THE ETHICS OF PUBLIC HEALTH NUDGES

A Dissertation
Submitted to the Faculty of the
Graduate School of Arts and Sciences
Of Georgetown University
in partial fulfillment of the requirements for the degree of
Doctor of Philosophy
in Philosophy

By

Yashar Saghai, M.A.

Washington, D.C.
December 10, 2012
THE ETHICS OF PUBLIC HEALTH NUDGES

Yashar Saghai, M.A.

Thesis co-advisors: Margaret Olivia Little, Ph.D.

Madison Powers, J.D., D.Phil.

ABSTRACT

There is growing interest in using non-coercive interventions to promote and protect public health, in particular “health nudges.” Behavioral economist Richard Thaler and law scholar Cass Sunstein coined the term nudge to designate influences that steer individuals in a predetermined direction by activating their automatic cognitive processes, while preserving their freedom of choice. Proponents of nudges argue that public and private institutions are entitled to use health-promoting nudges because nudges do not close off any options. Opponents reply that the nudgee has no opportunity to deliberate over her choice. The nudger controls the nudgee, who has no real freedom of choice.

In my dissertation, I salvage the concept of nudge from the charge that it merely pays lip service to freedom of choice, offer an alternative account of the moral import of nudges for the evaluation of public health policies, and provide an ethical framework for their justified use.

My argument proceeds in four steps. First, I argue that nudging mechanisms often involve some form of incomplete deliberation, and do not always bypass deliberation. Second, I maintain that nudges preserve freedom of choice because they preserve the
choice-set and are substantially noncontrolling. I show that the debate over nudges is plagued by confusion between real nudges, which are easily resistible and therefore substantially noncontrolling, and influences that are not easily resistible yet activate the same cognitive mechanisms as nudges. Third, I reject the view that nudges are systematically preferable to more controlling influences, even when they interfere with non-consequential liberties. I show that the principle of the least restrictive alternative often invoked in public health ethics is applied with too wide a scope. Fourth, I argue that paternalism is not central to the ethics of public health nudges. Rather, the substantive problem with nudges is that their overall cumulative effect might weaken our decisional capacities or undermine their proper exercise. This is a reason to limit recourse to certain nudges. I finally show, contra certain Kantians and Millians, that we have no perfect duty to engage or enhance decisional capacities.
ACKNOWLEDGEMENTS

I would first like to warmly thank my committee members who have provided me with tremendous support over the past several years, much beyond my wildest dreams.

Let me start with my co-advisors, Maggie Little and Madison Powers, who encouraged me to find my own voice and defend my views. Maggie provided insightful comments on a countless number of drafts, and spent many hours discussing these ideas with me. Her constructive criticisms helped and challenged me to tighten my arguments. John Cage's motto, "Begin again!" is truly hers. Her support and availability has been invaluable. If my writing style is most of the time closer to the standards of clarity of analytic philosophy than to a Google Translate rendition of Sartre's *Critique de la raison dialectique*, it is largely thanks to her.

Madison's mentorship is one's those unexpected gifts that life sometimes offers you. He opened my eyes to a number of issues, and is no doubt the source of my deep interest in public health ethics. I try to emulate his combination of philosophical sophistication and attention to the broader picture. Not only did he guide me out of the maze of my research project, but he also offered me to co-write an article with him and Ruth Faden. He placed a lot of trust in me, and I did my best not to be unworthy of it. Madison and Ruth both enabled me to experience at an early stage the life of a professional scholar. Collaborating with them was an honor and a daily joy. I am also appreciative that they agreed that I include our co-written article as part of this dissertation. I hope I can someday accommodate my students half as well as Maggie, Ruth, and Madison did.
I would also like to thank my readers, Bob Veatch, Peter Ubel, and Daniel Andler. Bob's generosity, intense love for ideas, and characteristic toleration for dissent taught me a lot. Peter gave me very helpful feedback on bioethical issues, and naturally, on my use of empirical literature on the psychology of medical decision-making, a field of research he contributed to shape. He made sure I refrained from philosophizing 20,000 feet off the ground. I will always owe a debt of gratitude to Daniel who nurtured my early interest for the philosophy of cognitive science and the ethical ramifications of the literature on rationality. He believed in this project from day one and lent it unwavering support.

Philosophy is essentially a dialogical endeavor and this project, perhaps more than most dissertations, has benefited from the feedback of a number of people. My ideas got sharpened thanks to written comments by Richard Ashcroft, Luc Bovens, Jerry Dworkin, Kalle Grill, Edmund Pellegrino, Alan Wertheimer, Brynn Welch, and anonymous referees of *Public Health Ethics* and the *Journal of Medical Ethics*. I also thank the editors of those journals, Angus Dawson and Russell Powell.

Over the years, I have benefited from conversations with many scholars, including David Armstrong, Paul Dolan, Bennett Foddy, Agomoni Ganguli-Mitra, Avner Offer, Marianne Promberger, Gérard Réach, Gina Rini, Simon Rippon, and Jo Wolff. The generosity of many academics is astonishing. The feedback I got from the audiences of conferences at which I presented my works-in-progress helped me think through the ethics of nudges from multiple perspectives.

My deepest gratitude goes to the French professors who encouraged me to apply to a U.S. doctoral program and supported my desire to get acquainted with analytic philosophy and English-speaking bioethics: Olivier Abel, Pierre-François Moreau,
Suzanne Rameix, and André Tosel.

My friends Philip Clark, Lydia Durand, Marshall Gagne, Sasha Lezhnev, and Jennifer Luff gave me precious proofreading and editing assistance. Marshall, in particular, tirelessly proofread the material I sent him, at times on a daily basis.

I’d also like to thank my mom, Farideh Fozoonmayeh, for her support. She believed in my ability to eventually succeed, although "a philosopher is a man who has to cure many intellectual diseases in himself before he can arrive at the notions of common sense" (Wittgenstein).

Last but not least, none of this would have been even conceivable without Anastassia Solovieva. Nastya's hedonistic outlook, and at times stoic attitude, as well as her loving support, led this project to fruition. Thank you!
To B&B
Please do not confuse *nudge* with *noodge*. As William Safire has explained in his “On Language” column in the *New York Times Magazine* (October 8, 2000), the “Yiddishism noodge” is “a noun meaning ‘pest, annoying nag, persistent complainer.’ . . . To nudge is “to push mildly or poke gently in the ribs, especially with the elbow.” One who nudges in that manner—“to alert, remind, or mildly warn another”—is a far *geshrei* from a *noodge* with his incessant, bothersome whining. *Nudge* rhymes with *judge*, while the *oo* sound in *noodge* is pronounced as in *book*.

Thaler and Sunstein, *Nudge*

A man journeyed to Chelm in order to seek the advice of Rabbi Ben Kaddish, the holiest of all ninth-century rabbis and perhaps the greatest *noodge* of the medieval era. “Rabbi,” the man asked, “where can I find peace?” The Hassid surveyed him and said, “Quick, look behind you!” The man turned around, and Rabbi Ben Kaddish smashed him in the back of the head with a candlestick. “Is that peaceful enough for you?” he chuckled, adjusting his *yarmulke*… [T]he point of this tale is that this man has nothing better to do with his time than journey around and get on people's nerves. For this, the Rabbi bashes his head in, which, according to the Torah, is one of the most subtle methods of showing concern.

Woody Allen, *Getting Even*
TABLE OF CONTENTS

Chapter I. Introduction: Debates, Concepts, and Nudging Mechanisms 1

1. What the Nudge Debate Is and How to Address It 1
2. Inconsistent Use of “Nudge” in the Literature 12
3. Nudging Mechanisms, Thaler and Sunstein-Style 16
4. Towards an Alternative Account of Nudging Mechanisms 24
5. Conclusion: Why Nudging Mechanisms Matter 32

Chapter II. Salvaging the Concept of Nudge 34

1. Introduction 34
2. Conditions for the Preservation of Freedom of Choice 38
3. Easily Resistible Nudges 50
4. An Amended Concept of Nudge 68
5. Conclusion: The Moral Import of the Concept of Nudge in Healthcare 75

Chapter III. Liberty, Mill, and the Framework of Public Health Ethics 77

1. Introduction 77
2. Public Health Ethics, the “Harm Principle” and Core Liberal Commitments 80
3. Mill’s Tripartite Division of Liberties 81
4. Mill on Liberty and Self-Determining Lives 85
5. A Nuanced Millian Framework for Public Health Ethics 92
<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>3. Decisional Capacities and the Kantian and Millian</td>
<td>178</td>
</tr>
<tr>
<td>Case Against Nudges</td>
<td></td>
</tr>
<tr>
<td>4. Nudges and the Duty Not to Weaken Decisional Capacities</td>
<td>183</td>
</tr>
<tr>
<td>or Undermine their Exercise</td>
<td></td>
</tr>
<tr>
<td>5. Nudges and the Duty to Engage Decisional Capacities</td>
<td>187</td>
</tr>
<tr>
<td>6. Nudges and the Duty to Enhance Decisional Capacities</td>
<td>190</td>
</tr>
<tr>
<td>7. Concluding Thoughts</td>
<td>193</td>
</tr>
<tr>
<td>Appendix A: List of Examples</td>
<td>196</td>
</tr>
<tr>
<td>Appendix B: Letter from Angus Dawson, Editor-in-Chief of</td>
<td></td>
</tr>
<tr>
<td>Public Health Ethics</td>
<td>202</td>
</tr>
<tr>
<td>References</td>
<td>203</td>
</tr>
</tbody>
</table>
LIST OF CHARTS

Chart 1.1: A dual process model of thinking and decision-making

(adapted from Kahneman (2003, 1451)) 18

Chart 1.2: A tri-process model of thinking and decision-making

(adapted from Stanovich (2011, 62)) 31

Chart 2.1: The continuum of influences from controlling to noncontrolling

(adapted with modifications from Faden and Beauchamp

(1986, 259, fig. 2)) 45

Chart 5.1: The overlapping territories of population-wide health interventions 145

Chart 5.2: Public health intervention pathways 163
LIST OF TABLES

Table 2.1: Taxonomy of deliberate directive interventions 72

Table 5.1: Two conceptions of justified public health interventions 146
CHAPTER I
INTRODUCTION:
DEBATES, CONCEPTS, AND NUDGING MECHANISMS

1. What the Nudge Debate Is and How to Address It

The nudge debate

Health nudges have recently become the center of public debate. In the United States (US), the “Science of Behavior Change” is a new priority for the National Institutes of Health (NIH SOBC Workgroup 2009). In the United Kingdom (UK), Prime Minister David Cameron is advised by the famous “Nudge Unit”—a team of behavioral economists who favor non-coercive policies of the nudge-kind (Cabinet Office Behavioral Insight Team 2010; Dolan et al. 2011). The use of population-wide health nudges that target groups of people has gained particular momentum in public health, health policy, and health promotion—where actions and policies can affect individuals’ health behavior and use of the healthcare system. Nudges are expected to efficaciously, gently, and cheaply steer patients, health care providers, and other stakeholders toward health-promoting behavior without forcing them to change their ways. Examples of interventions usually labeled “nudges” include:
Anti-Suicide Bridge Flyers. On a bridge where frequent suicide attempts occur, the Seoul City government puts flyers that read: “Did you eat anything yet?” (see, URL = http://www.theworld.org/2012/09/bridge-signs-used-in-south-korea-anti-suicide-efforts/).

Cafeteria. A cafeteria manager places healthy food at eye-level at the beginning of the food queue. Unhealthy food comes last and is least visible. The customer is then more likely to purchase healthy food (Thaler and Sunstein 2008, 1-3; Just and Wansink 2009).

Clinical Trial Enrollment. All research participants who have completed a clinical trial are, by default, contacted repeatedly to participate in future research, unless they actively signal their unwillingness to participate. This opt-out approach to recruitment to clinical trials is expected to increase the rate of participation in medical research, and thereby promote health. It relies on a foot-in-the-door technique (Junghans et al. 2005).

HIV-Test Cash Transfer. In Malawi, residents who pick up their HIV-test results receive 10% of their daily wage in cash (Balz March 16, 2011).

Organ Donations Opt-Out. The state makes organ donor status the default option with the possibility of opting out (Halpern, Ubel, and Asch 2007; Thaler and Sunstein 2008; Whyte et al. 2012).
Behavioral economist Dick Thaler and law scholar Cass Sunstein popularized the term “nudge” in an eponymous 2008 book. Even though they do not provide the reader with a technical definition of a nudge, the most plausible reconstruction of their view is that nudges are influence attempts that are expected to activate our automatic and unconscious cognitive processes, rather than controlled and conscious reasoning, and preserve freedom of choice (Thaler and Sunstein 2008). The concept of nudge draws heavily on the experimental literature in “behavioral sciences” (psychology of decision-making, social psychology, behavioral economics, and neuroscience).

Thaler and Sunstein justify the use of certain nudges on the basis of what they call “libertarian paternalism” (2003). This theory contends that public and private institutions are permitted to use nudges to steer people toward healthy choices that benefit them. The type of nudges libertarian paternalists defend would thus not only preserve freedom of choice (all nudges do), but they would also be designed to our benefit (sparing us harms due to irrational decisions), and would match our informed preferences (Thaler and Sunstein 2008, 10).

A “paternalistic” justification of any policy is controversial. At the same time, libertarian paternalism sounds like an oxymoron. That said, this chapter and the next are not concerned with the justification of nudges. Instead, they address several fundamental problems relative to the very concept of a nudge. First, many are not convinced this is a clear and precise enough concept. I try to clarify the contours of the concept of nudge. Second, Thaler and Sunstein do not provide sufficient arguments to show that influencing choice by activating automatic cognitive processes that bypass our deliberative capacities is conceptually (let alone empirically) compatible with freedom of choice (in a substantive sense).
Opponents claim that since nudges activate automatic cognitive processes, conscious deliberation is unduly bypassed (Hausman and Welch 2008). Therefore, the nudgees' opportunities and capabilities for actual dissent are severely restricted. They claim that nudges undermine “deliberative autonomy,” even if they are not directly coercive (2008). They are not a benign form of paternalism, assuming paternalism can ever be benign. Nudges, it is claimed, express deep disrespect for individuals who do not share the nudger’s values or do not assign the same weight to their health (or to different dimensions of health). Nudgees are deemed “irrational,” and their choices are insidiously influenced. Nudging is insulting. In the last analysis, nudges are an “assault on human dignity” and create the illusion of choice (White 2011).

Finally, many have objected that recourse to nudges illegitimately bypasses democratic debate about the ends of health policy. For those concerned with democracy, the attractiveness of nudges testifies to the evils of technocratic reliance on the expertise of social and cognitive scientists for the determination of the ends and means of public health policy that affect people’s life-style in profound ways. Contrary to what is advertised, nudges do not always track individuals’ preferences and ends. Even if nudges were in conformity with people’s preferences, however construed and measured, they would still be deeply problematic. There is no legitimate policy without democratic deliberative processes that allows people to publically discuss the ends they wish to pursue. In a liberal democracy, it is not even clear that the state should promote any nonconsensual ranking of values that assumes all individuals assign the same relative value to health or to different dimensions of health (see chapter 5).1

1 For background information about the ideological, political, and social context of the nudge debate in the US and the UK, see Burgess (2012); Foer and Scheiber (May 6, 2009); Oliver (July 20, 2011); Schlag
Dissertation outline

In this dissertation, I will specifically examine the ethics of using nudges to protect and promote public health, that is, a subset of population-wide health nudges. Although there is a scholarly debate about the reliability and generalizability of the experimental data on which nudge policies are based, I do not discuss this question in the dissertation. For the most part, I assume for the sake of argument that nudges are, to some extent, effective.

Chapter 1. In the remaining sections of chapter 1, I explain why the concept of nudge itself is in need of clarification by first pointing to the incoherent ways it is currently used. I next start the inquiry on the nature of nudges by exploring the mechanisms through which they operate, and show how this issue has important ramifications for the ethics of nudging.

Some scholars suspect that nudges may subject us to the control of others because of the mechanisms through which nudges operate, namely the activation of automatic cognitive processes. Critics rely on the connection Thaler and Sunstein explicitly establish between their concept of nudge and a dual-process theories of thinking and decision-making (2008, 20). According to this theory, the mind processes information through two distinct systems. “System 1” is automatic, unconscious, uncontrolled, heuristic, fast, and cognitively parsimonious. “System 2” is reflective, conscious, controlled, analytic, slow, and cognitively demanding. As automatic cognitive processes entirely bypass deliberation,

opponents of nudging claim nudges cannot genuinely preserve freedom of choice. The influencer does not offer the nudgee a real opportunity to resist the influence attempt.

I argue that multi-process theories of thinking and decision-making offer both a better explanation of nudging mechanisms and a more fruitful starting point for the ethics of nudging. More precisely, I show that nudges operate through “shallow cognitive processes,” many of which involve some form of incomplete deliberation, rather than no deliberation at all. I use Keith Stanovich's tri-process theory of reasoning and decision-making to illustrate this more granular approach to degrees and kinds of deliberation. Nudging mechanisms matter for the moral evaluation of public health nudges because we, at least sometimes, care about whether we are in a position to deliberate or not over certain choices and whether our theoretical and practical reasoning is reliable, no matter whether we phrase our concern in terms of “deliberative autonomy” or not.

Chapter 2. Chapter 2 pursues the trajectory of chapter 1 by offering a full account of the concept of nudge. More precisely, it addresses the challenge of reconciling the claim that nudges preserve freedom of choice with the fact that they activate shallow cognitive processes.

I first maintain that an influence genuinely preserves freedom of choice if and only if it is choice-set preserving and fully or substantially noncontrolling. Nudges are therefore best understood as influences that primarily work by activating shallow cognitive processes and are choice-set preserving and substantially noncontrolling. I next provide an easy resistibility criterion to ascertain substantial noncontrol. I show that under certain conditions the preservation of freedom of choice is compatible with the use of influences that activate shallow cognitive processes. I maintain that the debate over nudges is plagued
by confusion between real nudges, which are easily resistible, and influences that activate the same cognitive mechanisms as nudges, but are not easily resistible. Nudges are to be understood as a distinct, useful, and sufficiently precise concept. They enrich our taxonomy of influences classified according to the degree of control they exert on the influencee.

The account of nudges I defend in chapter 2 is not a moralized one. As a consequence, the remaining chapters explore various ethical problems about the use of nudges, so understood, to protect and promote public health. The focus of chapters 3 and 4 is on the justificatory burden that government and its agencies should be seen to bear when they infringe on some liberties for the sake of public health. This is an important issue because proponents of public health nudges often assume that, from a moral standpoint, nudges ought to be systematically preferred to more controlling interventions, such as the use of coercion or the elimination of choice.

Chapter 3. In chapter 3, I present a framework of public health ethics inspired by John Stuart Mill's understanding of self-determination that I have elaborated in conjunction with Madison Powers and Ruth Faden (2012). The claim here is that the presumptive weight assigned to particular liberties should be indexed to how important they are for leading a self-determining life. I defend the view that liberties do not all have the same presumptive weight when the state interferes with liberties for the sake of public welfare protection and promotion. What we mean here by a self-determining life is a life that is, in its main contours, free from the exercise of power by other individuals and by social and political institutions. It is also a life endowed with enough material resources and opportunities to put the individual in position to actually elaborate and execute a “plan of life.”
overarching ethical goal is to secure a substantial degree of control over the broad shape of one’s life.

According to this view inspired by Mill, there is a distinction between three tiers of liberty interests classified according to their decreasing importance for leading a self-determining life: vital liberty interests that should be strongly shielded from state interference (first tier); consequential liberty interests that enjoy a presumption in their favor (second tier); and non-consequential liberty interests that enjoy no such presumption (third tier). If successful, this argument for the differentiation among presumptions due to liberties of various sorts removes one important line of defense for the view that even the most modest proposals for public health interventions that interfere with non-consequential liberties bear a heavy burden of justification.

