The Behavioral Welfare Paradox: Practical, Ethical and Welfare Implications of Nudging

David R. Just

With decades of behavioral economic research now achieving prominence, the last decade has seen the advent of behavioral policymaking. These efforts have been widely seen as successful in that they achieve policy goals without inducing backlash on the part of policy targets. Behavioral policies create a unique challenge to welfare analysis that has not been sufficiently addressed in the literature. The existence of behavioral effects creates a paradox, at once justifying the use of paternalistic policies and undermining the empirical foundations of welfare analysis. In this paper I explore the behavioral welfare paradox and its implications for economic policy prescription.

Decades after the introduction of behavioral economics as a subdiscipline, behavioral economics has come into its own as an applied policy tool. The UK was the first to make a serious national push to include behavioral economic considerations in policy when it launched the Behavioural Insights team within the Cabinet Office.[1] The United States followed suit, creating the White House Social and Behavioral Sciences Team (SBST), which was established by executive order in 2015.[2] The mission of each of these groups is to make government efforts more efficient by using behavioral nudges. Such nudges should not change the structure or function of government programs or policies, but simply make the programs function at a lower cost or higher yield through the use of behavioral interventions. With this mandate, behavioral economics became much more than a theoretical exercise. Now dozens of behavioral nudges have been implemented on a massive scale, affecting the lives of hundreds of millions.

David R. Just is professor in the Charles H. Dyson School of Applied Economics and Management at Cornell University. Correspondence*: David R. Just ▪ 210C Warren Hall ▪ Cornell University ▪ Ithaca, NY 14853 ▪ Phone 607.255.2086 ▪ Email: djust@cornell.edu.*

This article was originally prepared for the 2016 "Outstanding Public Service through Economics" award lecture held at the annual meeting of the Northeastern Agricultural and Resource Economics Association in Bar Harbor, Maine.

The views expressed are the author's and do not necessarily represent the policies or views of any sponsoring agencies.

[1] The group was privatized in 2013 and continues to partner with the UK government in policy analysis.

[2] Several other countries now have similar offices, including Germany and Mexico.

Agricultural and Resource Economics Review 46/1 (April 2017) 1–20

A comprehensive list of such interventions would be difficult to compile and would distract from the main points of this article. Many of these efforts have been summarized in the annual report of the SBST (2016), and the publications section of the Behavioural Insights Team website (BIT, 2017). This includes policies designed to ensure that children have access to free or reduced-price lunch by expanding automatic enrollment (Sunstein 2014, Ralston and Newman 2015), reframing information on how social security benefits may differ depending on when they are claimed (Brown, Kapteyn and Mitchell 2011), and tax forms that encourage greater redemption of tax credits (Lindsey 2010).

The success of behavioral approaches has come through the promise of improved results without backlash. Here the results are defined by the policy in question. In the case of enrollment in the school lunch program, better results are those in which more of those eligible for free or reduced-price school lunch are enrolled in the program. By definition, a behavioral nudge cannot alter the choice set but only alter the way in which the individual perceives that choice set. By not altering the choice set, the decision maker would have little reason to feel harmed by the nudge. If this is the case, we would expect that true nudges will create no backlash.

In a fundamental way, behavioral economic policy sets aside the long tradition of economics as a normative discipline. While much of economics is positive (being used to describe behavior), much of the work on policy analysis is normative in nature, seeking policies that maximize social welfare. Indeed, the foundations of behavioral economics are at complete and total odds with the foundations of welfare economics. There have been a few attempts to reconcile this rift (e.g., Bernheim and Rangel 2007, Kőszegi and Rabin 2007). Attempts to extend welfare theory to behavioral results must either set aside the empirical keystone of welfare theory – revealed preference – or they must limit revealed preference within the confines of a single behavioral treatment or context. Either of these routes raises significant questions about the validity of application as a policy tool.

In what follows I will discuss the implications of behavioral economics for welfare theory and also the implications of welfare theory for behavioral economics. I will argue that behavioral economics makes revealed preference arguments untenable in most situations in which behavioral policy can be implemented. For this reason, it is at once more important to make moral arguments for behavioral policy, and more important to recognize that policy is being made based solely on moral arguments.

Welfare Economics

The primary purpose of welfare economics is to evaluate the overall impact of economic policy on the wellbeing of the affected individuals. We generally use some set of assumptions about how to balance the tradeoffs in wellbeing of various actors in the economy. These assumptions are used to derive some

measure of overall welfare, which can be used to determine which policies will maximize welfare, or, more commonly, which will maximize welfare while imposing some other policy objective. For example, maximizing measures of welfare were used to justify broader free trade agreements in the 1980s and 1990s (e.g., Findlay and Wellisz 1982, Romer 1994). More often, there are political barriers to maximizing wellbeing. One group that loses due to the welfare maximizing policy may be politically powerful and prevent the feasibility of such a policy (Rausser 1982). Because maximizing welfare can often be infeasible, economists focus most often on maximizing welfare subject to either feasibility constraints (Downs 1957), or some policy objective, such as achieving a minimum threshold of use for biofuels (Lapan and Moschini 2012). Typically, several equivalent policies may be used to achieve the objective. In the case of biofuels, these could be taxes on traditional fuels, subsidies on biofuels, blending mandates, or caps on production of traditional fuels. Each has the possibility of achieving the objective of a minimum threshold of biofuel use, but the overall cost to consumers, producers, and taxpayers may be substantially different.

While individual contexts can differ in important ways, there are some general results from welfare theory that are important and have wide application. The two fundamental welfare theorems that relate competitive equilibrium to Pareto efficiency (Hicks 1939, Kaldor 1939) and the potential to achieve specific policy outcomes via competitive equilibrium (Hotelling 1938), are the most widely cited and adapted. Among the more prominent results is the principle of targeting (Bhagwati 1971) – that it is generally most efficient to achieve a policy goal using the most direct policy. Thus, if one is attempting to reduce emissions from manufacturing, taxing emissions results in greater welfare than taxing inputs or production during the manufacturing process. Taxing production, for example, can lead to production processes that produce more emissions but are cheaper for the relative quantities that prevail under the tax. While welfare theory generally considers all actors in a market or economy, for the purposes of this paper I will focus only on a consumer. Much of behavioral economics is concerned primarily with the behavior of individuals, and, in the policy context, how the individual's chosen behaviors can be influenced by context.

