals: Nutrition Standards for School Meals." *USDA*. United States Department of

Agriculture Food and Nutrition Service. August 2015. Web. 16 November 2016.

"State Sales Taxes on Regular Soda." *BridgingTheGapResearch.* Bridging the Gap, January

2014. Web. 18 April 2016.

Stoner, Lee and Jon Cornwall. "Did the American Medical Association make the correct

decision classifying obesity as a disease?" *Australas Med J* 7.11 (2014) : 462-464. Web.

14 April 2016.

"Study Predicts Tax on Sugar-Sweetened Beverages Would Not Result in Overall Job Loss."

RWJF. Robert Wood Johnson Foundation, 13 February 2014. Web. 20 November 2015.

Sturm, R., L.M. Powell, J.F. Chriqui, and F.J. Chaloupka. "Soda Taxes, Soft Drink

Consumption, and Children's Body Mass Index." *Health Affairs (Millwood)* 29.5 (2010) :

1052-1058. Web. 14 April 2016.

"Sugary Drinks and Obesity Fact Sheet." *Hsph.Harvard.edu.* Harvard T.H. Chan School of

Public Health: The Nutrition Source, n. d. Web. 18 April 2016.

"Summary of Polls on Nutrition Labeling in Restaurants." *Cspinet*. Center for Science in the

Public Interest, 2008. Web. 17 April 2016.

Thaler, Richard H. and Cass R. Sunstein. "Libertarian Paternalism." *The American Economic Review* 93.2 (2003) : 175-179. Print.

Thaler, Richard H. and Cass R. Sunstein. *Nudge: Improving Decisions About Health, Wealth, and Happiness*. New York: Penguin Books, 2008. Print.

Wadden, Thomas A. and Albert J. Stunkard. "Social and Psychological Consequences of Obesity." *Ann Intern Med* 103.6 (1985) : 1062-1067. Web. 17 April 2016.

Waldron, Jeremy. 2014. "It's All for Your Own Good." *Nybooks*. The New York Review of Books, 2014. Web. 29 November 2015.

Walley, Christine. "Searching for 'Voices': Feminism, Anthropology, and the Global Debate over Female Genital Operations." *Cultural Anthropology* 12 (1997) : 405-437. Web. 2 April 2016.

Wansink, Brian. *Mindless Eating: Why We Eat More Than We Think*. New York: Bantam Books, 2006. Print.

Wardle, Jane and Lucy Cooke. "The Impact of Obesity on Psychological Well-Being." *Endocrinology & Metabolism* 19.3 (2005) : 421-440. Web. 14 April 2016.

Weiner, Rachel. "The New York City soda ban explained." *The Washington Post*. The Washington Post, 2013. Web. 10 December 2015.

Winkler, Dan. "Persuasion and Coercion in Health: Ethical Issues in Government Efforts to Change Life Styles." *Paternalism*. Ed. Rolf Sartorius. Minneapolis: University of Minnesota Press, 1983. Print.

Wolf, Susan. "Sanity and the Metaphysics of Responsibility." *Responsibility, Character, and the Emotions: New Essays in Moral Psychology*. Ed. Ferdinand David Schoeman. Cambridge University Press, 1987. Print.

Wootan, Margo G. "Written Testimony of Margo G. Wootan, D.Sc., Before the Committee on

      Energy and Commerce, Subcommittee on Health." *Cspinet*. Center for Science in the

      Public Interest, 2004. Web. 17 April 2016.

Young, Angelo. "Soft Drink Sugar Tax: UK Joins Mexico With Levy on Sugary Beverages to

      Fight Obesity." *Business Insider*. International Business Times, 16 March 2016. Web. 18

      April 2016.