*Chapter 4.* Chapter 4 applies the results of the three first chapters of my dissertation to the ethics of public health nudges. I discuss what are the conditions, if any, under which nudges should be systematically preferred to more controlling influences. A widespread assumption among public health ethicists is that all morally justified public health measures (laws, policies, regulations) ought to respect the principle of the least restrictive alternative. This principle is usually understood as stating that public authorities must give very strong priority to the use of means that are the least restrictive of personal liberties to achieve their public health objective.

In chapter 4, I examine the problem of the function and scope of the principle of the least restrictive alternative and its ramifications for the ethics of nudging. I defend the view that this principle is best understood as a three-step decision procedure requiring public health authorities to:
(1) Determine the legitimacy of the public health goal they consider pursuing;
(2) Guarantee that the relationship between the means of the intervention and its ends is strong enough;
(3) Be in a position to demonstrate that they have minimized the infringement of that intervention in proportion to the importance of the liberty interest at stake.

As I explain, the application of the decision procedure varies depending on the type of liberty at stake, making use of the tripartite division of liberties introduced in chapter 3.

I maintain that when the state considers interfering with first tier liberties, it should give priority to the use of the least controlling influences. If nudges are available, they are preferable to coercive interferences. Naturally, if persuasive strategies are feasible, they are preferable to nudges.

When the state contemplates interfering with second tier liberties, PLRA's defeasibility conditions are weakened. Nudges should still be, as a rule of thumb, preferred to coercive interferences.

When the state contemplates implementing policies that interfere with third tier liberties, I argue that PLRA does not require minimizing infringement on liberty since no important liberty is at stake. This is a radical view. Even as a rule of thumb, the state should not consider the use of nudges as preferable to recourse to more controlling means.

Chapter 5. In chapter 5, I examine the justificatory burdens that are unique to the use of public health nudges. I begin by showing that population-wide health nudges and public health nudges are not coextensive. As a consequence, some population-wide health nudges should be assessed according to considerations of particular concern in health policy, clinical research, clinical medicine, and health promotion, rather than public health per se.
I then distinguish between the minimalist and the maximalist conceptions of what it takes to justify public health interventions. The minimalist conception has been developed around demanding standards for the justification of coercive governmental interventions to protect and promote public health. In public health minimalism, only combating major threats to health and securing the efficient delivery of genuine public goods justify use of the coercive force of the state. I argue why public health minimalism may have normative and conceptual resources to deem unjustified the use of many health nudges, although it is exclusively concerned with state coercion.

Public health maximalism is the view that a much broader range of interventions might be justified on a distinctively public health basis. On this view, public health is concerned with the protection and promotion of population-level health. More expansive justificatory considerations lend support to a variety of coercive or noncoercive actions performed by state and nonstate agents. The maximalist framework therefore applies more directly to the justification of public health nudges than does minimalism, and is more likely to back the use of health nudges for public health reasons.

My aim is neither to defend nor to assess the merits of public health minimalism and maximalism. It is instead to highlight their distinctive characteristics, the challenges they each face, and their ramifications for the justification of public health nudges. Public health minimalism and maximalism are the two ends of a spectrum. Public health ethicists may defend positions that have more or less affinities with each polar extreme.

Chapter 6. Chapter 6 focuses on two major problems with the use of public health nudges. First, I explore the role that paternalism can play in justifying or criticizing the use of public health nudges. I argue two points. First, nudges are not paternalistic by the classical
conception of paternalism. Second, nudges are paternalistic on expansive conceptions of paternalism, but these conceptions are misguided because they do not track what is potentially worrisome about paternalism. This is why, in my opinion, it is unhelpful to focus the nudge debate, as it so often has been, on paternalism, following the trend set by Thaler and Sunstein. We should instead be tackling a more diverse set of substantive issues. I next identify one of these substantive, though neglected, issues: we may feel uncomfortable with the use of nudges because nudges involve some kind of meddling with our decisional capacities.

In order to address this problem, I first examine possible Kantian and Millian arguments against nudges, based on the view that they interfere with decisional capacities in an objectionable manner. Some Kantian deontologists may argue that nudges express disrespect for the dignity of persons because the nudger does not present the nudgee with reasons for action and may seem to undermine the human capacity for commitment and responsibility. Nudges, they may worry, undermine the conditions for personhood. Some Millian consequentialists may claim that nudges undermine the conditions for human flourishing by preventing the exercise and development of decisional capacities. Nudges undermine the conditions for adulthood. According to proponents of these views, even self-nudges raise concerns.

Scholars with these worries might claim that public health agents should abstain from the widespread use of nudges, and from encouraging others to use them. They might argue that we have three duties with regard to decisional capacities: a duty not to in the long run weaken our decisional capacities or undermine their proper exercise; a duty to engage deliberative capacities when they are present, rather than bypass them; and a duty
to enhance decisional capacities. Some Kantians and Millians might contend that public health nudges violate these duties.

My response to Kantians and Millians who may have these worries is two-fold. First, I agree with them that agents indeed have a duty not to weaken our decisional capacities or undermine their exercise. I show which public health nudges this duty rules out, focusing on the cumulative effect that might result from the use of a myriad of non-coordinated nudges. Second, I argue, against Kantians and Millians, that there is no perfect duty to engage or enhance decisional capacities whenever one can. Therefore, I do not believe public health nudges that do not engage or enhance decisional capacities are necessarily wrong or even distasteful.

2. Inconsistent Use of “Nudge” in the Literature

Cacophony

There is surprisingly little consensus in the literature and public debate on what is meant by the metaphor of a “nudge.” Most authors cite Thaler and Sunstein and define a nudge as “any aspect of the choice architecture that alters people’s behavior in a predictable way without forbidding any options or significantly changing their economic incentives” (Thaler and Sunstein 2008, 6). Because this definition is not technical and is rather capacious, different concepts circulate under the “nudge” label.

To some, Thaler and Sunstein’s definition seems overly fuzzy to be of any use, encompassing anything but outright coercion. It is difficult, they claim, to distinguish a nudge from health promoters’ decades old persuasive strategies informed by social
marketing (Bonnell et al. 2010). A nudge could designate any non-coercive intervention that does not use strong incentives or disincentives, including information-provision or simply good design (e.g., user-friendly medical equipment). Marteau and colleagues conclude:

[T]here is no precise, operational definition of nudging. This may reflect a reality—namely that nudging is at best a fuzzy set intended to draw attention to the role of social and physical environments in shaping our behavior and not to inform a scientific taxonomy of behavior change interventions. (2011, 263)

The British House of Lords’ recent critical report of the “Nudge Unit” illustrates this point (2011). The report notes that the House’s Science and Technology Committee “received differing accounts of the Government’s use of ‘nudge’.” For example, the Sustainable Transport White Paper contrasts ‘nudging’ with anything that forbids or restricts choice. Norman Baker MP, Parliamentary Under Secretary of State for Transport, equated ‘nudging’ with a “broad range of non-regulatory interventions, such as the provision of bus and train timetables” (2011, 12). The authors of the House of Lords’ report provide their own definition of nudges as follows: they “prompt choices without getting people to consider their options consciously, and therefore do not include openly persuasive interventions such as media campaigns and the straightforward provision of information” (2011, 12). However, they do not use the term consistently either: a few pages earlier, they classify “provision of information” under the “nudge” category (2011, 10).

Similarly, many people simply assume that “nudge” refers to any intervention designed by behavioral economists (Blumenthal-Barby and Burroughs 2012) or a subset of such techniques they use, such as the use of defaults (automatic assignment to an option if
the individual fails to actively choose an option) (Nuffield Council on Bioethics 2007, 42). In all cases, the different conceptual borders of nudges are unclear or debatable.

To avoid imprecision, some work with a definition of nudge based on their satisfaction of second-order preferences (preferences over preferences, such as preferring not to prefer smoking). For instance, Theresa Marteau and her colleagues once suggested that “[n]udging could be defined as increasing the chances that people act in ways that, on reflection, they would have chosen themselves; this is variously described as acting on preferred preferences [i.e., second-order preferences] or acting consistently with deeply held preferences” (2008, 122). However, this is a characterization that overlaps more with what Thaler and Sunstein mean by libertarian paternalism than by their understanding of a nudge.

The confusion is recurrent in the literature and often times “libertarian paternalism” and “nudge” are used interchangeably. For instance, “strategies from libertarian paternalism” sometimes refer to interventions allowing patients to opt out of the “preferred choice” (Wheeler et al. 2011).

In fact, though, Thaler and Sunstein strongly reject the view that nudges are by definition in-line with the influencees' preferences. Granted, Thaler and Sunstein sometimes suggest that nudges pick out only interventions that are based on individuals' second-order preferences and are likely to benefit them (2008, 6). Examples include:

**Deposit Contract.** All primary care physicians of a healthcare system offer their patients the possibility to voluntarily deposit an agreed-upon sum of money with the physicians. The physicians will then return the money to the patients in small installments if the latter meet certain agreed-upon health objectives to improve their
health (e.g., losing weight, exercising, quitting smoking) (Thaler and Sunstein 2008, 232).

But elsewhere Thaler and Sunstein claim that nudges do not always coincide with the nudgees' second-order preferences (preferences about other preferences, creating a hierarchy among preferences) and are unlikely to benefit them. Consider an example they label “nudge”:

**Nazi Ballots.** In 1938, the Nazis organized a referendum asking German voters whether they supported (1) the reunification of Austria and the German Reich, (2) and the party of “our leader, Adolf Hitler.” To say “Yes,” voters had to check a large circle right in the middle of the ballot; to say “No,” they had to check a small circle on the right side of the ballot (Balz, August 11, 2010).

In the light of these inconsistencies, is “nudge” a catchy metaphor or a construal capable of playing a precise and useful function in our public discourse? Can we debate the ethics of health nudges, if we do not even agree on what they designate?
3. Nudging Mechanisms, Thaler and Sunstein-Style

Nudges and dual process theories of thinking and decision-making

The theoretical model that Thaler and Sunstein use to construe the idea of nudge is borrowed from behavioral economics. In this section, I will provide a short conceptual history of the origins of nudge theory.

Behavioral economics is a school of thought within economics, which strives to bridge the gap between economics and psychology. Mainstream economists build up models of consumer choice and economic decisions, based on the assumption that an ideal chooser would reason according to rules of logic and the norm of expected utility maximization. That is, it is assumed that when making decisions under risk, a rational individual tries to maximize her utility (a quantifiable but subjective value) by taking into account the consequences of her choices (their magnitude) and their associated probabilities. This model is normative and descriptive (actual individuals’ behavior approximates the theory’s predictions), but it starts with normative assumptions and axioms.

By contrast, behavioral economists build models of consumer choice and economic decisions based on evidence from psychological experiments about the actual patterns of reasoning and decision making of average individuals. These experiments show that the response of average individuals on tasks designed to measure reasoning and decision-making performance often do not even approximate the response predicted by normative theories of decision-making (Gilovich, Griffin, and Kahneman 2002; Wilkinson 2008). In their popular writings, Thaler and Sunstein contrast “Humans” (the average individual) and “Econs” (the utility maximizing ideal economic agent).
Importantly, Thaler and Sunstein elaborate their concepts of nudge by drawing heavily on two interpretations of the experimental results of behavioral economists. The first explains the mechanism by which nudges work; the second focuses on the connection between those mechanisms and the debate about norms of rationality.

First, for Thaler and Sunstein, a dual-process theory of reasoning and decision-making can explain the experimental results about human systematic decision-making deficits. The theory they favor claims that the mind processes information using two distinct systems (Thaler and Sunstein 2008, 20; Kahneman 2003 and 2011). “System 1” is automatic, unconscious, uncontrolled, heuristic, fast, cognitively parsimonious, and provides quick responses that are adaptive only in environments similar to those for which this system has evolved. In a different environment, the response they provide is often off the mark. “System 2” is reflective, controlled, analytic, conscious, slow, cognitively demanding, rule-governed, and it provides responses that are more accurate. While System 1 is adaptive, the goals it pursues are by and large fixed by evolutionary processes (e.g., automatic fear reaction to a threat, such as a tiger). System 2 is slower, but its goals are less fixed, so it allows us to adapt our behavior to a broader range of circumstances and to plan our actions ahead of time and take into account their consequences (e.g., preparing for contingencies if one is likely to face a tiger). With regard to the experimental tasks mentioned above, subjects who provide responses that do not match what mainstream economic theory predicts rely on System 1 processing, whereas those who reach the expected responses rely on System 2. Chart 1.1 borrowed from Daniel Kahneman summarizes the model set out by this theoretical approach:
Chart 1.1: A dual process model of thinking and decision-making  
(adapted from Kahneman (2003, 1451))

The second interpretation of the experimental results concerns the status of the norms of rationality. For some, these experiments show that humans are often and predictably irrational, since their mode of reasoning and final conclusions deviate from the norms of reasoning and decision-making espoused by mainstream economics (Kahneman 1994). Others claim that these results can either be explained away by features of the task that induce humans to make certain assumptions about the problem at hand, or they require us to revise our norms of rationality that are simply not suitable for “mortals” with limited cognitive capacities (Gigerenzer 2008). I will say no more about the “great rationality debate” (Andler 1995; Stein 1996) in cognitive science and the extent to which mainstream economics models need to be revised (see Glimcher 2011, chapter 5 for an insightful recent overview). The main point is that Thaler and Sunstein belong to the first group of
theorists. For instance, relying on System 1 for one-time consequential decisions involving trade-offs between short-term and long-term gratification or complex decisions requiring probability estimates, often leads to normatively inadequate decisions. These decisions are “irrational” in the sense that they are suboptimal as compared to the standards of expected utility maximization.

These two interpretations of the experimental results on human reasoning and decision-making lead Thaler and Sunstein to defend the view that public policy and law should take into account the effects of System 1 on our decisions and behavior. The driving idea of nudge theory is to externally activate System 1 processes to increase the likelihood of certain behavior. Of course, one could use System 1 mechanisms to coerce someone to do something. For instance, if you have a phobia of spiders, I can threaten to put a tarantula on your face, if you do not comply with my demand. Thaler and Sunstein need to argue that one can use the predictable effects of System 1 and at the same time preserve freedom of choice. This is the heart of nudge theory. I make this point explicit in the next section.

Thaler and Sunstein’s theory was initially designed to solve the problem of government regulation of the economy. Traditionally, many economists favor the free market and believe that the economic system should be arranged in a way that would maximally satisfy individuals’ preferences revealed by their choices. One lesson we can draw from behavioral economics is that choices do not necessarily reveal preferences (Ubel 2009). What people “choose is strongly influenced by details of the context in which they make their choice, for example default rules, framing effects (that is, the wording of possible options), and starting points” (Sunstein and Thaler 2003, 1161). If choices do not track preferences because preferences are unclear, unstable, and not well ordered, then the
standard free-market approach to the regulation of the economy is questionable from within. However, Thaler and Sunstein do not favor coercive regulations by the law either. They wish to find a consensual middle ground between regulatory interventionism and the free-market, such that it is “acceptable to those who are firmly committed to freedom of choice on grounds of either autonomy or welfare” (2003, 1160). Thus, they first used the term “libertarian paternalism” to defend the use of tools from behavioral economics (which had, back then, no generic name) to steer individuals in a welfare-promoting direction in-line with what they would choose if they were fully rational, while preserving freedom-of-choice (Sunstein and Thaler 2003). Politically, their theory was centered on interventions that influence people’s choices in subtle ways and their justification.

**The Non-Neutrality Thesis and freedom of choice**

The introduction of the word “nudge” in Thaler and Sunstein's 2008 book changes the focus of their theory. This term is distinct from libertarian paternalism (what they call a “political philosophy”). According to Thaler and Sunstein, it is a generic name for any intentional or non-intentional feature of the natural, built, or social environment that influences choice by relying on automatic cognitive mechanisms, while preserving freedom of choice:

A nudge, as we will use the term, is any aspect of the choice architecture that alters people’s behavior in a predictable way without forbidding any options or significantly changing their economic incentives. To count as a mere nudge, the intervention must be easy and cheap to avoid. Nudges are not mandates. Putting the fruit at eye level counts as a nudge. Banning junk food does not...In accordance with our definition, a nudge is any factor that significantly alters the behavior of Humans, even though it would be ignored by Econs. Econs respond primarily to incentives. If the government taxes candy, they will buy less candy, but they are not influenced by such “irrelevant” factors as the order in which options are displayed.
Humans respond to incentives too, but they are also influenced by nudges. (2008, 6-8)

The problem with this definition is that it shifts the focus from intentional interventions to *any* form of influence no matter its source. In one sense, a nudge designates a feature ("aspect," "factor") of the environment that produces significant effects on the individual navigating this environment. Thaler and Sunstein present this idea by using an abstract language that borrows from consumer decision analysis and information processing theories. According to this family of theories, the chooser’s behavior depends on patterns of information available in her choice situation that she exploits. It does not matter whether the choice situation is natural, built or social. A mountain-climber spontaneously grabs rocks with certain shapes suitable for the human hand. These rocks appear more salient to us because they are, so to speak, naturally calibrated to produce the smoothest interface between our mind and the world. Let us call this sort of features of our environment, or cues that trigger and effortlessly guide behavior, "behavioral triggers."  

Nonetheless, once we pay attention to the contexts in which the notion of nudge is put to use, it appears that it refers more narrowly to a subset of behavioral triggers: those involving deliberate interventions. This is most clearly evoked when Thaler, Sunstein, and Balz mention the activity of a "choice architect" who makes decisions about the organization of the choice environment.  

----

2 I use the term “behavioral” instead of “behavior” in order to emphasize the conceptual connection between these concepts and the framework of behavioral economics.  

3 The notion of choice architecture is often mentioned in the nudge theory but has not been the object of extensive discussion in the context of the ethics of nudges. The metaphors of choice architecture, the
A choice architect has the responsibility for organizing the context in which people make decisions…. Doctors describing available treatments to patients, human resource administrators creating and managing health care plan enrollments, marketers devising sales strategies, ballot designers deciding where to put candidate names on a page, parents explaining the educational options available to a teenager; these are just a few examples of choice architects. (Thaler et al. 2010, 2)

Let us call this subset of behavioral triggers, “behavioral triggering interventions.” Should we then define nudges broadly, as behavioral triggers that preserve freedom of choice or narrowly as behavioral triggering interventions that preserve freedom of choice? I will argue that the most plausible and coherent reading of Thaler and Sunstein’s concept of nudge must restrict them to behavioral triggering interventions.

Thaler and Sunstein could support the view that nudges designate freedom-preserving behavioral triggers in general by defending the following thesis:

**Non-Neutrality Thesis.** Our behavior is always under some form of unnoticed environmental influence.

While I endorse the **Non-Neutrality Thesis**, I contest the conclusions Thaler and Sunstein implicitly draw from it. They do not believe there are significant differences between unconsciously processed influences that have different sources: artificial or natural, deliberately put in place or random. For them, using nudge to cover intentional or

architecture of information, and default options all belong to a lexicon imported from cognitive science (in particular, computer science) to political science and the behavioral theory of large-scale organizations in modern societies (Jones 2001).
non-intentional behavioral triggers does not generate any confusion, but instead highlights the importance of the non-neutrality of any choice environment.

We need to resist this conclusion, which does not necessarily follow from the Non-Neutrality Thesis. What matters morally and politically is precisely deliberate interventions, not behavioral triggers in general. This is because we care about who alters the controllable features of our environment that affect our behavior, not natural causes over which humans have no say.

The freedom of choice feature of nudges squarely applies to intentional interventions. From the Non-Neutrality Thesis it does not follow that intentional interventions and random features of our environment are morally similar. The former require scrutiny, because we can change them, and because if others select behavioral triggers, they are to some extent in control of our behavior.