Consider the standard consumer problem, where the individual solves

$$\max_{x \in C} U(x|\Theta)$$

where $U(\cdot)$ is the utility function representing the individual's wellbeing as a single dimensional real valued function of their chosen consumption $x \in \mathbb{R}^n_+$ and the context in which that decision is made $\theta \in \Theta$. We will consider the context to represent all potential factors that could be manipulated by a behavioral nudge. The individual is restricted physically from selecting any choice outside of $C \subset \mathbb{R}^n_+$. This choice set is often just a representation of the

budget available and the prices of potential goods. In other cases, there may be physical limits placed on consumption by policy or external factors.

In terms of welfare, this model has simple implications for traditional policies. Traditional policies (those not engaging nudges) tend to change choice by changing the available choice set. If an economic policy makes the choice set C larger, then the policy could potentially make the individual better off by allowing superior choices. It is not possible that the individual could be made worse off by adding choices to C because the optimal choice from the smaller choice set is still available, and presumably the individual would reject any choices that are inferior to this outcome. On the other hand, if the policy reduces the choices available in C, then the individual is potentially worse off. Indeed, if the policy changes the individual's behavior by reducing C, then the individual must be at least weakly worse off. Of course, the policy may have potentially ambiguous effects if adding some choices while taking others away.

This simple theory of the consumer underlies much of how welfare engages policy on the consumer side. The key principle is that whenever the policy places a restriction on C, there must be some benefit given to other actors in the economy to justify the restriction. In other words, whenever we suggest a policy restricting consumer choice, a burden of proof must be met to ensure that the policy even has the potential to be welfare improving. This burden of proof is often met by arguing that the individual's choices result in externalities affecting the welfare of others who cannot directly influence the choice. In that case, we are reducing the choices available to one individual to enable a broader choice set for others. This is the common approach when we discuss environmental regulation or other policies where there is a clear set of actors affected by the individual's decision.

The simple notion of restricting an individual's choice only when it could have substantive benefits for others, however, is a mismatch for some prominent policy debates. Clear examples of the mismatch can be seen in many safety regulations. For example, consider the requirement to wear seatbelts in cars or helmets on motorcycles; while there is little doubt that these policies have saved lives (Cohen and Einav 2003), they have done so by restricting the choices of those whose lives have been saved. The economic impact on others has been relatively minor, but in some cases, negative. A 1968 federal statute required automobile manufacturers to provide seatbelts long before states considered requiring seatbelts be worn. New York was the first state to require the use of seatbelts in 1984. By 1995, every state except New Hampshire had some legal requirement to wear seatbelts while driving. Violations of these laws carry significant fines in many cases. While there are some minor externalities from failing to wear a seatbelt, this policy was argued primarily as a means to save individuals from their own poor decisions (e.g., Warner 1983). Indeed, some states contribute a portion of collected fines to a fund for those who suffer head injuries (presumably from

not wearing a seatbelt) – a transfer from those who were caught to those who were not – as a way of giving some of the money back to the poor souls who would drive without a seatbelt.

Many other policies appear to fall somewhere along this same spectrum of policies, where the clear target is to save the individual from their own poor choices, often with some minor impact on others. Several examples can be found in recent pushes for stronger nutrition policies. The most widely debated are sugar and fat taxes. Denmark made news for its rather large tax on fat (Stafford 2012), though it was repealed within a year. More recently, Mexico, Berkeley, CA, San Francisco, CA, and Boulder, CO, have all implemented per-ounce taxes on sugared beverages. While the impacts of these taxes have been debated (see Cawley and Frisvold 2015), their primary purpose was to reduce soda consumption as a means to reduce obesity and diabetes, primarily among poor populations (Schwarz et al. Forthcoming). Similar efforts have sought to eliminate the sale of sodas larger than 16 ounces from certain vendors (Grynbaum 2012), eliminate the inclusion of a toy in a child's meal that does not meet certain nutritional standards (Wiley 2013), or even change zoning laws to prohibit fast-food restaurants in poor neighborhoods (Sturm and Cohen 2009). While there are clearly some negative externalities from being overweight or contracting diabetes, the externalities are relatively minor compared to the direct impacts of many of these policies on store owners. This is perhaps less so when the public is responsible for the medical bills (McCormick and Stone 2007, Cawley and Meyerhoefer 2012) or when insurance pools are not allowed to consider weight status to determine rates (Bhattacharya and Sood 2007). Similar arguments could be made for the mandate provision in the Affordable Care Act (Koh and Sebelius 2010) requiring all individuals to purchase health insurance if they meet a certain income threshold. However, a strong case can be made that requiring healthy people to join the insurance pool lowers costs for everyone, enabling the general coverage of pre-existing conditions when starting or changing insurance carriers. Other policies in this realm of paternalism include safety regulations where goods have been banned or regulated because those who use them inappropriately can cause a danger to themselves, including pharmaceuticals and some cosmetics as well as recreational goods.

In each of these cases, economists working on policies have largely punted on welfare analysis, at least in terms of analyzing the impact of the measure on the individual decision-maker. For example, in health economics one often examines the impacts of policy in aggregate dollar terms while explicitly or implicitly ignoring whether the individual facing the impacts is the individual who is in control of the choices being abridged (Cawley and Meyerhoefer 2012). Indeed, in many of these policy debates it is predictably the individuals who are to be benefited who have resisted the most. If we consider again the policy model, the whole notion behind these policies is that it is somehow possible to improve the outcome for the individual by reducing the choice set C. While some lip service is given to impacts on third

parties, the primary arguments are directed at making the decision-maker better off for having fewer choices. We will refer to policies that have as their primary goal improving the welfare of the individual decision-maker by changing the decision-maker's selected consumption bundle to another that has been revealed to be inferior as being *paternalistic*.