Of course, it could be argued that in some sense, behavioral triggers that no agent has intentionally positioned in our environment (e.g., the shape of a rock) increase or decrease our freedom of choice (Carter 2003). This is a disputable claim, but even if that were the case, intentional interventions and natural causes would still be morally distinct. The source of limitations of freedom of choice matters. The center of gravity of Thaler and Sunstein’s theory is interventions that matter morally and politically in laws and public policy. It is better, I believe, to call a subset of behavior triggers, “behavioral triggering interventions,” to highlight the morally relevant fact that these are intentional interventions. All behavioral triggers affect our choices because they activate automatic cognitive processes, but only some of them truly affect our freedom.

It is not clear that nudges genuinely preserve freedom of choice because, by activating automatic cognitive processes, they might be controlling the influencee. Indeed,
they are assumed to simply bypass deliberation. As Christian Rostbøll notes, without allowing the target of the intervention to deliberate over her decision, the nudgees might be “treated as not being responsive to reasons but as mere objects that react in calculable and predictable ways to default rules, framing effects, starting points, and so on” (Rostbøll 2008, 94-95).

This is the core of the problem with nudges. Do nudges allow for the possibility of making meaningful choices? Given that the main obstacle for claiming that nudges genuinely preserve freedom of choice is that they bypass deliberation, this is the point at which I start the inquiry. In the next section, I argue that an alternative account of nudging mechanisms clarifies the way nudges affect deliberative capacities. I deny that nudges necessarily bypass deliberation. I maintain that they often involve some form of incomplete deliberation. If my view is correct, then the problem of the impact of nudges on freedom of choice can be framed more precisely.

4. Towards an Alternative Account of Nudging Mechanisms

I offer in this section the framework for an alternative account of nudging mechanisms. I show that if we abandon certain assumptions of dual-process theories of reasoning and decision-making, it becomes clear that, while some behavioral triggering interventions indeed bypass conscious deliberation, others (central to Thaler and Sunstein’s argument) activate an incomplete form of deliberation. In order to make this point, I explore recent work in tri-process theories of reasoning and decision-making that do not rely on a dichotomy between automatic, unconscious processes and controlled, deliberative
processes. These new theories introduce forms of deliberation that are incomplete (in a sense I explicate below). Although these theories do not yet offer sufficiently granular distinctions between degrees and kinds of incomplete deliberative processes, they provide a helpful starting point for rethinking nudging mechanisms and the locus of moral concern with nudges.

In contrast to dual-process theories, some scholars defend tri-process theories of thinking and decision-making (Frankish and Evans 2009, 21). I rely here on Stanovich’s version of that theory, but other tri-process theories would lead to a similar conclusion (Evans 2009). Dual and multiple-process theories are constantly evolving. There is no final conclusion in this domain, and future development may also help to better understand how behavioral triggers (in general) work. The main point I wish to make is that there are good reasons to believe that deliberation is not psychologically all or nothing.

In brief, Stanovich and most psychologists have abandoned the System 1/System 2 dichotomy, although Stanovich coined these terms and contributed to their popularity (Stanovich 1999; Kahneman 2011). This is because there are not just two ways of processing information, but a multitude of processing systems that have some commonalities. Hence, Stanovich uses the labels “Type 1” and “Type 2” processes to distinguish processes that entirely bypass deliberation from those that involve some form of deliberation. However, he divides Type 2 processes into two distinct subcategories (hence the “tri-process” name).

Type 1 processes refer to a set of systems that “operate autonomously in response to their own triggering stimuli and are not under high-level cognitive control” (Stanovich 2011, 19). Type 1 processes have some commonalities: they do not require a heavy load on central processing capacity, they are rapid and they operate in parallel without interfering
with one another. They include a variety of processes such as those involved in emotional regulation (or affective responses), encapsulated evolutionary adaptive quasi-modules (such as mind reading capacities), habit-systems (automatic firing of overlearned associations in implicit learning and conditioning), impulse-systems, etc. (Stanovich 2011, 63, Vlaev and Darz 2012). The response they provide is generally in the right ballpark but is not the sophisticated result of fine-grained analysis. Stanovich calls Type 1 processes “The Autonomous Set of Systems,” or TASS, since their processing does not require input from Type 2 processes. Cafeteria is a good example of an intervention primarily aimed to activate TASS processes, though it does not preclude deliberation. It works through unconsciously processed cues.

Type 2 processes are a set of processes that are relatively slow and computationally more expensive than Type 1 processes and are largely serial (there is little parallel processing). They can override Type 1 processing responses by inhibiting them, and eventually substitute a better response. The important point of Stanovich’s theory is that there are two distinct sets of subsystems in Type 2 processing. The two subsystems are called the “Algorithmic Mind” and the “Reflective Mind”. The Reflective Mind groups systems that deal with the evaluation of beliefs and desires, macro-strategies, and long-term plans that are not evolutionarily tightly determined (“hard-wired”) and have a capacity to decontextualize the problems the individual attempts to solve. Importantly, they are in charge of macro-control of other processes: they can trigger inhibition of the Algorithmic Mind, or of TASS, when the response they provide is not sufficiently good.

The most important set of processes for us is the Algorithmic Mind. Its main function, once activated by an order from the Reflective Mind, is to provide a better response to the problem than what TASS processes provide.
This occurs due to two forms of deliberation. First, the mind can explore a broad, if not always exhaustive, range of hypotheses for solving a problem. This form of “complete” deliberation is computationally demanding but leads to accurate responses. Second, it can deliberate through what Stanovich labels “Serial Associative Cognition” (SAC), which is “cognition that is not rapid and parallel (in the manner of the systems contained in the autonomous mind) but is nonetheless inflexibly locked into an associative mode that takes as its starting point the model of the world that is most easy to construct” (65). In other words, this type of processing involves some form of deliberation. However, it is incomplete because it relies on a single model of the world that it does not question by forming alternative hypotheses and models. Although it is not as rapid as TASS, SAC is still much faster than full-blown deliberation. It provides the individual with an easy way to reason without using a lot of cognitive capacity. It is serial like full-blown deliberation but associative like TASS processes. Importantly, whereas the System 1/System 2 classifications bundled in the same category purely automatic processes and heuristics, Stanovich’s theory clearly shows that the later belong to a separate system. Heuristics (and biases) involve some form of deliberation. After all, they are mental shortcuts.

This type of processing, I argue, is crucial for understanding how behavioral triggering interventions work. Some activate TASS processes (as Thaler and Sunstein also suppose, such as priming effects or gut reactions to affective salience, impulses, etc.), while others activate SAC processes (such as framing effects, anchors, loss aversion, the status quo bias when defaults that are interpreted as recommendations). SAC processes passively accept the formulation of the problem given to them or the one that is most easily constructed from easy to recall information.
A few illustrative examples will help make this point explicit. They are intended to make the TASS/SAC distinction explicit, and I will not evaluate them from a moral standpoint here.

(a) Interventions in which TASS processes are primarily involved:

**Smaller Plates.** Using 10 inches instead of 12 inches plates results in a 22% reduction of calories per serving (Wansink 2006).

**Hand Washing.** To increase the rate of hand washing after toilet use in Ghana, a graphic TV campaign induces disgust and fear of contamination if people do not wash their hands (Curtis, Garbrah-Aidoo, and Scott 2007; Vlaev and Darz 2012).

In the first example, people eat less because they use habitual visual cues with relation to container size, not internal cues about fullness or reasoning, to determine how much to eat.

The second example was experimented with in Ghana where soap use after visiting the toilet was originally as low as 3%. It resulted in a 13% increase in hand washing (better than usual argument-based campaigns designed to change health-related beliefs). This can be explained by reliance on the impulse system: a conditioned stimulus (toilet) is associated to an unconditioned stimulus (impulsive feeling of disgust) (Vlaev and Darz 2012).
(b) Interventions in which SAC processes are primarily involved.

**Positive Surgery Framing.** A patient is told that surgery comes with a 70% chance of survival, he or she thinks: “If I get surgery, I have a 70% chance of survival. This is pretty high! I guess I should go ahead with surgery…”

**Negative Surgery Framing.** A patient is told that surgery comes with a 30% risk of death. He or she thinks: “If I get surgery, I have a 30% risk of death. Gosh, that’s really bad! I don’t want to die… I guess I should not go ahead with surgery…”

The first example used framing effects (Stanovich 2011, 66). The patient given information in these terms will not consciously deliberate about the problem at hand but will passively accept the way a problem is framed and quickly reason from that starting point. In both cases, there is reasoning and conscious deliberation, but it is incomplete. The patient never considers the other way of framing the problem. He or she never compares risks of death and chances of survival from surgery with probability of death or survival if he or she forgoes surgery or gets an alternative treatment. In brief, he or she does not deliberate by considering the broader set of relevant hypotheses. Incomplete reasoning and deliberation are nonetheless still clearly a form of conscious deliberation. This is an issue that scholars working with dual-process theories of reasoning and decision-making tend to neglect because of the taxonomy of cognitive process they use. As a consequence, the debate on the ethics of nudges has not sufficiently focused on what might be specifically worrisome with interventions activating incomplete forms of deliberation. I discuss this problem in chapters 2 and 6.
No doubt Stanovich’s theory needs to be refined, and there are many degrees of deliberative sophistication between SAC and exhaustive consideration of all relevant hypotheses. In addition, many interventions may combine TASS and SAC processes. All I intend to show is that there is some space between no deliberation and full deliberation. Some major types of behavioral triggering interventions (framing, anchors, defaults, loss aversion, etc.) rely on cognitive processes involved that do not fall under the category of automatic cognitive processes that bypass deliberation. In the remaining chapters of this dissertation, I will not use Stanovich’s idiosyncratic lexicon (TASS and SAC processes) but will refer to “nondeliberative” and “incompletely deliberative” cognitive processes (bundled up into the category of “shallow cognitive processes”) to denote a type of granular understanding of the architecture of the mind. My thesis does not depend on the particulars of Stanovich's theory.

Another important point here is to note that psychologists consider the mind a “cognitive miser”. As Stanovich puts it, when faced with a problem, the cognitive miser applies two lexically ordered rules:

Rule 1: Default to TASS processes when possible.
Rule 2: If Type 2 processing is necessary, default to SAC processing.
Chart 1.2 explains how Stanovich's model works:

Chart 1.2: A tri-process model of thinking and decision-making
(adapted from Stanovich (2011, 62))

In sum, many behavioral-triggering interventions involve conscious deliberation. That is why some nudges seem akin to persuasion, yet they are distinct from pure rational persuasion. The latter would engage Type 2 processes more fully. For instance, through rational persuasion, partners in a dialogue may engage in a critical examination of the proposed solution to a practical or theoretical problem. Such critical stance would require (1) the algorithmic mind to generate hypotheses and represent their ramifications, and (2) the reflective mind to assess the logical validity of an argument (and the like), having
recourse, for instance, to metacognitive tools. These complex operations can be contrasted to reliance on just one way of representing a problem and its solution.

In order to have a compelling theory of nudges, we therefore need to rethink how behavioral triggering interventions activating incompletely deliberative cognitive processes impact freedom of choice.

5. Conclusion: Why Nudging Mechanisms Matter

In this chapter, I have shown that the debate about the ethics of nudges has suffered from a lack of conceptual precision. Not only does the literature on nudges use inconsistent ostensive definitions, but even Thaler and Sunstein, who coined the term, do not provide us with an adequate definition. The most sympathetic reading of Thaler and Sunstein is that a nudge is for them either a behavioral trigger or a behavioral triggering intervention that influences the target by activating System 1 cognitive processes. Behavioral triggering interventions have an additional feature: they preserve freedom of choice. This definition entails that nudges necessarily bypass conscious deliberation. I have shown that for many authors, bypassing deliberation is a source of major problems as regards Thaler and Sunstein’s claim that nudges have only benign negative effects.

Next, I argued that Thaler and Sunstein (and their opponents) are mistaken about the mechanism by which nudges necessarily work. While all nudges activate shallow cognitive processes, these processes are of two kinds: some are similar to System 1 processes, while others involve an incomplete form of conscious. If I am right, then the charge that nudges raise moral concerns because they always bypass conscious
deliberation is incorrect. However, this may not be good news for nudge theory. Indeed, my account may be read as raising even more concerns about nudges. Nudges still sometimes bypass conscious deliberation. Even when they do not, it might be thought that deliberately activating suboptimal forms of reasoning that are channeled and not question the starting point of these processes is morally highly problematic. In the next chapter, I offer an amended definition of nudges that answers this objection and addresses other deficiencies in Thaler and Sunstein’s construal of nudge.
CHAPTER II

SALVAGING THE CONCEPT OF NUDGE

1. Introduction

Although Thaler and Sunstein never provide a technical definition of nudge, we can reconstruct their view. For Thaler and Sunstein, A nudges B when A makes it more likely that B will φ (2008, 1) by activating B’s automatic cognitive processes (2008, 19-20) while preserving B’s freedom of choice (2008, 5-8). The concept of nudge matters for the moral evaluation of actions and policies because it is designed to pick up efficacious influences that preserve freedom of choice, yet bypass the deliberative capacities of those influenced. The moral import of the concept of nudge thus primarily arises from the conditions requisite for an influence to truly preserve freedom of choice. Consider a sample of health-affecting influence attempts labeled “nudge” in Thaler and Sunstein's book or its official online companion edited by John Balz (who worked as a researcher on Nudge while completing his Ph.D. under Sunstein) in the Nudge Blog, besides those already cited in chapter 1 (Cafeteria, Deposit Contract, HIV-Test Cash Transfer):

* A shorter version of this text is forthcoming in the Journal of Medical Ethics (Saghai forthcominga), with open-peer commentary by Richard Ashcroft (forthcoming), Luc Bovens (forthcoming), Gerald Dworkin (forthcoming), Brynn Welch (forthcoming), and Alan Wertheimer (forthcoming), as well as my response (Saghai forthcomingb). Thanks to the editors for giving me the permission to use these texts.
**Asparagus-Lovers.** An investigator suggests to research participants that they liked or loved asparagus during childhood the first time they tried it, creating a false memory, and a false belief about the taste of asparagus. Subsequently, participants report increased general liking of asparagus, greater desire to eat it, and willingness to pay more for it (Balz May 14, 2009).

**Generic Medication.** Medicare beneficiaries are given generic medication by default but are offered the option of getting the brand-name drug (Thaler and Sunstein 2008, 169).

**Less Than You Think.** University campuses in Montana organize an alcohol consumption-reduction campaign accurately stating that 81% of Montana college students have four or fewer alcoholic drinks each week. The campaign underlines the fact that the majority of students binge-drink less often than what most students assume (Thaler and Sunstein 2008, 68).

**Paternal Competition.** In an Indian village, health professionals post children’s medical test results in a public place, creating competition among fathers to improve their children’s health (Balz January 27, 2009).

**Quadruple Bypass Burgers.** The Heart Attack Grill offers free “Quadruple Bypass Burgers” (8,000 calories) to people weighing 350 pounds or more (Balz June 2, 2010).
Although some of these interventions may unambiguously preserve freedom of choice, skeptics argue that others do not. For them, preserving freedom of choice requires more than avoiding the use of outright coercion, because we also care about “the control an individual has over his or her evaluations and choices” (Hausman and Welch 2010, 128).

Some scholars suspect that nudges may subject us to the control of others because of the mechanisms through which they operate. As explained in chapter 1, according to Thaler and Sunstein, the mind processes information through two distinct systems (Thaler and Sunstein 2008, 20). System 1 is automatic, unconscious, uncontrolled, heuristic, fast, and cognitively parsimonious, while System 2 is reflective, conscious, controlled, analytic, slow, and cognitively demanding. If nudges harness the automatic processes of System 1, as Thaler and Sunstein tell us, then their influence is efficacious precisely because the influencee does not deliberate over her choice. Why shouldn’t we therefore believe that nudgees are controlled by nudgers, and have no real opportunity to easily resist the attempt at influence? (Anderson 2010; White 2011, chapter 5). Unless this question is answered, nudge is nothing but a catchy metaphor.

In what sense do nudges preserve freedom of choice? Thaler and Sunstein seem to waver between two views: either nudges preserve freedom of choice simply because the nudger does not foreclose any option, or nudges preserve freedom of choice in a more substantive sense. If Thaler and Sunstein defend the former view, many influences will qualify as nudges, and policy-makers will have strong reasons to consider them as prominent regulatory tools. The drawback of this first view is that nudges’ moral appeal will be weak for those concerned with the control individuals have over their choices in the face of attempted influence. If Thaler and Sunstein choose the latter view, nudges will have
enhanced moral traction, but their place in the regulatory toolbox will be more humble because fewer influences will qualify as nudges.

In this chapter, I argue that influences activating cognitive processes that bypass the deliberative capacities of the reflective mind may preserve freedom of choice in a morally robust enough sense, centered on the issue of control. The psychological literature on control and the metaphysical literature on free will partly overlap and cross-reference (Frankfurt 1988; Fischer and Ravizza 1998; Di Nucci 2008; Suhler and Churchland 2009; Faucher 2012). However, in this chapter, my aim is not to address the metaphysical question of freedom of the will, rather it is to elaborate a conceptually sound and morally appealing ecumenical understanding of freedom of choice that can be used in public policy debates (for parallel projects, see Carter (2004) and White (2010)). I will therefore remain neutral on the free will debate while engaging the literature on control. Of course, whether nudges are compatible with free will is a problem that deserves to be addressed separately (Di Nucci 2008; Ashcroft forthcoming).

I thus intend to salvage the concept of nudge from the charge of paying lip service to freedom of choice, but the rescue operation cannot be performed without clarifying Thaler and Sunstein's understanding of the concept of nudge. In a concluding section, I explore the healthcare contexts in which nudges may have priority over more controlling influences and set the stage for next chapters of my dissertation.

Let me add a word of caution. This chapter focuses on the nature and function of nudges, regardless of whether they benefit the recipient, third parties, or the nudger. Therefore, I will not address here worries about paternalism that have attracted much

---

1 Examples include nudges promoting vaccination programs that do not benefit the person getting immunized.
attention in the nudge debate but return to this problem in chapter 6. Many believe nudges and Thaler and Sunstein's “libertarian paternalistic” interventions are co-extensive. This is not the case, although there are overlapping problems. The confusion arises in part owing to Thaler and Sunstein defending libertarian paternalism (Sunstein and Thaler 2003), before they coined the term “nudge” (Thaler and Sunstein 2008). The editor of the *Nudge blog* makes the distinction between nudge and libertarian paternalism explicit:

> It’s important to point out that nudging complements a libertarian paternalism outlook about public policy, but the two are distinct concepts. Libertarian paternalism is intended as means to help people make decisions that make them better off as defined or judged by themselves – not by a government or private authority. While the nudges cited in the book are intended to do exactly this, nudging takes place in [a] variety of realms where the nudger’s explicit goal is to promote [the nudger's] own welfare (think of almost any consumer marketing strategy or retail store layout). (Balz May 1, 2008)

Libertarian paternalism is a justificatory strategy for a subset of nudges satisfying two conditions: (1) they are performed for the benefit of the nudgee; (2) they satisfy an informed desire welfare criterion. *Generic Medication*, for instance, does not satisfy the first condition (because it is primarily designed to contain the cost of healthcare), and therefore cannot be supported solely on paternalistic grounds.

1. **Conditions for the Preservation of Freedom of Choice**

To quality as a nudge, an influence must preserve freedom of choice, but there are numerous conceptions of what freedom of choice entails (Carter 2004). The selection of the adequate conception depends on the kind of moral work we want the concept of nudge
to do for us. I take it that in the policy contexts this concept has been discussed, its moral role is to preempt concerns about interference with choice.

**The Choice-Set Preservation Condition**

Given this function, at least when we have alternative courses of action open to us, an influence leaves us free to choose. Can we go further and claim that an influence preserves freedom of choice if and only if those influenced have unlimited choice, or even the most extensive set of feasible alternatives? The answer is no. As Alan Wertheimer notes, we “always choose from among a limited set of options” (Wertheimer 1987, 10). So long as we interact with one another, the actions of others will affect the range of options that are open to each of us at any time either because others intentionally restrict our choices, or because restrictions on our choices are the unintended consequences of others’ choices. It seems too high a bar to claim that an influence preserves freedom of choice if and only if the influencee has unlimited choice, or even the most extensive set of feasible alternatives. The choice-set must be preserved in a relative sense:

**The Choice-Set Preservation Condition.** $A$ preserves $B$’s choice-set when the choice-set is unaltered or expanded, compared to a baseline representing $B$’s situation prior to $A$’s influence attempt. (Cf. Thaler and Sunstein 2008, 15)

One may object that we need to measure the availability of options in order to make any choice-set preservation claim, but that such measurement is deeply problematic. In line with Isaiah Berlin, I argue in response that we can reasonably aim at an approximation of options, rather than precise quantification. Berlin persuasively argues that “the method for
counting [alternatives] can never be more than impressionistic; possibilities of actions are not discrete entities like apples, which can be exhaustively enumerated” (2002, 177, n. 1). The problem of individuation of choice raises an obstacle to more ambitious plans.²

To illustrate the point Berlin makes, consider the following pair of cases:

**Toothpaste Choice.** $A$ reduces from 30 to 20 the number of toothpaste brands $B$ may choose from.

**University Choice.** $A$ reduces from 30 to 20 the number of universities $B$ can decide to attend.

I take it that in **Toothpaste Choice** $B$’s choice is still preserved, though $B$ has, *prima facie*, fewer choices. By contrast, in **University Choice**, we intuitively feel that $A$ has reduced $B$’s choices. One of the reasons is that there are qualitative differences between the two sets of options, the range of University choices is more consequential for leading one’s life than the range of toothpaste brands. But there is also a *quantitative* difference between these two cases. This is because in **University Choice** “fecund” options are restricted. Joel Feinberg defines fecund options as those leading to other options (for instance, career prospects) (1984, 208). Restricting those upstream choices has consequences on an unspecifiable number of branching downstream choices. “Limited” options are not gateways to other options (toothpaste brand choices are the endpoint of a decision tree). This is why restricting fecund options has a greater impact on the overall quantitative

---

² For a discussion of related points, see Dan-Cohen (2002, 127) and Caplin (2012).
estimate of our choices than restricting limited options.³

Though I believe we can roughly compare choice-sets, I am skeptical about attempts to develop formal methods for precisely quantifying choices. Enumerating options and weighing the type of restriction or costliness of choice within each competing choice-set is the type of analytic approach a number of philosophers and social choice theorists advocate (Sugden 1998, Carter 1999). Given how hard it is to track down the ramifications of fecund options, I adopt a more holistic approach.

Thaler and Sunstein endorse the Choice-Set Preservation Condition, and sometimes explicitly reject the libertarian view that freedom of choice entails “maximizing unfettered liberty of choice” (Sunstein and Thaler 2003, 1162, footnote 11). Elsewhere, they make nominal efforts to accommodate libertarians, for example by defending libertarian paternalism. But conceptually, unlike true libertarians like White (2011), Thaler and Sunstein do not claim that influences over choices should be minimized, or that only influences that have been expressly consented to preserve freedom of choice. I agree with them. We have no reason to adopt a libertarian standard for the preservation of freedom of choice. We may care for freedom of choice without thinking that every human interaction is an assault on freedom and human dignity unless it has been expressly consented to. Such a standard would rule out the most trivial human informal practices. We cannot preemptively make every influence explicit; the role of critical dialogue is precisely to bring out influences whose presence or absence bothers us. I maintain in the final section of this chapter that the libertarian view would fail to protect liberties that matter most, treating all liberties as though they were on a moral par.

³ University Choice is an example of partial choice elimination (A preemptively removes the possibility to φ from B’s choice-set).
Thaler and Sunstein claim that, “[t]o count as a mere nudge, the intervention must be easy and cheap to avoid” (2008, 6, see also 237). They gesture toward an understanding of easy avoidability in terms of the cost of a choice. However, they do not provide an exhaustive typology of costs, which in the economics literature is rather open-ended. Subcategories of cost include material, financial, physical, transactional, opportunity, temporal, deliberative, and cognitive costs. To be easily avoidable, the cost of rejecting the nudger’s preferred option should be trivial, that is, null or low. In addition, Thaler and Sunstein confess they “do not have a clear definition of ‘easily avoided’” (2008, 248-249). I suspect this comment hints at the fact that Thaler and Sunstein are aware that the influencee might not be in a position to easily opt out of the arrangements they favor, because nudges often unconsciously alter the way we perceive the options we are offered, to make them seem either more attractive or more repulsive. My suggestion is therefore to revise our understanding of what a nudge is by making explicit this second condition for the preservation of freedom of choice that I shall call “substantial noncontrol,” to guarantee that the influencee can easily resist the influencer.

The Substantial Noncontrol Condition

Choice preservation is a necessary but not a sufficient condition for preserving freedom of choice. An influence attempt preserving freedom of choice should not pressure

---

4 Thaler and Sunstein write,

If steps are taken to increase people’s cognitive effort—as by placing fruit at eye level and candy in a more obscure place—it might be said that the “cost” of choosing candy is increased. Some of our nudges do, in a sense, impose cognitive (rather than material) costs, and in that sense alter incentives. Nudges count as such, and qualify as libertarian paternalism, only if any costs are low. (2008, 8).
us to perform any action against our will. The influence should be *noncontrolling*.  

Ruth Faden and Tom Beauchamp elaborated such a theory in the context of informed consent. Robert Nelson, Tom Beauchamp and colleagues have recently expanded this theory (2011), and have partly operationalized it into an empirical measure of the perception degrees of noncontrol and self-control, dubbed the “Decision Making Control Instrument” (Miller et al. 2011). I propose to build on their work beyond the context of informed consent.

Faden and Beauchamp start their exploration of this issue by providing a counterfactual definition of control (I quote the authors):

> Person A's action [fully] controls an action X of person B if A gets B to do X through irresistible influences that would work even if B, left to his or her own ends, in no way wanted to do X. (1986, 258)

When one is subject to fully controlling influences, one is dominated “by the will of another.” According to this analysis, coercion is always controlling. By contrast, persuasion is never controlling (since one willingly accepts a belief as one’s own because of the reasons the persuader has offered).  

But there is a third, very broad category of

---

5 In “Choice, freedom, and freedom of choice,” Ian Carter attempts to clarify freedom of choice in an overlapping but not identical way. For him, “A person has freedom of choice if she lacks constraints on the reasoned selection and performance of one or more items on the action-menu” (2004, 69). I am not convinced that freedom of choice necessarily involves *reasoned* selection, that is, choosing rather than picking (see Ullmann-Margalit and Morgenbesser 1977).

6 I do not claim that the controlling-noncontrolling continuum exhaustively characterizes these influences, or that influences should always be classified in this manner no matter what criterion we use. For instance, Alan Wertheimer claims that coercion is characterized by two necessary and jointly sufficient conditions: (1) the
influences that can be situated on a continuum. Faden and Beauchamp call “manipulation”
the forms of influence that do not use the means typical of coercion (threats) or persuasion
(reasons, arguments) and in which there are degrees of control or noncontrol. Manipulatory
influences are more or less difficult to resist. Certain offers, a subset of manipulative
influences for Faden and Beauchamp, are even welcome. This characterization is
conceptually helpful, though I will not use the language of manipulation to characterize
influences less controlling than coercion but more controlling than rational persuasion
because I believe the term “manipulation” carries a negative valence that will obscure my
argument. Faden and Beauchamp, by contrast, use this term in a neutral, non-derogatory,
and non-normative sense.

Faden and Beauchamp claim that influences can be situated on a continuum from
fully controlling to fully noncontrolling. There are, however, two thresholds on the
continuum: some influences are substantially controlling, others are substantially
noncontrolling. Chart 1 illustrates their view:

---

threat prevents “exercising free will and judgment” and (2) is wrongful (1987, 30). The continuum I use
classifies coercion only according to the first condition, but does not reject the second condition. Hence, I
believe the criterion for external noncontrol (an easy resistibility test) need not rely on any implicit moral
assumption (contra Bloche1996, 71). That is, the question of noncontrol and easy resistibility is different
from the degree of pressure a person ought to resist. In this dissertation, by Coercion, I mean issuing a threat
to make someone worse off if that person refuses to comply with one’s demand (Wertheimer 1987). One
example of coercion is Medical Abandonment: Physician-investigator A threatens to abandon patient-
subject B if B refuses to enroll in A’s clinical trial.

7 My own taxonomy contains coercion, compulsion, elimination of choice, at one end of the continuum,
rational persuasion at the other, and several types of interventions situated on the continuum, such as
incentives and disincentives. See section 4 for more details.
Notice that the thresholds are signaled by dotted lines and the territory of substantially controlling or noncontrolling influences is shaded and expands around the thresholds, not after them. This is because one cannot establish clear-cut demarcations on the continuum.8

8 For Faden and Beauchamp, substantial noncontrol is one of three conditions of autonomy, thus a substantially autonomous action needs to be performed without substantially controlling influences. They claim that an action is autonomous if it is (1) intentional, (2) performed with understanding, (3) is not subject to controlling influences (238). Notice that they replace the language of voluntariness by that of noncontrol: “[w]e substitute a conception of noncontrol that does not have the history and connotation that burdens the terms ‘freedom,’ ‘voluntariness,’ and independence” (1986, 257). Nelson et al. have recently reintroduced the language of voluntariness in a similar theory. They analyze the concept of voluntariness in terms of two necessary and sufficient conditions: intentionality and substantial noncontrol (2011). According to this theory, an autonomous action is voluntary and performed with understanding. These disputes have no
Faden and Beauchamp define full control (1986, 258), but not substantial noncontrol, which I suggest to characterize as:

**The Substantial Noncontrol Condition.** $A$’s influence to get $B$ to $\phi$ is substantially noncontrolling if $B$ can easily resist $A$'s attempt to get her to $\phi$.

To summarize, an influence preserves freedom of choice if and only if it preserves the choice-set and is substantially noncontrolling. Critics might reply that, libertarianism to the side, I should also add the preservation of liberty and autonomy to these conditions. There are, however, good reasons for avoiding the introduction of liberty and autonomy, but also voluntariness, to explicate the preservation of freedom of choice.

---

9 Choice-set preservation and substantial noncontrol apply to nudges for the following reason. From the psychological perspective that behavioral economists adopt, choice sets cannot be exclusively defined in terms of their formal properties or quantity (e.g., \{a, b, c, d\}). For instance, the ordering of options in a choice-set is a formal property that does not give us reasons for a particular choice as long as it does not introduce any hierarchy. But we know humans usually think differently: psychologically, any ordering matters. Some choices may go unnoticed; others are singled out (salience effects). Choice-sets are percepts. Hence, what matters is how the influencee perceives her set of options. As the way I perceive my options can deeply affect my evaluation of my options, my deliberative capacities may be positively or negatively affected by behavioral triggering interventions. One can have substantial control over another person’s deliberation just by changing the way he or she perceives the choice-situation and the weight of each option.
Liberty and autonomy

Autonomy and liberty are tightly connected concepts. By not using them in the definition of nudges, I avoid the ambiguities and conceptual slips that plague the nudge literature. For instance, Thaler and Sunstein explain the preservation of freedom of choice in terms of the preservation of liberty (2008, 5). But liberty is a broader and normatively richer concept than choice-preservation (as defined above), which is narrow and descriptive.

Claiming that an influence preserves freedom of choice if and only if it is liberty preserving (rather than choice-preserving) and autonomy preserving (rather than substantially noncontrolling) raises other problems because the articulation between autonomy and liberty is a matter of dispute. For some, liberty and autonomy are clearly distinct concepts because “self-determination [that is, autonomy] can be limited without limiting liberty” (Dworkin 1988, 105). For others, “autonomy” is used to explicate liberty, or at least one of its dimensions. Take G. C. Jr. MacCallum’s classic characterization of liberty as a triadic relation: (1) an agent is (2) free from certain constraints, (3) to do or be certain things (1967). Depending on the type of constraints and their source, some will mean by “liberty” what others call “deliberative autonomy”: (1) an individual’s choices are not (2) substantially controlled by involuntary internal factors or external factors that distort, channel, or put pressure on our cognitive capacities for (3) making decisions. This is why I do not equate the preservation of freedom of choice with the preservation of liberty and autonomy.

Given these considerations in favor of putting aside liberty and autonomy, the onus is on those who wish to define the preservation of freedom of choice in terms of liberty or autonomy to explain why they are dissatisfied with choice-set preservation and substantial
noncontrol. Adopting the view I suggest does not commit us to the contentious claim that the preservation of freedom of choice is a sufficient, or even a necessary, condition for the justified use of nudges. Nonetheless, whether an influence preserves the choice-set and is substantially noncontrolling matters for the moral evaluation of actions and policies, if we care about the degree of control exerted by others over at least some of our choices, actions, and preferences.

**Resistibility and easy resistibility**

Critics may also object that substantial noncontrol is an unhelpful condition because it cannot be readily ascertained. In this section, I explore the criterion of easy resistibility that is at the core of substantial noncontrol. Here again, Faden and Beauchamp’s work on resistibility is a useful starting point. They note that the capacity to resist an influence is subjective: it depends on each person’s psychological vulnerabilities. Resistibility is a criterion for testing the degree to which an influence is controlling. For public policy purposes, Faden and Beauchamp defend an “objective” interpretation of easy resistibility, which depends on the predictable reaction of the “average (or normal, reasonable, etc.) person” (1986, 260). But the authors rely heavily on the concepts of resistance and resistibility without offering an analysis of these concepts, a lacuna they readily

10 Faden and Beauchamp contrast the objective interpretation of easy resistibility with its objective interpretation:

A related problem is that a criterion of “subjective irresistibility”, like any subjective criterion, is difficult to implement in practice. We offer no suggestions as to how to measure or test whether a given threat would prove irresistible to any particular individual. Such assessments doubtless have to be made for policy purposes on the basis of evidence and predictions about how most people would respond to a given threat under the circumstances—an objective criterion—and how the individual in question may differ in relevant abilities, life experiences, and values from most people. (1986, 342)
anticipate (1986, 360-361). To supplement their view, I define these concepts and
distinguish resistibility from easy resistibility:

**Resistibility.** A’s influence on B is resistible if and only if B can oppose the
pressure to get her to φ.

**Easy Resistibility.** A’s influence on B is easily resistible if and only if B can
relatively *effortlessly* oppose the pressure to get her to φ.

These definitions are consistent with, and inspired by, the psychological
literature.\(^\text{11}\) What distinguishes resistibility from easy resistibility is the amount of effort
the influencee needs to exert in order to oppose the influencer’s pressure to make it more likely that she will φ. If my explication of freedom of choice is compelling, the next task is
to provide theoretical and empirical evidence to the effect that at least some influences activating automatic cognitive processes (to use Thaler and Sunstein's words) are easily resistible.

\(^\text{11}\) For instance, Eric S. Knowles and Jay A. Linn write in the introduction of an edited volume entitled, *Resistance and persuasion* that,

Resistance has acquired a dual definition in psychology. On the one hand, it defines an outcome: the outcome of not being moved by pressures to change. On the other hand, it identifies a motivational state: the motivation to oppose and counter pressures to change… To some extent the outcome and motivational definitions are theoretically linked. A motivation to oppose would promote the outcome of not changing. However, the two definitions are also not completely overlapping. (2004, 5)

My understanding of easy resistibility brings together recent research in the psychology of persuasion (in the sense of influence), the psychology of attention (pre-attentive monitoring processes), and dual-process theories of decision-making (capacity for inhibition).
3. Easily Resistible Nudges

Let me first return to the way nudging mechanisms have been characterized. It is assumed that these influences activate automatic cognitive processes that always bypass deliberation. But in chapter 1, I argued that nudging mechanisms involve “shallow cognitive processes,” not “automatic cognitive processes” (in the sense of System 1 processes). This amendment to Thaler and Sunstein's theory matters because the conditions for easy resistibility vary depending on whether deliberation is entirely bypassed (through nondeliberative cognitive processes) or partly bypassed (through incompletely deliberative cognitive processes). Whether and how an influence engages our deliberative capacities is also a consideration relevant to the moral evaluation of actions and policies, although I argue in chapter 6 that there is no duty to always engage or enhance others' deliberative capacities.

The point I wish to make in this section is that there is sufficient psychological evidence to believe that influences that bypass deliberation and are covert (that is, unannounced) may still be easily resistible.\(^\text{12}\) These influences raise particular concerns for

\(^{12}\) Remark that influences activating nondeliberative cognitive processes may satisfy the easy resistibility criterion when they are *overt*, that is, when the influencee is told ahead of time that she will be steered toward *qing*. She will then have the opportunity to pay attention to her choice and figure out if she accepts the nudger’s guidance or not. But even then, she may quickly go off guard, and simply rely on her shallow cognitive processes that will be cued by her choice environment arranged by others. I thank Richard Ashcroft for bringing this point to my attention.
those who believe the lack of conscious deliberation undermines control over one's choices and therefore compromises genuine freedom of choice.

Often influences proceeding primarily through nondeliberative cognitive processes are covert (typically Cafeteria). I argue that those influences may satisfy the conditions necessary for easy resistibility. I also maintain that influences primarily relying on the activation of incompletely deliberative cognitive processes may also satisfy the conditions for easy resistibility, though in a different manner.

To reformulate the question at hand, what are the conditions necessary for easy resistibility? Note that the locution “easy resistibility” might be somewhat misleading because it seems to erroneously describe a characteristic of an influence *per se*, considered acontextually. Being able to easily resist an influence is not a characteristic inherent to the influencee either, irrespective of the context that enables, facilitates or frustrates the exercise of certain capacities. In my view (and I believe, in Faden and Beauchamp's view), easy resistibility is a concept that designates the interface between the influencee's psychological makeup and the force of the influence others exert on her in a particular context.

I maintain that *B* can easily resist *A*’s influence if and only if *B* is not subject to influences, or put in circumstances, that would either significantly undermine the relatively effortless exercise of the capacities that enable *B* to oppose *A*’s pressure to get her to φ, or weaken those very capacities. My use of the language of capacities is not meant to be technical and designates the mindware that enables people to process information in certain ways in order to do certain things. One exercises those capacities when one actually processes information for a particular end.
The empirical literature in cognitive science identifies at least two sets of cognitive processes necessary for such form of control: (1) the capacities for goal-conflict recognition, and (2) the capacities for goal-conflict resolution (Stanovich 2010). Let me explain each in turn before I examine circumstances or influences that are susceptible to negatively impact those capacities or their exercise. These conditions are necessary but perhaps not sufficient for easy resistibility.

**Goal-conflict recognition: Conscious and nonconscious**

Contemporary theories of self-control (in a broad sense) converge in models that distinguish goal-conflict recognition and resolution. Their aim is to describe how and when lower-level cognitive processes are overridden by higher-level cognitive processes, enabling the individuals to control their behavior and optimally allocate their finite cognitive resources.

There are various labels for, and models of the capacity for goal-conflict recognition. But goal-conflict recognition always involves processes that monitor the output of various lower-level cognitive processes to detect a conflict or a mismatch between the agent's explicit or implicit goal (broadly understood) and her behavioral response (Stanovich 2011, 79). The recognition of a conflict in one's internal states during information processing leads to strategic adjustments. Examples include situations in which nondeliberative cognitive processes fail to smoothly lead to an adaptive response to our environment. This is the case when we ride a bike on autopilot and suddenly notice an unusual problem on the road that cannot be addressed if we rely on our automatic skills (nondeliberative cognitive processes consolidated through learning). Sometimes reliance on heuristics (incompletely deliberative cognitive processes) to solve a problem does not
lead to a definitive answer to the problem we try to address (our goal), and therefore goal-conflict recognition sends a signal to initiate higher-level processes, such as more formal and systematic reasoning (if we are lucky enough to have the requisite mindware) (Stanovich 2010, 76). Importantly, goal-conflict recognition processes can be either conscious (Thompson 2009) or nonconscious (Suhler and Churchland 2009; Faucher 2012).