Nudging

Thaler and Sunstein (2003) introduced the concept of libertarian paternalism to rationalize policy goals that appear to undermine the individual's own preferences. Libertarian paternalism is a philosophy that retains the essential element of welfare theory – that any restriction on choice must be compensated by benefits to others. However, paternalistic policies are still permitted so long as they can be achieved without changing the choice set, with behavioral interventions used to change the context and framing of the decision rather than the choice set itself. Richard Thaler and Cass Sunstein (2008, p. 6) argue that "Libertarian paternalism is a relatively weak, soft, and nonintrusive type of paternalism because choices are not blocked, fenced off, or significantly burdened." Rather, the choices are given a different frame or context that (at least in terms of selected outcome) aligns the preferences of the individual with the policymaker. In terms of the model of individual welfare, the policymaker now selects $\theta \in \Theta$, in hopes of improving the welfare of the individual, and thus solves

$$\max_{\theta \in \Theta} V(x^{*}(C, \theta))$$

where $V(\cdot, \cdot)$ represents the wellbeing of the individual, and $x^*(C, \theta)$ is the solution to the consumer problem presented earlier. This exercise leaves the individual in charge of their own selected outcome, and ensures that the outcome they select will be revealed as preferred to all other outcomes for at least some context parameter. Underlying this philosophy is the notion that we can identify contextual changes that do not directly affect welfare of the individual, but rather simply change how the individual processes the information regarding which choices are available and which choices are most desirable. In other words, the utility function that is maximized by the individual is no longer considered to contain all the information about the individual's resulting wellbeing. Instead, it contains the individual's perception of their utility function for a given context. Other contexts may change their judgment, and some contexts may result, more or less, in an accurate perception. For a moment, let us set aside the obvious question of how one determines the social welfare function; this will be addressed directly in subsequent sections.

There are a variety of different mechanisms through which a behavioral nudge may operate. However, there are perhaps four mechanisms most often

cited in the literature, with the most sizeable and predictable impacts on behavior. The most famous of behavioral results are based primarily on the framing of decisions. Changing the framing of the decision simply presents all information about the choice set using a different reference point. For example, you might go out of your way to save $4,000 on a new car, though a 10% discount may seem less exciting. Similarly, a 40% discount on a clock might sound more enticing than saving $10. If the car cost $40,000 and the clock cost $25, then the dollar and percentage discounts are identical in both cases. However, the statements frame identical discounts using different comparisons to make the discount sound larger or smaller (Chen, Monroe, and Lou 1998). Small-percentage discounts on a large-ticket item will sound larger when framed in total dollar amounts. Alternatively, large-percentage discounts on small-ticket items will sound unimpressive when framed in total dollar amounts. Framing can also highlight the costs or the benefits of an action, or offer different comparison groups, to lead one to change their evaluation of an outcome.

A second and related mechanism to influence choice is by taking account of "rule of thumb" decision-making. In many decision contexts, an individual is unlikely to put much thought or effort into determining their preferred outcome. Instead they will use some heuristic or rule of thumb that, at best, approximates their best choice. Food choice provides two very stark examples. Food consumers have a hard time estimating how much food they should eat at a meal, and many simply consume everything on their plate, then stop – this is often called the "clean plate club" – but plates can be vastly different sizes, leading people to consume substantially more when given larger plates than when given much smaller plates (Wansink and van Ittersum 2013). This particular nudge is closely related to reframing, as the food is compared to the size of the plate when determining what to eat. However, a second example demonstrates that this mechanism can be distinct from simple framing. Wansink and Hanks (2013) have shown that individuals are significantly more likely to choose the first item in a buffet line. This may be because individuals are displaying some sort of satisficing behavior rather than maximizing. Thus, if the first item in the line is good enough to meet the individual's desire, they will select it, without first seeing what they may be giving up later in the line. Rules of thumb used in different contexts are unique and can vary widely. Thus, making use of a rule of thumb to alter decisions may require intricate knowledge of the decision process.

A third behavioral mechanism is manipulating the perceived choice set. In many circumstances, the choice set is vast and would require substantial time to discern all possible choices. In such a context, highlighting a few of the potential choices can lead the individual to consider only the highlighted choices, or at least consider them more prominently (Just and Wansink 2009). Online retailers such as Amazon.com use this technique by highlighting specific items on their landing page, and grocery stores make use of this technique by placing featured items on endcaps where they are more

visible. In each case the individual may perceive a different choice set than is actually available based upon how the items are presented.

Finally, nudges can operate by suggesting social norms. In this case, rather than recasting what choices are available, the nudge can recast the social acceptability of choices. For example, individuals are more likely to choose a "regular"-sized food item than one labeled "double" even if they are the same size (Just and Wansink 2011). Similarly, utility bills that indicate when you have used more electricity than your neighbors can be effective at reducing overall use (Ferraro and Price 2013).

While there are myriad potential mechanisms through which nudges may operate, each has become policy relevant based on several claims, some of which are verifiable. Nudges have claimed to have sizable impacts on behavior that are unintuitive to those steeped in the rational choice doctrine. These big changes are claimed to be achieved without altering the choice set in any meaningful way. In other words, the individual's realized wellbeing for any outcome should not be adversely altered by the nudge, though the perception of that utility in relation to other choices might be. It is often claimed that nudges are so innocuous that the individual would not mind the nudge even if it were revealed (Jung and Mellers 2016). Presumably, each nudge is controllable, in that a policymaker can use it to guide individuals to choices that are preferred from the policymaker's perspective. Moreover, the nudge itself must be relatively inexpensive to implement so that there is little welfare loss in the process of implementation.

In thinking about nudges and welfare, consider the oft-cited example of nudging to encourage organ donation (Johnson and Goldstein 2003). Johnson and Goldstein noted that there were wide disparities in the organ donation rates of many countries that one may consider to be relatively similar culturally and developmentally. In Denmark only 5% of adults are organ donors while in Sweden 85% of adults are organ donors. The difference appears to be linked to the fact that in Denmark one must opt in to being an organ donor, while in Sweden one must opt out if they wish to avoid being an organ donor. Similar statistics play out in several other countries. More than 98% of adults are organ donors in Austria, Belgium, France, Hungary, Poland and Portugal, which all feature opt-out policies. This is an interesting case from a welfare perspective. In this case one may argue that there is little utility to the organ once the donor has passed (barring religious objections), while others may benefit substantially. Hence there may be a substantive argument from a traditional welfare policy perspective to subsidize or mandate organ-donor status. In this case, however, there appears to be a way to achieve the same target without altering the choice set. Allowing the choice to opt out of organ donation leaves the choice set intact, while also shifting the choice of individuals in line with an outcome that provides substantial benefits to others. The cost of implementing the opt-out policy appears not likely to cost more than the opt-in policy. Hence, it would be hard to argue that this nudge policy is not welfare-improving (though see Hansen 2012).