Let me start with conscious processes. There are several models of goal-conflict recognition processes that detect an anomaly, and bring the anomaly to our conscious attention. For instance, according to Valerie Thompson, “stimuli that produce a strong Feeling of Rightness [i.e., an intuition that the decision is correct] are not likely to be further scrutinized,” whereas those that produce a feeling of dysfluency are more likely to trigger scrutiny (Thompson 2009, 187). Those feelings then activate internal representations of goals and intentions, etc. There is a large, though dispersed, body of literature marshaling evidence that at least when people already have strong and settled enough preferences, aims, or beliefs, they are likely to become aware of an anomaly, that is, recognize a conflict and attempt to consciously resist the influence attempt (Knowles and Linn 200; Sagarin and Cialdini 2004, 262). At least when there is sufficient conflict between the influencer’s and the influencee’s goals, the influencees’ goal-conflict recognition processes might be activated, and the influencee may be in a position to deliberate over her choice and resist the attempt to modify the course of her behavior. Those who maintain that control over our choices can only be the result of conscious deliberation have therefore no reason to believe that nudges are always incompatible with

---

13 In Stanovich's theory, for instance, what he calls preattentive monitoring processes monitor the output of various cognitive processes (see chart 1.2).
freedom of choice, as they understand it. But of course, we do not have to agree with them that control needs to be conscious. This is the point I argue next.

According to some cognitive scientists, certain goal-conflict recognition processes can be entirely nonconscious. As Christopher Suhler and Patricia Churchland write, “nonconscious processes can support a robust form of control and, by extension, [...] consciousness is not a necessary condition for control” (2009, 341). These processes are flexible and efficient in performing various functions, including goal-pursuit in the face of distractions or obstacles in the environment. Experiments in nonhuman animals (such as rats) have, for instance, shown that mammals can delay gratification by pursuing a distant goal and inhibiting their prepotent responses to a stimulus, without relying on conscious deliberation.\(^\text{14}\)

A good example of this kind of nonconscious goal-conflict recognition in humans comes from the literature on stereotyping. One of the most famous (and saddest) results in

\(^\text{14}\) For an overview of the empirical literature, including the anatomical (brain regions and pathways) and physiological (molecular mechanisms) dimensions of self-control, see Suhler and Churchland (2009) and Faucher (2012). Suhler and Churchland defend a view about the conditions for robust nonconscious self-control and goal pursuit that converges with my views about the capacities enabling easy resistibility. They write,

The implications of the neurobiological account of control developed here for the interpretation of social psychological results can be summarized as follows: quite simply, most of the patterns of behavior described in the social psychology literature do not fall outside the realm of control (see also Ref. [3]). The reason is that although the effects studied by social psychologists are mediated by situational factors and (often) by nonconscious processes, evidence indicates that the requirements for control set out earlier are typically met. First, the brain structures essential for control functions are intact (the anatomical condition). And second, the circumstances studied are usually within the typical range encountered in the evolutionary past of humans, and thus, levels of various neurochemicals on which the proper functioning of the anatomical structures depends can reasonably be expected to be within their normal ranges (the physiological condition). (2009, 346)

They also note that factors like stress have distinct effects on the physiological states necessary for self-control (344).
social psychology is that we are prone to automatic racial stereotyping. For instance, due to implicit categorization and stereotypes even individuals who claims they have no racial bias are still more likely to mistake a harmless object for a gun when a Black person holds it than a White person. This effect is produced when subjects are asked to make a snap judgment, when time-pressure disables the capacity to consciously deliberate over one's choices. These automatic stereotypical associations are hard to eliminate even when people would prefer not to have them. But Brandon Stewart and Keith Payne (2008) propose an original method that can help to correct negative stereotyping. It consists of keeping in mind a counter-stereotypical intention and to implement it by harnessing automatic cognitive processes. The technique consists of associating a counter-stereotypical sentence with a stimulus whose presence will activate the implicit goal of negative stereotype avoidance:

According to [our] theory, an if-then format links an environmental stimulus with a specific response (e.g., when I see a Black face, I will then think “safe”). The creation of this link increases the accessibility of both the environmental stimulus and the specific thought or intention to which it was linked. (2008, 1337)

Importantly, the intention to which the stimulus is linked remains unconscious when one makes the snap judgment. The idea here is that by habituating myself to other automatic associations, I am able to regain self-control and not be as much affected by irrelevant situational factors. My counter-stereotypical stance may therefore remain resilient even in the face of deliberate attempts to influence me in adopting racist stereotypes. This is, I think, a common experience in many situations of resolute choice, when I have not only the intention to change my behavior (e.g., my eating habits) but also have taken steps to
implement my resolution through the benefits of forming habits.\textsuperscript{15} For instance, Robert Nozick proposed to adopt a general principle of behavior to change unwanted habits:

The mark of a principle ("Never eat snacks between meals"; "Never smoke another cigarette") is that it ties the decision whether to do an immediate particular act (eating this snack, smoking this cigarette) to the whole class of actions of which the principle makes it part. This act now stands for the whole class. By adopting the principle, it is as if you have made the following true: if you do this one particular action in the class, you will do them all. (1993, 17)

This strategy belongs to the family of strategies instantiated in the counter-stereotypes example: Achieving self-control through the use of automatized habits, rather than through deliberation. At some point, deliberation is no longer necessary for (and may even undermine) resolute choice; control is achieved through nonconscious monitoring of one's environment to detect features of one's environment or influences that might lead us astray from the smooth pursuit of our wholeheartedly endorsed goals.\textsuperscript{16}

\textsuperscript{15} For a compelling study of this issue in the medical context of therapeutic observance, see the work of French physician and philosopher Gérard Réach (2007; 2008).

\textsuperscript{16} One possible objection against my view is that when we do not care about φing, we do not even need to exert resistance in the face of an influence attempt. I think the objection covers two distinct cases. In case 1, I am indifferent between φing and not φing but disposed to do either. In that case an influence relying on our shallow cognitive processes is perhaps not easily resistible if I am not aware that it takes place and my goal-conflict recognition processes are not activated. I would need to exert some form of resistance in order to not be likely to φ. In case 2, I am indifferent between φing and not φing but not disposed to φ or not φ (e.g., buy a car of brand X rather than Y). I simply don't care about φing or not φing. In the second case, I do not need to oppose any resistance because from my perspective, I am not being subject to any influence at all.
**Goal-conflict resolution: Inhibition**

Naturally, the above-mentioned examples do not solely require conscious or nonconscious goal-conflict recognition but also goal-conflict resolution. I refer here in particular (though not exclusively) to our capacity to inhibit our propensity to do what the influencer wants us to do. Inhibition is the capacity to stop a cognitive process once it is triggered (Stanovich 2011) and it is a necessary component of resistance. Some call “willpower” any type of inhibitory capacity (Baumeister and Tierney 2011), but willpower primarily refers to impulse control, a subset of inhibitory capacities.\(^\text{17}\) These capacities also include the capacity to inhibit erroneous but spontaneous forms of reasoning, for instance (Stanovich 2011).\(^\text{18}\) Sometimes we are in a position to easily resist an unwanted influence (by recognizing a conflict and inhibiting a response), but also need to find a better solution to the problem at hand. For instance, suppose that a political campaign aims at discrediting a candidate by spinning statistics. You notice that there is something not quite right there and you are able to easily inhibit your propensity to believe what you are told. However, you are still left with a problem if you don't know enough about statistics to come up with a better way to process the statistical data (e.g., calculus involving base rates), and you

---

\(^{17}\) Baumeister uses the term willpower to cover any type of inhibitory process, disregarding potential functional and neurological differences between those processes. The psychological literature on inhibition is connected to the neuroscience research on executive function.

\(^{18}\) Though psychologists and neuroscientists have started studying inhibitory capacities, there is no general theory about the nature of those capacities, and whether they all function similarly. All we know is that there are many of them, and that there are large interindividual differences with regard to them (Stanovich 2011). But, in practice, until we get a more precise picture of patterns of interindividual differences, we can assume that all are endowed with those capacities to some degree, unless they have a specific impairment (due to mental disease, a genetic condition, or injury).
may fall back on using a misleading heuristic and form a false belief or an unjustified true belief.

The exclusion criterion for the absence of easy resistibility

From a practical perspective, there is fortunately an exclusion criterion ruling out influences that are not easily resistible: the influence should not create circumstances that weaken those capacities, undermine their exercise, or attempt to steer away an individual who is already in a cognitively strained situation. To qualify as “relatively” effortless, the influencee should not be expected to have acquired unusually high skills enhancing her goal-conflict recognition or resolution capacities. In the context of influences exercised on groups of individuals, effortlessness is relative to what we can expect of normal individuals (within that group) with limited inhibitory resources that can be depleted (Baumeister and Tierney 2011). There is accumulating data about circumstances that undermine people’s goal-conflict recognition and resolution capacities. Typically, the main factors are time pressure, stress (and stress-inducing poverty), submission to perceived authority, fatigue, anxiety, cognitive load, and distraction (Knowles and Linn 2004; Boush et al. 2009; Suhler and Churchland 2009; Baumeister and Tierney 2011; Stanovich 2011).

19 For instance, we know that obedience to authority and perceived social norms may numb our capacity to detect the most striking gap between what we are told to do (and actually do), and what we normally believe we should be doing. I leave the exploration of this important issue for another time.

20 For an overview of the literature on those skills, see Boush et al. (2009, 162). I return to this problem in chapter 6. Some will argue that instead of nudging individuals, consumers should be enabled to acquire those skills if they wish, and to make use of those skills when they deem necessary.
To see how the criterion of easy resistibility is serviceable for differentiating substantially controlling influences from substantially noncontrolling ones, consider:

**Pharmaceutical TV Ad.** In a TV ad for medication, the list of side effects is read in a monotonous voice while displaying images of butterflies and happy people.

This ad activates nondeliberative cognitive processes by providing the mind with enjoyable and attention-grabbing stimuli irrelevant to the audio disclosure. Its strategy is to confuse the audience through the association of those stimuli with positive affects, rather than the negative affects typically associated with worrisome side effects.21 This influence is resistible, but it is not easily resistible. It is not a nudge.

What about influences that involve incomplete deliberation, relying on mental shortcuts to cause the endorsement of a belief, the formation of an intention, or the alteration of a preference? Unlike influences activating nondeliberative cognitive processes, these influences have informational content. If an influencer weakens the

---

21 Boush et al. discuss these “dazzling” tactics (2009, 44, 48). In connection to these techniques, a session was organized on “Magic, the Brain, and the Mind” at the 39th annual meeting of the Society for Neuroscience held in Chicago (2009). Professional magicians Apollo Robbins and Eric Mead demonstrated their ability to divert attention of an audience of neuroscientists who know the mechanisms for fooling the brain inside out and were aware they were going to get tricked. On the neuroscience of magic, see Macknik and Martinez-Conde (2010).
required capacities or undermines their exercise by recourse to deception,\textsuperscript{22} concealment,\textsuperscript{23} or misrepresentation\textsuperscript{24} of material information,\textsuperscript{25} the influence is not easily resistible.\textsuperscript{26}

\textsuperscript{22}E.g., Deceptive Guards: Guards bring a prisoner to falsely believe his cell door is securely locked (by mimicking door-closing motions), whereas it is in fact unlocked, and he could walk out of the cell any time (Dworkin 1988, 105).

\textsuperscript{23}E.g., Phase 1 Clinical Trial: A, an investigator, does not provide B, a research participant, with information about the risks of enrolling in a phase 1 randomized clinical trial.

\textsuperscript{24}E.g., Tobacco Doubt Inoculation: Tobacco company A designs a marketing campaign inoculating public B with doubt about the scientific validity of research on the health risk of their products by cherry-picking incomplete scientific results (Boush et al. 2009, 72).

\textsuperscript{25}I join Faden and Beauchamp who maintained a similar position with regard to the “manipulation of information.” However their understanding of this category and their reasoning to reach this conclusion are different from my position. Faden and Beauchamp’s taxonomy of manipulation is not based on the psychological process involved. Manipulation is divided into (1) the manipulation of options (incentives and disincentives), (2) the manipulation of information (targeting cognitive processes without recourse to rational persuasion), (3) psychological manipulation (targeting affective mental processes, not cognition). (Note that Nelson et al. (2011) have dropped the category of psychological manipulation from their list.) The reason Faden and Beauchamp advocate the view that the manipulation of information is compatible with autonomy only if there is sufficient understanding is that they believe the easy resistibility criterion does not apply because the influencee is unaware of the influence attempt. Their belief that awareness is a necessary condition for easy resistibility, and that it is sometimes absent, leads them to struggle with squaring understanding with noncontrol. A further problem with their position is that they pursue two projects at once: a. determining conditions under which an action is autonomous (it requires understanding, whether one is under the influence of others or not); b. determining conditions under which an influence is noncontrolling.

When discussing informational manipulation, their position is ambiguous: they might claim either that the easy resistibility criterion is the only criterion allowing to test for noncontrol, and in that case, it simply does not apply to manipulation of information (362), or they maintain that understanding is an alternative to easy resistibility, and therefore minor manipulation of information is “compatible with substantial noncontrol and
For instance, **Less Than You Think** qualifies as a nudge, but a misleading campaign inflating or spinning the number of students who do not often binge-drink would not be conducive to letting them make their own decision (assuming for the sake of argument that students minimally trust their university's health authorities in the first place). This is because students are less likely to be critical of the information they are exposed to when they expect the authority communicating with them to be truthful. An influence’s not being generally *easily* resistible does not mean that it is not resistible at all. Some individuals may even easily detect the misleading or deceitful content of the influence and may be in a position to easily resist it. Recall that when a policy applies to groups of individuals, easy resistibility is assessed for the average person. Whether an information-rich influence is easily resistible depends on how skillful the target population is. Mathematicians are more likely than the rest of us to easily resist an influence attempt based on misleading statistics and get it right. In certain circumstances, we may be responsible for increased vigilance based on the realistic expectation that the influencer may use means that weaken our capacities required to easily resist him or her, or undermine the exercise of those capacities.\(^{27}\)

---

26 See Robert Proctor and Londa Schiebinger’s work on “agnostology,” the study of culturally induced ignorance or doubt (2008). I'm grateful to Daniel Andler for having introduced me to this literature.

27 I join Faden and Beauchamp who maintained a similar position with regard to the “manipulation of information.” However their understanding of this category and their reasoning to reach this conclusion are different from my position. Faden and Beauchamp’s taxonomy of manipulation is not based on the
To put the point differently, my claim here is not that a deception-based influence is never easily resistible (Dworkin forthcoming). Rather, it is that by using deception the influencer attempts (successfully or not) to undermine the exercise of the capacities enabling easy resistibility. Some individuals will be able to resist even the most cunning forms of deception, but what is important when it comes to cases such as Asparagus-Lovers is the predictable effect of the influencer on the average person (what Faden and Beauchamp call the objective interpretation of easy resistibility).

psychological process involved. Manipulation is divided into (1) the manipulation of options (incentives and disincentives), (2) the manipulation of information (targeting cognitive processes without recourse to rational persuasion), (3) psychological manipulation (targeting affective mental processes, not cognition). (Note that Nelson et al. (2011) have dropped the category of psychological manipulation from their list.) The reason Faden and Beauchamp advocate the view that the manipulation of information is compatible with autonomy only if there is sufficient understanding is that they believe the easy resistibility criterion does not apply because the influencee is unaware of the influence attempt. Their belief that awareness is a necessary condition for easy resistibility, and that it is sometimes absent, leads them to struggle with squaring understanding with noncontrol. A further problem with their position is that they pursue two projects at once: a. determining conditions under which an action is autonomous (it requires understanding, whether one is under the influence of others or not); b. determining conditions under which an influence is noncontrolling. When discussing informational manipulation, their position is ambiguous: they might claim either that the easy resistibility criterion is the only criterion allowing to test for noncontrol, and in that case, it simply does not apply to manipulation of information (362), or they maintain that understanding is an alternative to easy resistibility, and therefore minor manipulation of information is “compatible with substantial noncontrol and informed consent” (263). In a related manner, Miller et al.’s Decision Making Control Instrument does not ask participants any question about whether they judge their understanding of the ramifications of the consequences of alternative options in the choice situation to have been undermined by the influence of others, and as a result, whether the influencer had substantial control over their decision (see their questionnaire 2011, 735-736).
Compare a situation in which an influence activates incomplete deliberation with an attempt to persuade us rationally:

**BSE Prevention Campaign.** A health campaign attempts to create a positive attitude toward breast self-examination (BSE) by using a persuasive message: “By performing BSE you are able to detect breast cancer in an earlier and therefore more treatable stage” (Ruiter et al. 2001).

**BSE Loss-Aversion Campaign.** A health campaign attempts to create a positive attitude toward breast self-examination (BSE) by using a loss-aversion message: “Don’t let the chance of detecting breast cancer go!”

In both cases, the aim is to either change the influencee’s belief about the usefulness of BSE and/or to incite her to form the intention to adopt the habit of doing BSE. The first campaign is primarily based on a simple form of instrumental rationality (if one wants the health benefit, [given that $X$ is true] then one needs to do $Y$). The second campaign contains less informational content and triggers a less logically structured chain of thoughts. Loss-aversion triggers negative feelings stronger than the prospect of health gains and activates a dormant and simple decision rule or heuristic, “I should not let a chance go if I can” that it associates with a specific belief and intention.

When rational persuasion occurs, the influencee comes to believe or form the intention to $\phi$ while being in a position to *epistemically evaluate* the merits of the reasons
the persuader offers in support of his or her view.\textsuperscript{28} Criteria necessary for epistemic evaluation are complex, and an exploration of those criteria goes beyond the scope of this dissertation. For my purposes, we can simply state that the influencee needs to use a reliable form of theoretical or practical deliberation to reach her conclusion, given her values and preferences. The persuadee is then in a position to epistemically evaluate the persuader’s claims. Rational persuasion is therefore fully noncontrolling. We sometimes say that the force of an argument is normatively “irresistible,” but this is different from the

\textsuperscript{28} I take it that $A$ rationally persuades $B$ when $A$ induces $B$ to willingly believe, or form the intention to $\phi$, primarily by presenting her reasons to $\phi$. My understanding of rational persuasion is different from the use of the term “persuasion” in the interdisciplinary literature on persuasion and compliance gaining.\textsuperscript{28} For instance, Robert H. Gass and John S. Seiter (2010) use “persuasion” to cover coercing, brainwashing, indoctrinating, manipulating, compliance gaining, advising, convincing, educating, and propaganda. For instance, they write:

Persuasion involves one or more persons who are engaged in the activity of creating, modifying, reinforcing or extinguishing beliefs, attitudes, intentions, motivations and/or behaviors within the constraints of a given communication context. (2010, 33)

In the psychological literature, “persuasion” is by and large synonymous with “influence,” no matter whether the effect is produced through the use of force, threat, shallow cognitive processes, or arguments. I use other terms to cover various types of influences. I am here concerned only with “rational persuasion.” My definition is also different from other definitions proposed in the philosophical literature (cf. Faden and Beauchamp 1986, 261-262; Wertheimer 1987, 292; Beauchamp and Childress 2009, 133). To qualify as a rational persuasion, an influence attempt does not need to be exclusively based on the presentation and acceptance of reasons. Reasons have to play the primary role in bringing about change, but they are compatible with appeals to emotions. Acts of communication rarely happen without the involvement of emotional processes, especially when one is making a decision and needs to assign a value to outcomes. When rational persuasion occurs, the persuader does not control the persuadee. The influencer does not exercise any power that would undermine the voluntariness of the influencee’s change in belief or intention.
claim that the rational persuader exercises an irresistible pressure on us to change our beliefs.

The problem of the reliability of a form of deliberation raises specific concerns when mental shortcuts (heuristics or biases) are used, since they may not be optimal forms of deliberation. Whether, and under what conditions, they are reliable is open to debate. Mental shortcuts allow us to extract information (cues) and statistical regularities from the environment, and make faster, but less reason-based, decisions by using an incomplete, effortless, and fast type of deliberation.