Other policies are potentially more difficult to analyze. Consider the Smarter Lunchrooms program that is now used in more than 30,000 schools nationwide (Gabirelyan et al. 2017). The key parts of this program make use of visibility and convenience (among other tools) to encourage school children to take and eat more fruits and vegetables as part of their school meal (see Just and Wansink 2009). The approach is relatively low in cost, and has been shown to increase intake of fruits and vegetables by sizable percentages (Hanks, Just, and Wansink 2013). Moreover, the approach does not alter the offerings on the school lunch line, and therefore does not directly affect the choice set. On the other hand, fruits and vegetables themselves cost money. Therefore, to make the welfare case, we would need to know that somehow either the children or someone else was benefiting from this program to a greater degree than the added cost of produce. Welfare theory gives us very little to work with in this regard. Before Smarter Lunchrooms is introduced, some children choose not to eat the produce, while, after introduction, they do. It is not possible to directly compare the children's preference in the two cases. It may be possible to say something about the preferences of school or federal officials or of the parents of the children. However, such preferences are (rightly or wrongly) usually ignored in welfare analysis.

Even without a clear welfare argument, the context of the Smarter Lunchrooms approach provides some welfare justification. Federal regulations since 2010 have required that each reimbursable meal include a fruit or vegetable. This requirement was costly (Newman 2012), increasing the cost of the meal by approximately $0.17 per tray (Just and Price 2013). However, simply providing the fruits and vegetables does not motivate children to eat them, and hence substantial amounts of this expenditure can be wasted. Smarter Lunchrooms approaches can improve the effectiveness of this policy with minimal additional cost – a sort of behavioral second-best approach.

An even more dubious case can be made for a nudge explored in Just and Wansink (2014). They found that individuals eat less when a food item is given a name that sounds big (like "double-size"). In addition, individuals are willing to pay more for an item that sounds larger, so one could imagine a policy in which larger names are given to food items without changing the choice set of available items. This policy may encourage individuals to eat less, though again it is difficult to use any standard version of welfare theory to demonstrate that the individual would be better off for this reduction. On the other hand, if such a policy encourages food producers to raise prices, it may improve the profits of producers at the consumers' expense. Consumers may pay more for the same products and eat less of them. Under the constraint of finding a policy to encourage the consumption of less food, there may be some way to argue a welfare improvement over more directly paternalistic policies. However, without such a constraint, it is difficult to see how a clear argument could be made to justify such a policy using traditional welfare theory.

Behavioral Choice and Empirical Welfare

Behind each of the proposed or implemented nudge policies is substantial behavioral literature exploring why choices may be influenced in ways that do not alter the choice set. While the reasons posited for the behavioral observations vary widely, the common theme is that individuals do not appear to make choices by optimizing in a way that is approximated by standard economic models in the particular applications explored. Rather, individuals appear to value some things that we would normally consider irrelevant, or to base choices on arbitrary or random factors that should seemingly have no impact. While observing such behavior is enough to tell us that behavioral decisions do not resemble our simple models, we still have very little idea what drives decisions in any general sense. In contrast to the standard economic approach of finding a simple model of choice based on first principles, the behavioral approach is piecemeal. This piecemeal approach is difficult to reconcile, particularly when it comes to broad constructs such as welfare theory, anchored in basic assumptions about choice.

To see this point, consider the economic arguments against paternalism. The entire argument is that eliminating preferred options from the choice set will make the individual worse off by leaving them with a choice that is inferior. This rationale assumes that the policy itself will leave the individual preferences unchanged. Several behavioral studies demonstrate an additional effect that is ignored by this policy approach. The behavioral literature has long known of reactance, which is defined as any sort of rebellion against a threat to one's freedom. The early work on reactance examined the response of college students to signs prohibiting graffiti in bathrooms (Pennebaker and Sanders 1976). When the signs were authoritarian, proclaiming stiff penalties, the researcher observed an increase in graffiti. Alternatively, when the signs were more persuasive and less authoritarian, graffiti decreased. The key problem that reactance and other emotional responses to policy pose for economists is the prospect that preferences are endogenous to policy (Just and Hanks 2015). If this is the case, then we face some fundamental problems with pursuing the use of empirical welfare economics to form policy.

Empirical welfare economics is based on two fundamental assumptions. The first is that choice is a result of preferences that at once represent wellbeing and behavior. Behavioral economics literature takes its departure from the observation that people do not truly maximize a utility function that both measures their wellbeing and guides all of their choices. Instead, there are several indications that individuals choose to do things that are not in their own best interest. For example, individuals often pay substantially more for services than required (Della Vigna and Malmendier 2006). Such choices seem to unambiguously leave the individual worse off than they could be.

Behavioral Axiom 1: *Selection by an individual does not imply that a consumption bundle results in greater wellbeing than any other bundle in the choice set.*

Behavioral Axiom 1 leads directly to proposition 1.

Proposition 1: *If Behavioral Axiom 1 holds, then it is feasible to improve welfare by restricting or altering individual choice.*

Second, welfare theory is based on the premise that preferences are stable. Under this premise, while we may find an individual at a different point on their utility surface when consumption bundles are changed, the shape of their utility surface does not change. Kahneman, Knetsch and Thaler (1991) show a simple example of how this fails intuitively in their famous mug experiment. They find that randomly assigning one to own a mug increases the mug's value to the individual, demonstrating that changing the bundle can change preferences.

Behavioral Axiom 2: *Preferences are not invariant to consumption bundles or context.*

Accepting Behavioral Axioms 1 and 2 creates an interesting paradox in behavioral welfare theory.

Consider Figure 1, which depicts the standard conceptualization of a paternalistic intervention at the individual level. Prior to the paternalistic intervention, the individual obtains utility u_0^* from choosing consumption bundle x_0^* (trivialized in the figure to a single dimension) where the choice set $C_0 = \left[0, x_0^*\right]$. An individual meeting the standard assumptions of welfare theory would always choose the option that gives them the greatest wellbeing. A paternalistic policy that restricts choice in this case might change the choice set to $C_1 = \left[0, x_1^*\right]$, inducing a choice of x_1^* and a utility of u_1^*. Unambiguously, imposing this restriction on choice must lead to reduced wellbeing if it alters choice, and if the utility function is strictly monotonic. But, if this is the case, then paternalistic policies could not be justified. The individual could not be made better off by restricting their choice if they always choose the option that yields the greatest wellbeing. While such has been the refrain from neoclassical economists for decades, this has not yet deterred policymakers or activists.