An intervention may create or reinforce a “benign” environment where I can reliably fall back on incompletely deliberative cognitive processes to make decisions without needing complete deliberation. As Stanovich puts it, “a hostile environment for heuristics is one in which there are few cues that are usable by heuristic processes, or there are misleading cues” (Stanovich 2011, 21). Asparagus-Lovers is an example of an arranged hostile environment designed to bring about health benefits. The deceived participants are under the substantially controlling influence of the investigator, who has created an environment to misinform participants about their own taste. 29 The experimental design of Asparagus-Lovers is intended to mislead subjects into believing that no one is trying to influence their food preferences. The questionnaires subjects had to

---

29 Another study showed that merely imagining an event increased the probability one would believe the event actually took place in one’s life. The recollection holds no matter whether the event actually happened to us, whether it was a relatively frequent event making self-attribution at least plausible (“Having a milk tooth extracted by a dentist before the age of 6”), or whether the event never happened to anyone in one’s country according to medical records (“Having a nurse remove a skin sample from my little finger”) (Mazzoni and Memon 2003).
fill out were allegedly part of a study of the relationship between food preferences and personality, and researchers cleverly “planted hints to suggest that [they] were studying childhood obesity.” (Laney et al. 2008, 298). So, Luc Bovens's claim that “I can easily recognize that pressure is being exercised when I am told that I used to like asparagus” is empirically false (forthcoming).

In contrast to a hostile environment, an influence providing us with accurate information creates a “benign” environment where we can reliably trust our fast and frugal shallow cognitive processes to make sufficiently good decisions without the need for slow and resource-consuming full-blown deliberation. Reliance on heuristics might be our natural tendency or the result of intentional training. It is often a smart and rational strategy for allocating cognitive resources.

One motive for arranging the choice environment so as to guide choice is that people are sometimes unable to access and evaluate reasons for action even if the compelling reasons were offered to them as part of a rational persuasion strategy. This is a well-known problem in risk communication. Trade-offs need to be made between the accuracy of the scientific information experts want to communicate to the public and the less accurate but more graphic information that the lay public might better understand, appreciate, and act upon.\textsuperscript{30} Of course, the judgment that lay public fails to grasp complex

\textsuperscript{30} In the literature on decisional capacity, those capacities are usually divided into four subcategories: understanding, appreciation, reasoning, and choice (Charland 2011). Allen Buchanan and Dan Brock write about appreciation that “to appreciate the nature and meaning of potential alternatives—what it would be like and “feel” like to be in possible future states and to undergo various experiences—and to integrate this appreciation into one's decision making” (Buchanan and Brock, 1989, 24). Nudges might facilitate appreciation by stirring emotions in order to generate a more appropriate affective response to a risk situation, when significant but distant risk is underestimated.
scientific issues involving risk assessment depends in part on what the normative standards of acceptable risk are. However, sometimes the public may not understand and appreciate the magnitude and likelihood of adverse events, even in the absence of disagreement about levels of acceptable risk (e.g., to future generations and distant people).

For instance, in the climate change debate, presentation of possible scenarios in probabilistic terms is unlikely to be sufficiently understood and appreciated to enable citizens to engage in an informed public deliberation. Climate scientists might instead provide the public with some detailed and vivid scenarios that are more likely to engage them. But this strategy comes at a cost: it is much less accurate in the presentation of the likelihood of possible specific events, like polar bears losing their habit if the Arctic ice melts (Kitcher 2010). The problem, then, is how to create a benign decision environment that would allow the public to reliably trust their fast and frugal shallow cognitive processes without, for instance, being misguided by the activation of the representativeness heuristic that would generalize certain detailed non-representative scenarios (Kahneman 2011, 146-155). This heuristic is roughly the tendency to imagine that an event is similar in characteristics to its parent population (e.g., polar bear extinction to the effects of climate change in general). It is more accurate than chance guesses would be. The drawback is that its activation may lead us to overestimate the reliability of this mental shortcut to accurately predict the likelihood of another event. The public may conclude that these scenarios illustrate the only outcomes that matter, neglecting, for example, the probable impact of global warming on the spread of infectious diseases. Creating a benign decision environment is therefore all but easy.

To conclude, I have shown that under some conditions, influences activating shallow cognitive processes preserve freedom of choice. These conditions might be hard to
satisfy, but when they are satisfied nudgees are in control of their choices: they have a real opportunity to dissent from the nudger.

4. An Amended Concept of Nudge

The implication of this argument is an amended definition of a nudge:

**Nudge.** *A* nudges *B* when *A* makes it more likely that *B* will \( \varphi \), primarily by activating *B*’s shallow cognitive processes, while *A*’s influence preserves *B*’s choice-set and is substantially noncontrolling (i.e., preserves *B*’s freedom of choice).\(^{31}\)

Let me add a few comments to this definition. A nudge increases the *ex ante* probability that some individuals in a group, or an individual over time, will \( \varphi \). It is a probabilistic success term. This means that the success condition for a nudge is not that the nudgee actually \( \varphi \)es but that his or her chance of \( \varphi \)ing is higher under nudge conditions than under the *status quo*. For instance, it may have increased by 1% or 40%. In the latter case, the nudge is both successful and highly effective. Why do some people not \( \varphi \) under nudge conditions? I see three main possibilities. First, they might not have their shallow cognitive processes activated at all. Second, they successful inhibited those processes. Third, their

\(^{31}\)\(A\) and \(B\) stand for individuals, institutions, or populations. Though \(A\) and \(B\) are typically distinct entities, they can be the same entity at different times (self-nudge). **Deposit Contract** is an example of a self-nudge.
nudge-activated shallow cognitive processes got disrupted for unrelated reasons (e.g., the performance of competing tasks).

A nudge is also an intentional action (of the “choice architect,” to use Thaler and Sunstein's words) because a person’s freedom of choice can be infringed or preserved only by other agents. Although random features of the natural, social, or built environment, or unintentional behavior of agents, can also activate our shallow cognitive processes, they are morally and politically irrelevant, and therefore the individual has not been nudged in the sense I use this term.

My account of nudge amends Thaler and Sunstein’s definition on two points. First, it clarifies what is meant by the preservation of freedom of choice by elaborating on the importance of substantial noncontrol. Second, it introduces a more fine-grained understanding of non-deliberative and incompletely deliberative nudging mechanisms. Is this proposal revisionist? The response depends on whether these amendments bring our understanding of nudge closer to the function we want this concept to play in our public debates over health influences. While it certainly reduces the extension of the concept, it clearly argues for a more robust understanding of freedom of choice that has real moral traction.

Influences that activate shallow cognitive processes without satisfying the substantial noncontrol condition are not nudges because they fail to track influences that preserve freedom of choice in a robust enough sense. To facilitate the conversation, I suggest adding to the taxonomy of influences what I shall call “behavioral prods”:
**Behavioral Prod.** $A$ prods $B$ when $A$ makes it more likely that $B$ will \( \varphi \), primarily by activate $B$’s shallow cognitive processes, while $A$’s influence preserves $B$’s choice-set but is substantially controlling.

Behavioral prods are not easily resistible. **Asparagus-Lovers** is a paradigmatic example of a behavioral prod. Because behavioral prods use the same means (i.e., default-setting, framing, anchoring, etc.) as nudges and preserve the influencee’s choice-set, they are often confused with nudges when commentators narrowly focus on the influencing technique employed, neglecting the degree of control the influence exercises.\(^{32}\)

My distinction between nudges and behavioral prods explains why we are reluctant to call many business-oriented influences “nudges.” Nudging is often not the only way a person or an institution can influence behavior. More controlling influences can be used. In those situations, opting for nudges demonstrates self-restraint, an attitude in tension with the profit-maximizing goals of business-oriented practices. The title of a typical sales technique handbook speaks for itself: *The irresistible offer: How to sell your product or service in 3 seconds or less*. Another book targeting slightly more patient readers is titled, *The Science of Influence, How to get anyone to say ‘Yes’ in 8 minutes or less*. Salespeople are not always successful in crushing resistance, but it would be naive to ignore their goal. Marketers often prefer prodding to nudging.

---

\(^{32}\) Of course, the reader does not have to adopt the technical terminology I suggest. If some prefer to use the term “nudge” to refer very broadly to influences activating shallow cognitive processes, they could distinguish controlling from noncontrolling nudges.
Critics could object that health promoters use the very same marketing techniques. This is correct. But the focus of my concern is not on a set of techniques (Bonnell et al. 2011). Influences promoting unhealthy behavior are often more difficult to resist than those spurring consumers to healthy behavior. I assume for the sake of the argument that we agree on what counts as healthy behavior. For evolutionary reasons, we are simply more inclined to eat fatty and sugary food and avoid exercise than the opposite (Ubel 2009). When self-control is at stake, influences reinforcing our natural propensities are more likely to end up substantially controlling our behavior than those reorienting our impulses and habits or facilitating choices aligned with our second-order desires. Health promoters are more often nudgers than prodders. But when nudge policies target heterogeneous populations, a health-promoting policy designed to nudge a segment of the population may have an unintended, though predictable, stronger impact on another, psychologically more vulnerable, segment of the population. From a moral standpoint, health promoters are responsible for the overall reasonably predictable effects of their policies.

In sum, many influences are distinct from nudges when one varies three parameters:

1. The degree to which the influencers control influencees' behavior;
2. The range of options influencers make available to influencees;
3. The primary means the influencers use. Combining these parameters, Table 2.1 summarizes some of the most salient types of deliberate directive interventions that matter for the moral evaluation of actions and policies.  

33 There are undeniably connections between nudging and social marketing (French 2009).

34 I offer these definitions only as a starting point for discussion (cf. Faden and Beauchamp 1986; Wertheimer 1987; Nuffield Council on Bioethics 2007; Grant 2011; Blumenthal-Barby and Burroughs 2012).
<table>
<thead>
<tr>
<th>Type of influence</th>
<th>Degree of control</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>Choice elimination</td>
<td>Fully controlling</td>
<td>A preemptively removes the possibility to φ from B’s choice-set.</td>
</tr>
<tr>
<td>Compulsion</td>
<td>Fully controlling</td>
<td>A uses physical force to get B to φ.</td>
</tr>
<tr>
<td>Coercion</td>
<td>Fully controlling</td>
<td>A threatens to make B worse off if B refuses to φ.</td>
</tr>
<tr>
<td>Behavioral prod</td>
<td>Substantially controlling</td>
<td>A makes it more likely that B will φ, primarily by activating B’s shallow cognitive processes, while A’s influence preserves B’s choice-set but is substantially controlling.</td>
</tr>
<tr>
<td>Disincentive</td>
<td>Substantially controlling or Substantially noncontrolling</td>
<td>A increases the probability of getting B not to φ, primarily by raising the monetary or nonmonetary cost of φing.</td>
</tr>
<tr>
<td>Incentive</td>
<td>Substantially controlling or Substantially noncontrolling</td>
<td>A increases the probability of getting B to φ, primarily by providing B with some monetary or nonmonetary benefit.</td>
</tr>
<tr>
<td>Nudge</td>
<td>Substantially noncontrolling</td>
<td>A makes it more likely that B will φ, primarily by activating B’s shallow cognitive processes, while A’s influence preserves B’s choice-set and is substantially noncontrolling.</td>
</tr>
<tr>
<td>Rational persuasion</td>
<td>Fully noncontrolling</td>
<td>A induces B to willingly believe, or form the intention to, φ, primarily by presenting her reasons to φ.</td>
</tr>
</tbody>
</table>

Table 2.1: Taxonomy of deliberate directive interventions

Note that not all forms of speech-based influences are reducible to rational persuasion (e.g., deception, flattery, etc.) (Grant 2011, 70). Incentives may be distinguished from rewards and disincentives from punishments by introducing considerations of merit or desert (Grant 2011, 70). Mill introduced a version of this continuum, but insisted on the ends (compulsion/coercion versus rational persuasion). He also made some insightful comments on the fluidity of the choice elimination/disincentive distinction when he writes, “every increase of cost is a prohibition, to those whose means do not come up to the augmented price” (1991, chapter V, 111). Similarly, an influence can often be described either as an incentive or as a disincentive.
According to this taxonomy of deliberate directive interventions,\textsuperscript{35} the concept of nudge is sufficiently precise to be distinguishable from other types of influences.\textsuperscript{36} It is therefore not excessively vague (Bonnell et al. 2011; Marteau et al. 2011). With these distinctions in hand, we are in a better position to see what is wrong with the sample of influences that I mentioned in the introduction. \textbf{Asparagus-Lovers} is not a nudge because it is deceptive and not easily resistible. \textbf{Deposit Contract} is a self-imposed disincentive deliberately selected not to be easily resistible. \textbf{Quadruple Bypass Burgers}, and \textbf{Paternal Competition} are designed to arouse strong affective responses and the power of social norms (conformity or male competition) in order to substantially control the influencee. From Thaler and Sunstein's original list, only \textbf{Cafeteria} and \textbf{Generic Medication}\textsuperscript{37} qualify as nudges.

\textsuperscript{35}This taxonomy is not an “intervention ladder.” These ladders classify interventions according to their degree of restrictiveness, assuming that more restrictive interventions have a heavier burden of justification than less restrictive ones (Nuffield Council on Bioethics 2007, xix). The taxonomy of degrees of control neither affirms nor denies that the burden of justification increases with the degree of control.

\textsuperscript{36}Ruth Grant levels two objections against typologies that do not sufficiently distinguish forms of control (or power, to use her terms) from degrees of control (or power) (2011, 68). Her first objection is that persuasion should not be viewed as necessarily weaker than coercion, since deception might be more controlling than a threat to use force that lacks credibility. This objection does not undermine my argument, since my focus is on rational persuasion, which is certainly more controlling than coercion. In addition, I conceive of all categories in my typology as success terms, whereas Grant confuses successful coercion with attempted coercion. Her second objection is that incentives (or “bargaining,” in her terms) cannot be assigned a particular degree of control falling squarely between coercion and rational persuasion. I agree with Grant on this point, and I maintain that incentives and disincentives are sometimes substantially controlling, and sometimes substantially noncontrolling.

\textsuperscript{37}For the sake of argument, I suppose that there is no penalty attached to opting for a brand name.
Although we have gained in precision, there is room for ambiguous or borderline cases. First, we may wonder whether an influence plays a primary or secondary role in explaining its effect. I have noted that nudges “primarily” activate shallow cognitive processes: that is, these processes are always activated in the presence of a successful nudge (not merely a nudge attempt) and play a major explanatory role with respect to the effect produced by the influence. Secondary processes may also be activated and play a minor explanatory role. Occasionally, it is difficult to disentangle the effects of several processes, and therefore to distinguish a nudge from a persuasive influence or an incentive. Second, it is sometimes disputable whether an influence is an incentive, a nudge, or a prod. Consider a clear-cut incentive:

**Gym Bonus.** A company offers bonus salary ($5,000 a year) to employees if and only if they go to the gym on a regular basis.

Why is this different from **HIV-Test Cash Transfer**? Both cases have a common point: a recipient is offered some benefit for accomplishing a pre-determined behavior (behavioral conditionality). But in **Gym Bonus**, the benefit is consequential even for a wealthy professional. An incentive may introduce new reasons for action to motivate behavior change. **HIV-Test Cash Transfer** perhaps is not an incentive because its effect is not primarily explained by the magnitude of the benefit it provides. If it occurs just once, receiving 10% of one’s daily wage in cash may not be significant enough to fully explain its effect, even for disadvantaged populations. The likeliest explanation for the influence’s efficacy is that it triggers the tendency to overvalue proximal and immediate benefits. Is
**HIV-Test Cash Transfer** a nudge, an incentive, or a behavioral prod? The answer depends on how misleading the overvaluation of the benefit is.

**5. Conclusion: The Moral Import of the Concept of Nudge in Healthcare**

My account of nudge emphasizes two considerations that matter for the moral evaluation of actions and policies: the degree to which others control our choices and engage our deliberative capacities. I do not, however, beg the question of the moral permissibility of nudges by claiming that an influence has to be substantially noncontrolling to be morally permissible, or even that less controlling influences are systematically preferable to more controlling influences, by virtue of the principle of the least restrictive alternative (see chapter 4). Policy makers and proponents of rival moral views may use this amended concept of nudge without having to endorse questionable moral assumptions.

The ethics of nudging is particularly intricate when powerful public or private institutions alter minute aspects of the choice environment to generate population-wide health gains. This chapter focuses on conditions under which one has control over particular choices, no matter how consequential they are. This is why both those who endorse and those who reject the view that all liberties are on a moral par, and deserve the same presumption in their favor, can use my account of the concept of nudge in their debates over the ethics of nudging.

In chapter 3, however, I defend the view that all liberties are not on a moral par, a view I have developed in conjunction with Madison Powers and Ruth Faden (2012). According to this view, the contribution of particular liberties to leading a self-determined
life is the criterion we use to assign liberties different presumptive weight when a state policy interferes with them for the sake of public welfare protection. The overarching ethical goal is to secure a substantial degree of control over the broad shape of one’s life. Chapter 3 is a general presentation of this framework and does not discuss the health nudges. It puts in place many conceptual tools and normative arguments I use in subsequent chapters.

Chapter 4 applies the results of the three first chapters of my dissertation to the ethics of public health nudges. I argue there that the principle of the least restrictive alternative implicit in many defenses of nudges is invalid when public authorities, pursuing a legitimate public health goal, interfere with liberties that have no presumptive weight. Nudges are not systematically preferable to more controlling influences. I also explore the conditions under which nudges are preferable to more controlling influences.
CHAPTER III

LIBERTY, MILL, AND THE FRAMEWORK OF PUBLIC HEALTH ETHICS*

1. Introduction

Public health policies implemented by the state can affect a cluster of related moral interests that individuals have in personal autonomy, the exercise of choice, living under social and institutional arrangements that foster self-directing lives, and securing control over private matters. Popular as well as academic discussions of public health policies bearing on matters as diverse as reproduction, immunization, smoking, automobile safety, and dietary risks of disease often proceed with the assumption that there is a presumption in favor of individual liberty whenever it competes with public health goals.

Traditional medical ethics debates similarly focus on the place of autonomy among other competing principles and values within the context of patient-clinician relationships. However, the history of clinical medical ethics offers an imperfect model for thinking about public health ethics (although traditional lines between clinical and other aspects of health promotion are blurring at a rapid pace). Unlike clinical medical ethics, public health ethics directly engages the moral relationship between the state and

* This chapter is taken from Powers, Faden, and Saghai (2012). I made a few minor changes for the sake of clarity. I thank my co-authors and the editor-in-chief of Public Health Ethics, Angus Dawson, for giving me the permission to reuse this text (see Appendix B).
the individual and thus falls within the domain of political philosophy. Here, a common assumption is that any form of state coercion poses a uniquely troubling moral threat to personal freedom and autonomy. It is often concluded that there is a general presumption in favor of liberty such that a burden of proof lies with the proponent of any state intervention that interferes with its exercise.

We challenge this familiar way of framing public health issues by rejecting a general presumption in favor of liberty. We aim to modify how the balancing of some liberty and public health interests should proceed by taking the thumb off the liberty end of the scale, that is, by arguing against a burden of proof for some liberty interests.

In section 2, we examine briefly how the work of J.S. Mill has been employed to defend a general presumption of liberty in public health on the grounds that it provides the best account of the unique, core commitments of the liberal tradition. In section 3, we argue that there are good reasons to think that Mill rejected the kind of general presumption of liberty that this line of argument defends. Though we do not offer a definitive interpretation of Mill, we think that Mill makes an insightful distinction among liberties that should be treated as immune or quasi-immune from interference (in the sense of being strongly shielded from it) (first tier liberties); liberties that deserve a presumption in their favor (second tier liberties), and liberties that do not deserve such a presumption (third tier liberties). This tripartite division of liberties represents a fine-grained, nuanced framework for evaluating public health interventions rather than the familiar alternative that assumes the appropriateness of a general, undifferentiated presumption of liberty. In section 4, we argue that what is of focal importance for Mill in protecting liberty is captured by the essential role that the value of self-determination
plays in human well-being. Section 5 takes a step back and makes the case for a more complex and nuanced Millian framework for public health ethics, a view building upon arguments presented elsewhere as part of a comprehensive theory of the moral foundations of public health and health policy (Powers and Faden 2006).

The overall purpose of our reconstruction of Mill’s line of argument is to demonstrate the reasons that contemporary philosophers interested in the ethics of public health have for endorsing a political framework of this sort and to show how that framework links up with similar strands of argument found in the work of recent liberal political theorists. There are, of course, other important arguments made on behalf of a strong or even conclusive presumption against state interference with the kinds of choices that we deny are due any presumption in the balancing process. While such views need to be addressed as well, our argument for the differentiation among presumptions due to liberties of various sorts, if successful, removes one important line of defense for the view that even the most modest proposals for public health intervention bear a heavy burden of justification.

---

1 One prominent line of argument not discussed herein is the claim that because the state should be neutral among competing conceptions of the good, even what we take to be matters of considerably less consequence than major life plans or the overall contours of a person’s most basic life choices merit the same presumption against state interference. The literature on political neutrality is vast and the meaning and plausibility of a neutrality constraint is contested (Wall and Klosko 2003). Moreover, it is an open question whether such a constraint, if justified, would apply equally to “non-basic liberties” as a requirement of respect owed by the state to its citizens who hold differing conceptions of the good. These issues are discussed elsewhere (Powers and Faden 2006, 445).
2. Public Health Ethics, the “Harm Principle” and Core Liberal Commitments

The most frequently quoted excerpt from the famous passage in Mill’s *On Liberty* summarizing the Principle of Liberty asserts that “the only purpose for which power can be rightfully exercised over any member of a civilized community, against his will, is to prevent harm to others” (Mill, 1991a, 13-14). This formulation of Mill’s Liberty Principle is one that some (but not Mill) call the “Harm Principle.” Numerous contributions to the public health ethics literature put this aspect of the Millian framework front and center, and recent discussions demonstrate its continuing theoretical and practical importance.

Some argue for the need to displace the “Millian paradigm” (Jennings 2009), seeking to dislodge the protection of individual liberty interests and autonomy rights from their perceived privileged position. Others attempt what might be described as supplementation. The Nuffield Council Report, for example, proceeds by specifying additional conditions under which state promotion of public health objectives might be justified on grounds beyond the “harm principle.” Among these additional reasons are considerations of justice, such as the importance of reducing health inequalities (Nuffield Council on Bioethics 2007). The underlying assumption of many such arguments is that the centrally defining core commitments of the liberal tradition, informed by Millian arguments and their modern equivalents, require a highly restrictive attitude toward all public health interventions. The resulting interpretation of Mill’s Principle of Liberty, as applied to public health, is thus a general and stringent presumption in favor of liberty in relation to other interests, defeasible only for reasons pertaining to harm to others.
In the next two sections, we will show that this restrictive construal of the centrally defining, core commitments of liberalism, and of Mill’s own arguments as they pertain to the regulation of public health, are misguided. In particular, it is the defense of a blanket presumption in favor of liberty in all matters of state action for the sake of the public’s health that we challenge as a commitment required by liberalism in general or by a nuanced understanding of Mill’s own arguments.

3. Mill’s Tripartite Division of Liberties

Consider first the fact that early in the opening chapter of *On Liberty* Mill poses a question about the degree of protection against coercive interference that should be accorded to liberties of particular types. Initially, his claim is that at this stage in history it should be uncontroversial that liberties of conscience should be viewed as absolute in character (Mill 1991a, 16). No one else, including the state, has a morally relevant interest of the sort that earns it a place in any process of balancing against the interest individuals have in liberty of conscience. Mill announces that, although some will find it more problematic, he intends to argue in similar fashion for liberty rights of expression. He then considers whether the same absolute protection might attach to liberty of conduct as well, and at least for some forms of conduct, the ultimate conclusion he expects to defend also warrants absolute protection. Early on, then, we find clues to the basic idea that liberties of various sorts might warrant different degrees of threshold protection.
In his more specific elaboration late in *On Liberty*, Mill describes the Principle as composed of “two maxims” (Mill 1991a, 104), and he observes that the balance of weight of the various liberty interests covered by the Principle should proceed by way of “general presumptions” (Mill 1991a, 85). The upshot is that he ends up defending, first, a strong maxim that accords an absolute or conclusive presumption to some liberties, such that they are not subject to being overridden by any interests asserted by others. On such matters, he notes, the “individual himself is final judge” (Mill 1991a, 85). Mill notes, in broad fashion that “over himself, over his own body and mind, the individual is sovereign” (Mill 1991a, 13-14). Such a view allows for neither a calculus of comparative harm that might be caused to others by the exercise of choice within one’s sovereign domain of judgment nor the balancing of any other competing interest against liberty interests that are protected by what is, in effect, a conclusive moral presumption.

In contrast, the justificatory standard embodied in the second maxim merely establishes a presumption in favor of liberties of certain kinds, not immunity from interference. The reason is that the exercise of choice in some instances affects “certain interests” of others which ground competing rights claims, and therefore earn a place in some balancing process (Mill 1991a, 83). These lesser liberties are still weighty enough to warrant a moral presumption in their favor, but unlike cases involving matters over which the individual is sovereign, they properly may be subject to some balancing.

What has been often ignored is that a number of liberty interests, on Mill’s view, appear to fall outside the scope of either type of presumptive moral protection conferred by the Principle. Some choices lie within the jurisdiction of what Mill calls the doctrine of Free Trade, rather than the Principle of Liberty. The distinction and its importance are
often overlooked, but Mill is quite clear that the differences between the two are significant. He notes that “the principle of individual liberty is not involved in the doctrine of Free Trade” and that this doctrine rests on “grounds different from” those which underlie the principle of liberty (Mill 1991a, 105). The reason Mill claims that his Principle of Liberty is not involved in the moral evaluation of the types of state regulations that are governed by the doctrine of Free Trade is that, although the doctrine affects individual choices, it is focally concerned with balancing society’s interests in economic efficiency and society’s interests in public welfare protection. He offers some examples for which the doctrine of the Free Trade, rather than the Liberty Principle, is meant to apply. They include various forms of consumer product regulation and mandatory hazard labeling, state registries for the purchase of dangerous substances, laws regulating product adulteration, worker safety and wage laws, and a whole host of contract and other market place regulations (Mill 1991a, 105-108). Such regulatory targets are precisely the sorts of matters that many traditional public health interventions are designed to address.

The Free Trade doctrine assumes that government restriction on trade is “an evil” to be guarded against, but not because such restriction infringes upon a liberty interest in a manner that grounds either version of the presumption in favor of individual liberty. Mill is explicit in saying that “they are wrong solely because they do not produce the results which it is desired to produce by them” (Mill 1991a, 105). In these cases, therefore, where the only reason for state forbearance is economic efficiency or a failure to produce the desired improvement in public welfare, we believe that Mill does not endorse any fixed view about what, if any, moral presumption in favor of liberty is
required. Nor does he say anything to support the assumption that the moral analysis of public welfare regulation necessarily should proceed with a presumption in favor liberty. The reasons that ground the moral Principle of Liberty are not of the same type that ground the economic doctrine of Free Trade. In the former, the relevant reasons for any measure of state forbearance lie in the distinctly moral interests that individuals have in directing and shaping their own lives. In the latter case, state forbearance is assessed in relation to the balancing of societal interests in overall economic efficiency and societal interests in public welfare protection.

Embedded in Mill’s distinction among liberties that ought to be treated as if they were immune or quasi-immune from interference, those that warrant a presumption in favor of liberty, and those that fall under the doctrine of Free Trade is the premise that it is a mistake to suppose that liberties of all sorts are on a moral par. They are not appropriately viewed as being on a moral par in the sense that not all stand in an equally privileged position at the front end of a balancing process when liberties conflict with other competing interests.

Mill’s distinction among liberties, based on the differential threshold weight that each type of liberty should be accorded, does run counter to some other well-known views within the liberal tradition. Most obviously, within contemporary libertarian political theory there is an implicit moral parity claim regarding the singular standard of justification for all state interferences with liberty. Libertarians often speak as if liberties of all sorts enjoy the same high–often absolute–threshold weight in competition with all other goods or principles. Their singular political aim is to secure individual liberties, small and grand, whenever and however the state can intrude, against the uniformly evil
threat of any state interference with choice. However, it would be a mistake to equate the libertarian view with all of the liberal tradition. As we have seen, Mill, for one, does not endorse a fully general presumption of liberty. Nor does he endorse a single view of its stringency whenever such a presumption does apply.

4. Mill on Liberty and Self-Determining Lives

We now turn to our explication of the rationale behind Mill’s view that all liberties do not deserve either form of the presumption in their favor. If the Principle of Liberty establishes two types of moral presumption in favor of some individual liberties, and these distinctive moral reasons are different in kind from the public policy reasons in competition under the Free Trade doctrine, what exactly informs this tripartite division of liberties, and why does Mill take such great pains to distinguish them? The answer lies in scattered clues throughout the text of On Liberty and elsewhere, but the clues all point in the same direction. The key factor is that some individual choices merit some degree of strong state protection because of their connection to some crucially important aspect of individual self-determination as an ingredient of well-being. A variety of marketplace choices not implicating the value of leading a self-determining life in the same way do not fall squarely under either of the two moral presumptions afforded by his Liberty Principle.

Mill, of course, is notoriously elusive in his presentation of his case for the value of liberty, in part because his defense rests upon a view of how it figures within his
conception of human well-being, which he describes as “utility in the largest sense, grounded on the permanent interests of man as a progressive being” (Mill 1991a, 15). Even though an adequate reconstruction of Mill’s overall account of well-being (and whether Mill remains a utilitarian in good standing) is beyond the scope of our immediate task, there is enough evidence contained in the textual clues to venture a limited hypothesis about the heightened protection he wishes to claim for some liberties, but not for others.

The central lines of the argument in *On Liberty* focus on the interests that individuals have developing and executing a plan of life (Mill 1991a, 63, 65-66, 113-115) or participating in “experiments in living” by which individuals may test which modes of living contribute to their well-being (Mill 1991a, 63, 89-91, 117, 122-123). One strand of Mill’s argument emphasizes the claim that leading a self-directed life, one in which individuals exercise and further develop their capacities for determining their own destinies, is itself a constituent of well-being (Mill 1991a, 63). Mill also claims that when an individual is free to develop a plan of life, she develops her overall faculties to the fullest degree and thereby advances all other primary elements of her well-being (Mill 1991a, 63, 65-66, 70-71). Leading one’s own life, according to one’s own lights with regard to its overall shape, is in itself an important constituent of well-being, and it has added instrumental value in the ancillary development of other valuable human faculties that, when richly developed, contribute greatly to human well-being.

What Mill is focally concerned with in the protection of various liberties of thought, expression, and conduct is the vital interest each has in his own “mode of existence” (Mill 1991a, 75-76). Mill’s talk of a plan of life or experiments in living may
strike contemporary readers as overly rationalistic or as excessively deliberative, but the point of his odd locution does not lie in this direction. Mill’s focus is on the moral importance of a set of liberties that protect the kinds of choices that structure the course of one’s life in its most fundamentally defining ways. What the emphasis upon the formulation of an overall life plan – or way of life, or mode of existence - thus suggests is that the core concern of any scheme of protected liberties is not the perceived importance of providing all choices, no matter how quotidian, with either some heightened degree of protection or a blanket immunity from all state-related impediments, on the grounds, for example, that the risks of state oppression are ubiquitous and uniformly comparable. Instead, the argument turns upon the affirmative good Mill thinks is realized by social conditions facilitating the development and exercise of individual capacities, including capacities for self-direction, that are irreplaceable constituents of well-being (Mill 1991a, 63).

No general presumption in favor of liberty is necessary for the realization of the kind of uniquely important contribution to well-being of concern to Mill. To the extent that Mill’s general defense of liberty is predicated upon consequential life choices having a long arc of influence on overall life prospects, one need not have unobstructed choice in every marketplace decision or with regard to every possible course of action as a consumer. Subjecting every public welfare regulation to a presumption of liberty, as if every such regulation should be assumed to have the same sort of profound significance, is at odds with the very rationale upon which Mill attempts to categorize liberties according to the more nuanced degree of protection each deserves.
Critics might reply that any attempt to drive a wedge between the highly consequential, life-shaping choices and the quotidian choices of the sort typical of a host of marketplace decisions is contrary to the spirit of Mill’s overall theoretical ambitions to constrain the state. In defense of this view, critics might argue that definitive textual evidence to the contrary exists in *On Liberty*, where Mill says that the “most cogent reason for restricting government interference is the great evil of adding unnecessarily to its power” (Mill 1991a, 122). However, that conclusion would be wrong. The cited text occurs in the context in which Mill turns away from his consideration of coercive interferences with liberty to an evaluation of what might be objectionable about state action intended to help citizens by doing beneficial things that individuals would otherwise do for themselves. The evil of interference referenced in this passage is not the evil of coercive interference with individual choice, but rather the evil of well-meaning government action that has the unwanted effect of displacing individual initiative. Nothing in this often-cited passage supports the view that Mill’s political theory is overwhelmingly and single-mindedly focused on the evils of state oppression inherent in any form of liberty-limiting state action. In fact, where Mill is most explicit, he takes matters of public welfare regulation out of the domain of the Principle of Liberty and assumes that the centrally defining considerations to be balanced are economic efficiency and public welfare protection.

Our hypothesis, then remains intact. Mill is interested in defending a scheme of stringent liberty protections grounded principally in the special moral significance that attaches to choices that are consequential for an individual’s overall life prospects and focally related to control over the shape of a self-determining life as a whole. The proper
understanding of Mill’s comment on government evil, together with his clear exclusion of various marketplace choices from the scope of either of the two general presumptions arising under his Principle of Liberty, leads us to reaffirm the conclusion that the choices routinely implicated in public welfare regulation are without deep moral significance for the broad contours of a life that qualifies as substantially self-determining. In contrast, it is precisely the overall, major life-shaping choices that are of such deep moral significance that the Principle of Liberty is intended to protect.

Mill, of course, grants that state regulation of activities in the marketplace in the interests of public welfare may sometimes impinge on a buyer’s interest in forming a life plan, and in those instances, the level of appropriate front-end concern about interference with choices of a particular kind increases (Mill 1991a, 106). The contingent fact that regulation of market activities can take on the kind of added moral significance that may bring it under the scope of the Principle of Liberty does not alter the basic claim that, absent very special circumstances and the impact on an individual’s ability to determine for himself an overall mode of living, Mill does not view the ordinary marketplace choices of buyers and consumers as appropriately subject to the more demanding strictures of the Principle.

Critics might continue to press the worry that the textual clues in On Liberty are too slender to support the weight of the interpretative hypothesis we offer. However, additional support is articulated often more clearly and more forcefully in The Subjection of Women. In particular, it is in The Subjection of Women that we find a discussion of the “burden of proof” that rests with those who oppose liberty.
Those who oppose equal liberty for women are said to be under a “double burden of proof” (Mill 1991b, 472-473). It is important to see just what Mill has in mind when he makes this claim. He is not, for example, arguing that there is a burden of proof imposed upon anyone, who in any given instance, proposes to restrict the range of choices that might be available to an individual under some form of social arrangement. His target is something much bigger and quite different altogether than an argument for a presumption of deference to individual choice in any instance in which it might be restricted.

Mill’s point is about a way of life, and the necessary degree of immunity from comprehensive forms of external control that are of signal importance. The chief complaint Mill makes against the subjection of women, which he equates with the culturally pervasive, legally enforced regime of wholesale power of men over women, is an evil that he analogizes expressly to the kind of power exercised by absolute monarchs over their political subjects (Mill 1991b, 515-516). The absolute power of husbands over their wives within some legal systems is an equivalent form of despotism, now long rejected, and for the same reasons, the exercise of such power ought to be rejected in the domestic sphere as well (Mill 1991b, 509-510).

The thrust of Mill’s argument for equal liberties for women and men rests on the grounds that both have equal interests in making the kinds of life choices that matter centrally to a person who desires to be free from the control and domination of others. The same guarantee of independence that Mill takes as presumptively due to everyone, regardless of gender, is predicated upon the importance of the “same free development of originality in women which is possible to men” (Mill 1991b, 498). The conclusion is that
what is lost by living under either form of despotism is what he describes as “one of the principal sources of human happiness”, the personal satisfaction that comes from the free direction of one’s faculties (Mill 1991b, 578-582).

That Mill is concerned focally about the liberty to make consequential life choices is confirmed in his selection of examples of the kinds of morally objectionable impediments to equal liberty for women. Decisions about whether and whom to marry, admission to professions and careers without restrictions based on gender, control of property and personal assets, control over one’s own body and mind, the opportunity for education, for cultural advancement and contribution, and the opportunities for political expression and positions of authority are among the consequential sorts of life choices Mill enumerates. While some other routine choices in one’s daily life, when interfered with by the state, may be irksome, inefficient, ill-advised, and foolish, not all are of the sort that undermine one’s ability to lead what we call a self-determining life, or what Mill simply describes as a life in which no one is subjugated to the arbitrary and absolute will of another who more or less resembles a political despot in the kind and degree of authority exercised over them (Mill 1991b, 482-485). His brief for equal liberty is, as it is in *On Liberty*, an argument for a substantial measure of individual control over the shaping of one’s own life in the kinds of choices that structure the fundamental course of that life.

Moreover, the moral significance of protecting the ability of an individual to shape and direct the broad contours of her own life is claimed by Mill to be a good of incommensurable importance to well-being. The contribution to well-being that would be lost by the absence of a significant degree of self-determination exercised over the
primary determinants of one’s life prospects is so profound that no other increase in material well-being or other goods can compensate its loss (Mill 1991b, 512). In its essence, the heart of Mill’s argument for the burden of proof imposed on the opponents of equal liberty for men and women is his claim that the value of personal independence from forms of controlling influence intrinsic to both political and familial despotism is an indispensible ingredient of human happiness.

Mill’s defense of liberty as necessary for human well-being is thus very different from the one that supposes that every impediment to choice is so morally significant that it merits a case-by-case presumption against it. On the contrary, the central line of argument in The Subjection of Women, further illustrated by way of his primary examples, suggests that what is principally at stake within Mill’s political theory is not simply the value attached to making one’s own choices, whenever and however the opportunity might arise, say, for example, as consumers in the marketplace. Rather, what is of special value is a life of self-direction under conditions that permit individuals to determine the broad course of their own lives for themselves, free of the dominion of others who would, in the exercise of that dominion, rob them of an indispensible source of well-being.

5. A Nuanced Millian Framework for Public Health Ethics

We have argued that a suitably nuanced reading of Mill provides an important part of a framework for public health ethics, and for thinking about public welfare regulation more
generally. We begin with Mill not only because a particular interpretation of his arguments remains highly influential, but also because we think that the general thrust of his argument is largely right and that it is consonant with similarly persuasive claims made by recent theorists who are well within the mainstream of liberal political thought. However, we describe our own framework as Millian rather than claim it as Mill’s own view in every detail. Our argument is not based on the presumed authority of Mill, any definitive interpretation of Mill (assuming the many strands of his argument admit of such a thing), or the endorsement of all of his particular conclusions. For example, it is not our purpose here to defend or reject the view that there are any liberties that warrant absolute protection (as Mill believes), or that in all instances liberties of this sort can be readily distinguished from liberties that warrant a weaker presumption in their favor, or none at all. Rather, we maintain that Mill’s tripartite division of liberties and its rationale point in the right direction. In this section, we present two primary reasons why a more nuanced, and indeed, more complex Millian framework provides an attractive alternative to other public health ethics frameworks, and we conclude with a brief comment on how this line of argument connects with some other theories that challenge a different aspect of the kind of Millian framework that we reject.