The behavioral literature provides some justification for paternalism in general through Behavioral Axiom 1. If individuals make mistakes, and if those mistakes are systematic, it may be possible to make the individual better off by restricting their choices and putting attractive bad choices off limits. However, very few would be bold enough to suggest that revealed preference has nothing to do with wellbeing, just that there are some systematic

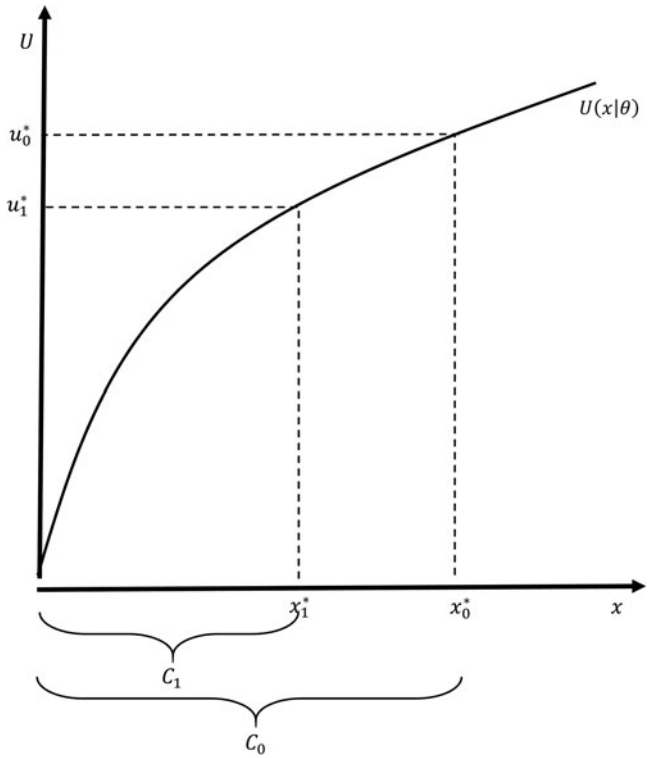

Figure 1. The Impact of Paternalistic Interventions on Individual Utility

errors and mistakes we may be able to address. In this case, we may face a situation much like that in Figure 2. Here, when left on their own and given the choice set $C_0 = [0,\ x_0]$, the individual chooses x_0 even when there are other options that would yield a higher level of wellbeing. This may be thought of as representative of behavioral problems such as overeating, or engaging in destructive behaviors. In some cases the individual may be aware that their behavior is limiting their wellbeing (for example, if according to Behavioral Axiom 2, destructive preferences are understood by the decision maker to be transient), while in others they may not be. In either case, now a paternalistic policymaker could impose the restriction on the choice set, $C_1 = \left[0, x_1^*\right]$ hopefully leading the individual to now select the wellbeing maximizing bundle x_1^*.

Justifying this paternalistic intervention, however, requires us to maintain two important assumptions. First, the policymaker must be able to identify which consumption bundles will improve welfare relative to what the individual would select on their own. Let us set this assumption aside for a moment. Secondly, we must also assume that the policymaker can formulate a choice set that will induce selection of such a welfare-improving bundle.

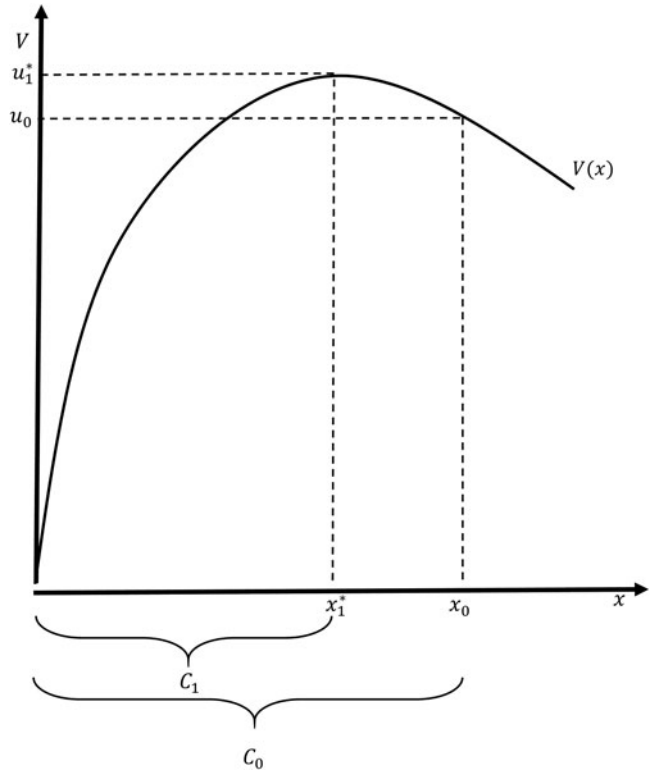

Figure 2. The Impact of Paternalistic Policies in a Behavioral Model

Given that the individual has a demonstrated inclination to choose consumption bundles that do not maximize wellbeing, this may not be a trivial assumption. If, for example, restricting the choice set as described now induces the individual to select $\tilde{x}_1$, then the paternalistic intervention would fail. This means that if we are to use behavioral results as a justification for traditionally paternalistic interventions, then we must have a thorough understanding of the behavior to claim that we can induce welfare-improving choices by restricting the choice set.

For example, if we wish to reduce children's sugar consumption in school, we may think that eliminating chocolate milk in school may be one way to achieve this goal. However, if the policy induces some children to abandon all milk (Hank, Just, and Wansink 2014), smuggle soda into school (Moore 2012), or overcompensate with greater consumption of chocolate milk after school, the policy could be self-defeating. The potential for reactance or other compensating behaviors that are outside the control of the policymaker may make traditional paternalistic policies particularly ripe for such miscalculations.

Proposition 2: *Given Behavioral Axiom 2, policymakers will not always be able to accurately predict a priori how individual decision-makers will respond to potential policies using standard economic theory.*

Proposition 2 is important in that it underlines the importance of measuring the actual impact of policies. Whether policies rely on behavioral nudges or not, responses may be behavioral and could be sizable.

Consider Figure 3, which depicts the potential effect of a libertarian paternalistic policy. As in Figure 2, when left on their own and given the choice set $C_0 = [0, x_0]$, the individual chooses x_0 even when there are other options that would yield a higher level of wellbeing. In this case, however, the policymaker leaves the original choice set intact, and uses a contextual nudge that leads the individual to choose x_0^* leading to a gain in individual wellbeing. As with the paternalistic intervention, this approach requires some assumptions to justify intervention. We must be able to find a behavioral intervention that leads individuals to the desired consumption bundle. This is, in fact, the primary purpose of policy-focused behavioral research.

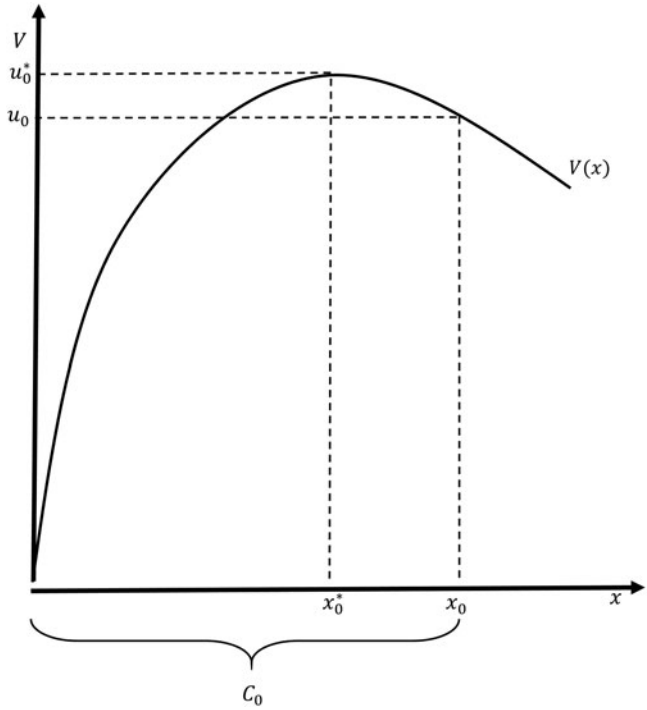

Figure 3. The Impact of Libertarian Paternalistic Policies on Individual Wellbeing

As with paternalistic interventions, we must also be able to identify which consumption bundles will lead to welfare improvements. The case is often made that because the choice set is not altered, we are perhaps not under the same burden (or same extent of burden) to demonstrate that the outcome is in fact better for the individual. Nonetheless, if the issue is of enough importance to require policy intervention, it is of importance to demonstrate that the intervention will in the least be beneficial. There is the potential that instead of leading one from a suboptimal point x_0 to x_0^*, we may be leading them from the optimal to some suboptimal choice. Much more likely, we may be leading the individual from one suboptimal choice to another suboptimal choice and it may not always be clear how to determine if it is an improvement. Returning to our milk example, we may consider a nudge that encourages children to select white milk over chocolate milk even when both are available (perhaps by placing white milk where it is more visible or convenient). These results could help reduce sugar consumption, unless selecting the white milk triggers the child to purchase cookies more often with their lunch. Even if we knew that reducing sugar consumption was welfare improving, operationalizing such policy in a nudge requires significant knowledge that the nudge will induce a bundle that improves overall welfare. Demonstrating such becomes much more complicated under a behavioral theory.

Proposition 3 (Behavioral Welfare Paradox): If Behavioral Axiom 1 and 2 hold, then revealed preference cannot be used as a measure of welfare for policy purposes.

Together, Propositions 1 and 3 yield the paradox of behavioral welfare theory. Welfare theory was built on the notion that we could learn what was better for an individual by observing their choices. If a bundle is chosen, then it must be at least weakly better for the individual than the other choices that were passed up. If we admit that selection does not imply the option is at least weakly better than other options that were available (Behavioral Axiom 1), then we no longer have any empirical basis to rely upon when determining what makes an individual better off (Behavioral Welfare Paradox). Behavioral theory at once admits that paternalistic interventions may make the individual better off (Proposition 1) and eliminate our ability to empirically justify which outcomes yield the best overall wellbeing. Indeed behavioral economics undermines the entire notion of welfare economics at the consumer level – that we can find externally verifiable measures of wellbeing that are useful in creating policy.

The Need for Behavioral Welfare Theory

With the growth of behavioral economics in policy, there has been a slow rise of opposition to nudge style policies. The opposition comes not just from

professional opinion writers, but also from academic economists and even from some prominent behavioral economists. In 2010 George Loewenstein and Peter Ubel wrote about the need to temper our expectations regarding the power of nudges. His essential critique is that if the change in behavior is truly meaningful, then the individual probably has strong preferences over the behavior, and a nudge is unlikely to be effective. While he holds out hope that some nudges will be important from a policy perspective, he throws some cold water on the unbridled enthusiasm that has been rampant among researchers and policymakers alike. This critique is obviously valid, and underscores the critical need for thorough study of the impacts of behavioral policies. Traditional policies are often rolled out with very little thought to measuring impacts. Such an approach with behavioral policy could very easily lead to snake-oil-type approaches that make policymakers feel good with little actual improvement. Indeed, measurement was one of the key missions of both Behavioural Insights and the SBST. Some of these agencies' greatest accomplishments to date have been demonstrations that some of the effects found in the literature do not actually manifest themselves when implemented, or only manifest themselves when policies are implemented in very specific ways.

A second critique has been championed by Cass Sunstein (Forthcoming), who argued that not all nudges are acceptable to the individual being nudged. Indeed, some appear to be onerous and would be rejected by the average individual if given the choice. He finds that in general, people prefer nudges to directly paternalistic policies. However, people are concerned by nudges that may lead to undesired outcomes through inattention, or nudges that address issues with which they disagree. One example Sunstein uses in his national survey is a question asking if individuals would like a default policy that all are enrolled in the Republican party unless they choose to opt out. Such a policy is clearly undesirable to those who oppose the Republican party. But even less-biased nudges can cause pushback or offense. In general, individuals appear to prefer nudges that are likely to lead to greater thought or care in making a choice, sometimes called behavioral interrupts (Just and Gabrielyan 2016), to nudges that operate on a subconscious level. While Sunstein finds little evidence that individuals have rebelled against nudges that have been used in practice, more evidence is needed. An under-addressed issue is whether nudges can create reactance when they are revealed to the individual who is subject to them. While the example of a default political affiliation is extreme, it demonstrates the potential for harm while not technically altering the choice set.

Welfare theory has generally relied on very strict assumptions to support policy. It is clear that such strict assumptions cannot hold, and without these strict assumptions we must also part with concrete and indisputable evidence of improvement. Welfare theory has lost its empirical moorings. Sunstein's approach offers one way forward, through surveying potential subjects of the nudge to determine their preference regarding whether they

would want the nudge or not. Further experimentation could help to ensure that nudges do not induce reactance once they are revealed to the target audience. Alternative theories and assumptions could also help to restore the usefulness of welfare economics in the context of a behavioral policy. For example, we may be willing to impose policies that reduce future regret. This would be consistent with policies designed to help alleviate self-control issues. Theoretically, this would require by axiom that individuals in the future hold some privileged position and are better able to judge welfare than earlier versions of the same individual.

Perhaps the most defensible way forward, however, is simply to argue that policy goals be determined based upon moral arguments rather than empirically identifiable measures of wellbeing. One aversion to such an approach is the acknowledged fact that we cannot all agree on what is moral. From the perspective of academic economists, we could focus on achieving specified goals at minimum cost or with minimum impact on choice, without the need to engage the moral arguments directly. On the other hand, acknowledging that policy objectives are based upon moral arguments could do much to help create an honest dialogue, both in political and academic realms. The notion that the right policy objectives are somehow empirically verifiable can create an atmosphere where instead of persuading others of their motives, policymakers simply cite studies and claim there is no other way to look at the issue. At the same time, the notion that we can empirically verify good policy objectives creates significant motives for academics in economics, as well as many other fields, to engineer study outcomes that can substitute for moral arguments.

Finally, it is important to acknowledge that not every potential context for nudging is equal. While there is no need to place any political party as the default party, some food item must appear first on the lunch line. Thus, in some cases we have no choice as to whether the government will nudge, but only of which nudge they will implement. This perhaps places a somewhat lesser burden on the policymaker to demonstrate substantive benefits for their implemented nudge.

Conclusion

While several have made attempts to extend welfare theory to the realm of behavioral policy (e.g., Bernheim and Rangel 2007, Kozegi and Rabin 2007). These attempts fail to bridge what is the most important and persuasive gap – that of replacing empirical welfare economics. When we admit behavioral models that separate choice from wellbeing, we are left with a paradox. At once, we find a general justification for paternalistic policies, but at the same time welfare economics becomes unmoored from empirical measures of wellbeing. We can justify intervention, but cannot argue empirically which outcomes are more desired. In this paper I have outlined the reasons why such a conundrum is important to address, and offered some very minor

attempts at directions we may look for solutions. In the meantime, it is unlikely that the pace of policy needs will slow down or cease, and it is important for us to make whatever attempts we can to address the behavioral welfare issue as we compare both traditional and behavioral policies.

We are embarking on a complicated and nuanced discussion about issues over which individuals may hold strong feelings and for which no current empirical approach will suffice. Perhaps the rudiments of welfare theory we have taught (and sold many policymakers on) oversimplified the impacts of economic policies. New tools are necessary to help sort out the complicated questions arising from policies we are already beginning to employ, as well as the behavioral implications of paternalistic policies we have dabbled with for more than a century.

References

Bhattacharya, J., and N. Sood. 2007. "Health Insurance and the Obesity Externality." *Advances in Health Economics and Health Services Research* 17(1) : 279–318.

Bernheim, B.D., and A. Rangel. 2007. "Toward Choice-theoretic Foundations for Behavioral Welfare Economics." *American Economic Review* 97(2): 464–470.

Bhagwati, J.N. 1971. "The Generalized Theory of Distortions and Welfare." In: Bhagwati, J.N. and C.P. Kindleberger, eds., *Trade, Balance of Payments, and Growth*. Amsterdam: North-Holland.

BIT. 2017. Publications, The Behavioral Insights Team, UK. Available at http://www.behaviouralinsights.co.uk/publications/ (Accessed January 11, 2017).

Brown, J.R., Kapteyn, A. and Mitchell, O.S. 2011. "Framing Effects and Expected Social Security Claiming Behavior" NBER Working Paper No. 17018, National Bureau of Economic Research, Cambridge, MA.

Cawley, J. and D. Frisvold. 2015. "The Incidence of Taxes on Sugar-Sweetened Beverages: The Case of Berkeley, California." NBER Working Paper No. 21465, National Bureau of Economic Research, Cambridge, MA.

Cawley, J. and C. Meyerhoefer. 2012. "The Medical Care Costs of Obesity: an Instrumental Variables Approach." *Journal of Health Economics* 31(1): 219–230.

Chen, S.F. S., K.B. Monroe and Y.C. Lou. 1998. "The Effects of Framing Price Promotion Messages on Consumers' Perceptions and Purchase Intentions." *Journal of Retailing* 74 (3): 353–372.

Cohen, A. and L. Einav. 2003. "The Effects of Mandatory Seat Belt Laws on Driving Behavior and Traffic Fatalities." *Review of Economics and Statistics* 85(4): 828–843.

Della Vigna, S. and U. Malmendier. 2006. "Paying Not to Go to the Gym." *American Economic Review* 96(3): 694–719.

Downs, A. 1957. "An Economic Theory of Political Action in a Democracy." *Journal of Political Economy* 65(2): 135–150.

Ferraro, P.J. and M.K. Price. 2013. "Using Nonpecuniary Strategies to Influence Behavior: Evidence from a Large-scale Field Experiment." *Review of Economics and Statistics* 95 (1):64–73.

Findlay, R. and Wellisz, S. 1982. "Endogenous Tariffs, the Political Economy of Trade Restrictions, and Welfare." In Bhagwati, J.N.,ed., *Import Competition and Response*, 223–244. Chicago, IL: University of Chicago Press., pp. 223–244.

Gabrielyan, G., D.S. Hanks, K. Hoy, D.R. Just and B. Wansink. 2017. "Who's Adopting the Smarter Lunchroom Approach? Individual Characteristics of Innovative Food Service Directors." *Evaluation and Program Planning* 60: 72–80.

Grynbaum, M. M. 2012. "Will Soda Restrictions Help New York Win the War on Obesity?" *BMJ* 10(345): e6768.
Hanks, A.S., D.R. Just and B. Wansink. 2013. "Smarter Lunchrooms Can Address New School Lunchroom Guidelines and Childhood Obesity." *Journal of Pediatrics* 162(4): 867–869.
Hanks, A.S., D.R. Just, and B. Wansink. 2014. "Chocolate Milk Consequences: A Pilot Study Evaluating the Consequences of Banning Chocolate Milk in School Cafeterias." *PloS One* 9(4): e91022.
Hansen, P. G. 2012. "Should We Be 'Nudging' for Cadaveric Organ Donations?" *American Journal of Bioethics* 12(2): 46–48.
Hicks, J. R. 1939. "The Foundation of Welfare Economics" *Economic Journal* 49(196): 696–712.
Hotelling, H. 1938. "The General Welfare in Relation to Problems of Taxation and of Railway and Utility Rates." *Econometrica* 6(3): 242–269.
Johnson, E.J. and D. Goldstein. 2003. "Do Defaults Save Lives?" *Science* 302(5649):1338–1339.
Jung, J.Y. and B.A. Mellers. 2016. "American Attitudes Toward Nudges." *Judgment and Decision Making* 11(1): 62–74.
Just, D.R., and G. Gabrielyan. 2016. "Why Behavioral Economics Matters to Global Food Policy." *Global Food Security* 11(1):26–33.
Just, D.R. and A.S. Hanks. 2015. "The Hidden Cost of Regulation: Emotional Responses to Command and Control." *American Journal of Agricultural Economics* 97(5):1385–1399.
Just, D.R. and J. Price. 2013. "Default Options, Incentives and Food Choices: Evidence from Elementary-school Children." *Public Health Nutrition* 16(12): 2281–2288.
Just, D. R. and B. Wansink. 2009. "Better School Meals on a Budget: Using Behavioral Economics and Food Psychology to Improve Meal Selection," *Choices* 24(3): 1–7.
———. 2011. "The Flat-rate Pricing Paradox: Conflicting Effects of 'All-You-Can-Eat' Buffet Pricing." *Review of Economics and Statistics* 93(1):193–200.
Just, D.R., and B. Wansink. 2014. "One Man's Tall is Another Man's Small: How the Framing of Portion Size Influences Food Choice." *Health Economics* 23(7): 776–791.
Kaldor, N. 1939. "Welfare Propositions of Economics and Interpersonal Comparisons of Utility" *Economic Journal* 49(195):549–552.
Kahneman, D., J.L. Knetsch, and R.H. Thaler. 1991. "Anomalies: The Endowment Effect, Loss Aversion, and Status Quo Bias." *Journal of Economic Perspectives* 5(1): 193–206.
Koh, H.K., and K.G. Sebelius. 2010. "Promoting Prevention through the Affordable Care Act." *New England Journal of Medicine* 363(14): 1296–1299.
Kőszegi, B. and M. Rabin. 2007. "Mistakes in Choice-based Welfare Analysis." *American Economic Review* 97(2): 477–481.
Lapan, H. and G. Moschini. 2012. "Second-best Biofuel Policies and the Welfare Effects of Quantity Mandates and Subsidies." *Journal of Environmental Economics and Management* 63 (2): 224–241.
Lindsey, V. W. 2010. "Encouraging Savings Under the Earned Income Tax Credit: A Nudge in the Right Direction." *University of Michigan Journal of Law Reform* 44(1): 83.
Lowenstein, G. and P. Ubel. 2010. "Economics Behaving Badly." *New York Times*, July 14. Available at http://www.nytimes.com/2010/07/15/opinion/15loewenstein.html (Accessed January 11, 2017).
McCormick, B. and I. Stone. 2007. "Economic Costs of Obesity and the Case for Government Intervention." *Obesity Reviews* 8(s1): 161–164.
Moore, G. 2012. "Feds' Flavored Milk Ban Spurs High-school Black Market in Chocolate Syrup." *Boston Business Journal*, September 13. http://www.bizjournals.com/boston/blog/mass_roundup/2012/09/chocolate-syrup-in-school.html (Accessed January 11, 2017).
Newman, C. 2012. "The Food Costs of Healthier School Lunches." *Agricultural and Resource Economics Review* 41(1): 12–28.

Pennebaker, J.W. and D.Y. Sanders. 1976. "American Graffiti: Effects of Authority and Reactance Arousal." *Personality and Social Psychology Bulletin* 2(3): 264–267.
Ralston, K. and C. Newman. 2015. "School Meals in Transition" *Economic Information Bulletin* No. 143, Economic Research Service, U.S. Department of Agriculture, Washington, D.C.. Available at https://www.ers.usda.gov/webdocs/publications/eib143/53570_eib143.pdf?v=42236 (accessed February 2017).
Rausser, G.C. 1982. "Political Economic Markets: PERTs and PESTs in Food and Agriculture." *American Journal of Agricultural Economics* 64(5): 821–833.
Romer, P. 1994. "New Goods, Old Theory, and the Welfare Costs of Trade Restrictions." *Journal of Development Economics* 43(1): 5–38.
Schwarz, M., D.R. Just, A. Ammerman and J. Chriqui. Forthcoming. "Appetite Self-regulation: Environmental and Policy Influences on Eating Behaviors." *Obesity*.
Stafford, N. 2012. "Denmark Cancels 'Fat Tax' and Shelves Sugar Tax Because of the Threat of Job Losses." *BMJ* 10(345): e7889.
Sturm, R. and D.A. Cohen. 2009. "Zoning for Health? The Year-old Ban on New Fast-food Restaurants in South LA." *Health Affairs* 28(6): w1088–w1097.
Sunstein, C.R. 2014. "Nudging: a Very Short Guide" *Journal of Consumer Policy* 37(4): 583–588.
——— Forthcoming. "Do People Like Nudges?" *Administrative Law Review*.
SBST. 2016. "Social and Behavioral Sciences Team Annual Report." Executive Office of the President, National Science and Technology Council, Washington, D.C..
Thaler, R.H., and C.R. Sunstein. 2003. "Libertarian Paternalism." *American Economic Review* 93 (2): 175–179.
Thaler, R.H. and C.R. Sunstein. 2008. *Nudge: Improving Decisions about Health Wealth and Happiness*. New Haven, CT: Yale University Press.
Wansink, B. and A.S. Hanks. 2013. "Slim by Design: Serving Healthy Foods First in Buffet Lines Improves Overall Meal Selection." *PloS One* 8(10): e77055.
Wansink, B. and K. van Ittersum. "Portion Size Me: Plate-size Induced Consumption Norms and Win-win Solutions for Reducing Food Intake and Waste." *Journal of Experimental Psychology: Applied* 19(4): 320.
Warner, K.E. 1983. "Bags, Buckles, and Belts: The Debate over Mandatory Passive Restraints in Automobiles." *Journal of Health Politics, Policy and Law* 8(1): 44–75.
Wiley, L.F. 2013. "Sugary Drinks, Happy Meals, Social Norms, and the Law." *Connecticut Law Review* 46(5): 1877.