Our first point relates to the legacy of liberal political theory. Although we do not work under the “default assumption” that liberalism should shape debates about public health practice and policy (Dawson 2009, 122), we do think it is useful to show that what we take to be a nuanced Millian view is not in conflict with this influential tradition. One reason to believe that this is not case is that the rejection of the general, undifferentiated
presumption of liberty is echoed in the works of other prominent liberal philosophers, including John Rawls.

One line of argument in Rawls might seem to reveal a major fault line between the Millian framework we endorse and a claim that is central to Rawls’s political theory. Lubomira Radoilska maintains, for example, that reliance upon considerations of social justice as grounds for public health restrictions on individual liberty “contradicts the priority of the basic liberties that defines political liberalism” (Radoilska 2009, 140). However, this claim misses the point of Rawls’s “priority of basic liberties” thesis, for he, like Mill, has a particular theoretical aim in mind in assigning lexical priority to some but not all liberties. The liberties identified by Rawls as deserving to be treated as immune from trade-off against any other welfare or opportunity-enhancing good constitute a special, defined class of liberties that may be sacrificed only for the sake of another basic liberty. This class of basic liberties is narrowly defined to include only some familiar civil and political rights, and it expressly excludes liberties that are denominated as non-basic. Non-basic liberties are ones pertaining to marketplace activities such as contract and certain forms of property acquisition and transfer (Rawls 1971, 273; Rawls 1993, 291-299). Rawls is clear that while basic liberties can be limited only for the sake of other basic liberties, non-basic liberties can be limited for a broader set of reasons, most importantly, considerations of distributive justice.

Rawls’s underlying assumption is thus that all liberties are not “on a par” (Rawls 1971, 273), and accordingly, he does not extend the categorical protection afforded to the basic liberties to every instance in which there is the prospect of state interference with choice (Rawls 1971, 273; Rawls 1993, 291-299). The explicit point that Rawls wishes to
make is that some liberties (basis and non-basic) have different starting points in the balancing process. The non-basic liberties Rawls identifies are the same sorts of marketplace liberties that Mill seemed to have in mind, and they include concerns about contracts and some types of property rights (Rawls 1971, 273).

Rawls might be read as endorsing some added threshold burden of proof for non-basic liberties (Freeman 2007, 78). For example, he notes that some moral considerations, in addition to those that are specific to the basic liberties, might provide the appropriate basis for a measure of added burden of proof on liberty infringement (Rawls 1993, 296). In some instances, some “general presumption” for non-basic liberties may be warranted, he says, but not for the same reasons that basic liberties are accorded their lexical priority (Rawls 1993, 291). However, Rawls’s explication of that remark reveals that he only intends to say that overriding non-basic liberties requires “good and sufficient reasons” (Rawls 1993, 291). This formulation does suggest that some justification is needed, but not that some sort of burden of proof or added threshold weight is necessary, or that the balancing begins with the scales tipped in any direction. A “good and sufficient reasons” approach signals nothing more than the analogous justificatory standard in legal theory in which a preponderance of the evidence is required in order to decide in favor of either side of a conflict. Unlike other legal standards that demand higher threshold weight, either by way of “clear and convincing evidence” or proof beyond a reasonable doubt,

---

2 For a discussion of similarities between Mill’s and Rawls’s views on private property, and the potential influence of the former on the latter, see Persky (2010). Neither of them endorses absolute libertarian property rights or liberty of contract, nor do they argue for in-principle liberty-based limits to interference with marketplace transactions.
there simply is no case made by Rawls for an assignment of a burden of proof in such matters.

Rawls’s argument, like Mill’s, makes clear that well within the liberal tradition are arguments in support of the claim that it is important to distinguish between those liberties that ought to be treated as if they were immune or quasi-immune from interference (Mill’s list partially overlaps with Rawls’s), liberties that benefit from a presumption in their favor, and those that lack such a presumption. In addition, Ronald Dworkin makes the similar point effectively in his famous remark that, “I have no political right to drive up Lexington Avenue… It will not do, in the one-way street case, to say that although I have a right to drive up Lexington Avenue, nevertheless, the government for special reasons is justified in overriding that right. That seems silly because the government needs no special justification – but only a justification – for this sort of legislation” (1977, 169). Here we join Rawls and Dworkin.3 Liberty interests matter in any case as concerns that trigger a need for justification, even if they do not loom so large as ones that figure centrally in the value of a self-determining life. The difference is that not all such concerns trigger a presumption in their favor.4

---

3 There are other liberal theorists who complain about the lack of nuance in some accounts of the weight of liberty interests. David Miller claims that a “line must be drawn between basic freedoms that people should have as a matter of right and what we might call bare freedoms that do not warrant that kind of protection,” that is, a presumption in favor of freedom unless there are “strong reasons” for restricting choice (Miller 2007, 204).

4 James Griffin argues similarly that the libertarian view “misunderstands the political value of liberty” inasmuch as it assumes implausibly that any and all liberties need comparable levels of protection in order
There are certainly borderline cases about which we may have a reasonable disagreement. This concession notwithstanding, the point remains that not all public health interventions – even when they take the form of state-imposed bans – implicate liberty interests of the sort that warrant presumptions in their favor.

The second reason we favor a nuanced Millian framework for public health ethics is that self-determination is not only an important element of well-being but an element of well-being that also matters centrally within a theory of social justice that we think of as providing the moral foundation for public health, more broadly (Powers and Faden 2006, 26-29). Our concern expressed under the rubric of “self-determination” is, as it is in Mill’s argument, the ability of a person to exert some substantial, though not perfect or complete, control over her or his path through life. Self-determination in this sense is not only a core element of well-being that is of value within a good life independently of all other constituents of well-being, regardless of what whatever else individuals might want. It matters crucially within a theory of justice because leading a self-determining life requires some measure of control over one’s fate such that an individual is not merely the instrument of the will of others or of social forces which she has had no role in shaping and which leave her without critical material resources or adequate political standing.

It is not necessary for leading a life that is more self-determining than another kind of life that there be some social structure in which all or as many choices as possible face fewer obstacles, impediments, or external influences. What matters is that we have some significant say over the general course of our lives. Put slightly differently, self-

---

for individuals to live as normative agents who place a high value on their ability to play a substantial role in shaping the course of their own lives (2008, 179).
determination is about living a life that is not in crucial respects under the domination and control of others or the tyranny of profound necessity. This is a further reason why the central political concerns of public health ethics should be focused on the value of self-determination, rather than based on the assumption that all choices are on a moral par. Only on such an assumption would a general presumption in favor of individual liberty is warranted.

The plausibility of our rejection of the blanket presumption in favor of liberty depends on how we might understand the value of self-determination that informs the distinction among the various classes of liberties. Our understanding of the value of self-determination can be brought into sharper focus by comparisons to similar, overlapping ideals expressed under different conceptual headings within other contemporary works of political philosophy.

Joseph Raz’s well-known ideal of the requirements of an autonomous life has obvious parallels to our account of self-determination. His ideal provides the rationale that he relies upon for the defense of various liberty protections, but in addition, that same rationale underpins the case for other forms of social organization said to be necessary for realizing the full value we should place on the exercise of the capacity for autonomy. Raz’s general aim is to identify the key constituents or conditions necessary for realizing the ideal of an autonomous life, and his central argument is that for those who value autonomy, they want not merely to have and exercise such a capacity, but to do so in ways that represent its role in contributing to a good life. The ideal is one in which persons are actually able to make something of their own lives according to their own lights, rather than merely being able to navigate an otherwise inhospitable world
with few impediments to choice, but also, with few valuable options from which to choose (Raz 1986, 369-374). Raz’s ideal thus identifies a rich set of concerns that are very similar to what we subsume under the value we assign to living a self-determining life.

James Griffin is similarly interested in identifying the necessary conditions for realizing the value of what he calls normative agency (Griffin, 2008). To be a normative agent in ways that matter morally (within his specific argument, to be the bearer of human rights) there must be some underlying capacity for self-direction, which he labels as autonomy. Like Raz, Griffin’s argument is that our moral interest lies with its successful exercise. Griffin also agrees with Raz that some underlying ideal of the successful exercise of human agency provides the ultimate rationale for the political protection of important liberties, but in addition, that very same rationale is used to underwrite some social guarantee of minimum provision of resources sufficient to make those choices “real” and not merely formally unconstrained. Our theory reaches the same conclusion, arguing that a self-determining life is not merely the ability to shape a life for oneself, but the ability to be self-directing under social conditions that make possible the realization of a sufficient level of well-being in all of its core dimensions (Powers and Faden 2006).

Both Griffin and Raz address issues of domination that our account of self-determination highlights. This feature of all three accounts is of crucial importance, and it is one that political theorists who work within the contemporary republican philosophical tradition (emphasizing political liberty conceived as non-domination or independence from arbitrary power) often criticize the liberal political tradition for neglecting
(Beauchamp 1988; Pettit 1997). The heart of the republican worry is that fewer external constraints on action may not account for all that an adequate political theory regarding human freedom should take into account (Pettit 1997). The republican claim is that any plausible ideal of a free person within organized society is incompatible with forms of domination that involve the control that some persons can exert over others, even if those in a position to exert control choose not to do so. Raz engages that autonomous agency requires independence from forms of external authority that have the power to exercise control over our lives, at will, and without limit (Raz 1986, 378). Griffin also notes the central importance of an individual’s independence from the domination of others, but unlike Raz, he subsumes that concern under his definition of autonomy, which is in turn a constituent of his more encompassing ideal of normative agency (Griffin 2008, 33, 151).

Raz and Griffin thus differ in terminology, but both accommodate worries about domination within their respective ideals, as do we. Our conception of the value self-determination incorporates overlapping concerns addressed by both Raz and Griffin, but given somewhat different emphasis in each account. Raz emphasizes the importance of forms of social organization that ensure an adequate range of options, focusing on the set of permissible restrictive social structures necessary to ensure immunity from the domination of others. Griffin emphasizes the importance of social guarantees of minimum levels of resource provision at a level sufficient for independence from profound necessity that undermines the successful exercise of normative agency (Griffin 2008, 33, 151). Our view is explicit in its incorporation of both forms of independence
that Raz and Griffin emphasize—from the domination of others and from the grip of profound necessity (Faden and Powers 2010).

For a variety of reasons, including the wide variation in philosophical usage surrounding autonomy and the risk of taking on board a variety of ancillary assumptions that we do not endorse, we prefer what we take to be the more straightforward language of self-determination. That said, the common threads across our account of self-determination, and similar ideals of an autonomous life in Raz and normative agency in Griffin, underscore our claim that what is at issue in public health ethics is not a general concern with all liberties but a focal concern with those liberties central to leading a self-determining life. Concerns of this sort are ones that other political theories also identify as centrally valuable.

---

5 There is much to be said about the moral significance of differences in how self-determination is affected adversely by domination and by social structural mechanisms that are not set in motions as part of a larger pattern of oppression or animus toward groups. We address these issues elsewhere (Powers and Faden 2006, 57-79).

6 Each theory brings its own set of assumptions that we do not necessarily endorse. For example, Raz at times suggests that because some choices are worthless, morally wrong, or repugnant they are not entitled to liberty protection (Raz 1986, 411). Griffin deploys his conception of the value of normative agency as the singular rationale for a theory of human rights. We make neither of these claims. Moreover, the bioethics literature, apart from political philosophy generally, has its own long-standing, internal controversies associated with autonomy and what appropriate respect for autonomy involves (Beauchamp and Childress 2009). We also refrain from using the language of negative versus positive liberty. These categories are far from enjoying universal agreement on what delineates them, and in our judgment, many familiar ways of marking the distinction tend to be more misleading than illuminating.
A final observation about the overall shape of our proposed Millian framework is that we concur also with those who argue that there are a plurality of justifications for liberty-limiting public health policies, other than prevention of harm to others (Dawson and Verweij 2008, 194). Although we do not argue the point here, we are sympathetic to Alan Wertheimer’s arguments that Mill’s own text contains a rather extensive catalogue of defeasibility conditions for presumptive liberty rights, apart from the prospect of harm to others (Wertheimer 2002, 38-58), including considerations of justice and the discharge of public burdens. The broader account of the moral foundations of public health and health policy, within which our conception of self-determination is embedded, concurs with those who seek to give a prominent place to multiple considerations of social justice (Powers and Faden 2006). Thus, we think that the best construal of a Millian framework for public health ethics is both more complex than one that supposes only one kind of reason for the defeasibility of any presumptive liberty right, and more nuanced in its recognition that all liberties are not on a moral par and thus, not entitled to either version of Mill’s presumption.

6. Conclusion

Mill deserves a prominent place in public health ethics, but for the right reasons. A broadly Millian framework for the ethics of public health is plausible only if it is as nuanced and complex as Mill’s own theory. It must begin with the recognition that all liberties are not on a moral par, such that liberties in need of the greatest protection are
those that are focally important to the value of self determination, and it must conclude with a recognition that even liberties that warrant a presumption in their favor can be defeated by a plurality of reasons. The focus of this chapter has been on the need for nuance that gives appropriate weight to liberty interests of all types, including those that warrant no presumption whatsoever, by arguing that they should not be restricted without adequate justification in the design of public health and other public welfare measures.
CHAPTER IV

NUDGES

AND

THE PRINCIPLE OF THE LEAST RESTRICTIVE ALTERNATIVE

1. Introduction

In this chapter I discuss under what conditions (if any) nudges should be systematically preferred to more restrictive influences. A widespread assumption among public health ethicists is that all morally justified public health measures (laws, policies, regulations) ought to respect the principle of the least restrictive alternative (thereafter, PLRA) (Childress and Gaare Bernheim 2008; Childress et al. 2002, Coggon 2012; Gostin 2008; Holland 2007; Kass 2001; Nuffield Council on Bioethics 2007; Upshur 2002). Though what this principle entails is far less clear than its name suggests, PLRA is usually understood as stating that public authorities must give very strong priority to the use of means which are the least restrictive of personal liberties to achieve their public health objective.

Proponents of PLRA disagree about its scope. As a principle of constitutional law, PLRA's function is to provide enhanced protection to fundamental liberties. In contrast, many public health ethicists believe that this principle should contain a wider
range of public health interventions.¹ Those who explicitly or implicitly endorse a wide-scoped PLRA conclude that when public health nudges are available, they always, all things considered, take priority over more restrictive alternatives (e.g., coercive measures).

In this chapter, I oppose wide-scoped versions of PLRA and defend the view that PLRA should instead be seen as a three-step decision procedure requiring public health authorities first to determine the legitimacy of the public health goal they consider pursuing; second, to guarantee that the relationship between the means of the intervention and its ends is strong enough; third, to be in a position to demonstrate that they have minimized the infringement of that intervention in proportion to the importance of the liberty interest at stake. As I explain, the application of the decision procedure varies depending on which liberties are deemed more important than others according to a theory. My understanding of PLRA can be used by many theorists (such as Rawlsians), but I make use of the tripartite division of liberties introduced in chapter 3. I argue, contra proponents of wide-scoped PLRA, that a defensible version of PLRA combined to the tripartite division of liberties would not demand that public authorities systematically prefer nudges to more controlling influences.

To argue this point, I first examine possible versions of PLRA (section 2), and argue that there is no compelling reason to adopt a wide-scoped version of PLRA (sections 3). Second, I present my model of PLRA and explore how we can plug in PLRA the self-determination-based tripartite division of liberties laid out in chapter 3

¹ PLRA might extend to rights and interests unrelated to freedoms, as Gostin mentions (2008). However, my focus in this chapter is limited to freedoms.
(section 4). Third, I defend the framework set up in this chapter and chapter 3 against possible objections (section 5).

2. Nudges and PLRA in Public Health Law and Ethics

Narrow PLRA in constitutional law

The primary source of PLRA is US constitutional law. The Constitution and the Bill of Rights prohibit the government from encroaching on fundamental rights and freedoms. Public health measures that count as state actions therefore are constrained by the Constitution, which protects individuals’ interests adversely affected by government’s acts (Gostin 2008, 114-116). Public health interventions led by private actors do not benefit from such protection unless there is a close nexus between private and public actors. In chapter 5, I return to the problem of assigning responsibilities and obligations to the private sector and civil society.

The function of PLRA is to protect fundamental liberties thanks to a system of retrospective judicial oversight of public health measures. Retrospective oversight is expected to modify policymakers' behavior, providing them with a reason to prospectively secure the protection of fundamental liberties when they select the goals and the means of public health policy.

The Constitutional protection is structured through a three-level approach to due process and equal protection. The basic idea is that different liberties and related rights trigger different levels of legal scrutiny of the constitutionality of public health measures.
These three levels of legal review are rational basis review, intermediate review, and strict scrutiny.

Rational basis review is the level of scrutiny that places a strong presumption in favor of government action. It requires that (1) the government’s goal be a legitimate public interest, and that (2) the means selected be reasonably related to the public health goal (Gostin 2008, 138). These requirements are weak in the sense that the government does not have to provide any compelling reason for its interest in performing certain actions, or evidence that the means it has chosen are strictly necessary for achieving its objectives. The Court defers to the expertise of the public health agency on these matters.

Rational basis review is important because “all public health regulation must at least comply with this minimum rationality standard” (Gostin 2008, 138). In other words, the presumption is strongly in favor of public health interventions as long as the state has an interest in protecting the public’s health (e.g. traffic safety, disease screening and prevention, mandatory treatment), and the means it chooses are not entirely disconnected from its goals.

The intermediate review differs from rational basis because (1) the government’s interest must be important, not simply legitimate, and (2) the relationship between means and ends must be substantial, not merely reasonable (Gostin 2008, 140). This heavier burden of justification lies on the state when its action involves certain worrisome interferences with individuals’ interests, such as discrimination based on gender (e.g., gender-based discrimination in marriage license regulations mandating syphilis testing for women only) (Gostin 2008, 140). Under intermediate review, the means selected to
implement a public health objective do not have to be necessary for achieving this objective.

The last level is strict scrutiny. Under this standard, the government has to respect PLRA. Under this understanding of PLRA, the government is required to demonstrate that (1) it has a “compelling interest” in achieving its goals, and that (2) its goals could not be achieved by less restrictive or discriminatory means (Gostin 2008, 141). PLRA is therefore a principle requiring a particularly high level of justification for the means through which public health authorities accomplish their role. Insofar as a compelling interest can be demonstrated and the measures are the least restrictive possible, the state does not have to renounce its public health goals in order to avoid interference with liberties.

I shall call the constitutional version of this principle, “narrow PLRA.” It is narrow because its scope of application is limited to infringements on fundamental liberties as defined by US constitutional law. Examples of such liberties include procreation, marriage, interstate travel, and child rearing (Gostin 2008, 141). Strict scrutiny is also triggered when the law discriminates against certain populations using “suspect classification” on the basis of race, ethnicity, or religion. In sum, strict scrutiny protects individuals from unfair discrimination or infringement on their fundamental liberties.

What makes a right or interest fundamental from a legal perspective is a complicated issue, given that the “Supreme Court, during its modern history, has used different forms of reasoning to determine if a right is ‘fundamental’” (Gostin 2008, 141).
Let me mention the three major forms of reasoning that the Supreme Court uses in that regard.

The first is appeal to the text of the Constitution, which mentions freedom of expression and religion, freedom from unreasonable search and seizure, equal protection from discriminatory or arbitrary policies and deprivation of liberty. To what extent some public health interventions interfere with those rights is an open question when it comes to compulsory vaccination (Gostin 2008), civil commitment of the mentally ill (Lin 2003), and quarantine (Parmet 2008).

The Supreme Court's second form of reasoning consists in the adoption of a dignity and autonomy-based criterion to demarcate fundamental liberties from nonfundamental ones. Certain liberties that involve “the most intimate and personal choices a person may make in a lifetime, choices central to personal dignity and autonomy,”² including the right to procreate, are thus considered fundamental.

Finally, the Supreme Court may appeal to rights “deeply rooted” in American history and tradition. These rights are also highly protected, even though they do not appear in the Constitution (Gostin 2008, 142).

It is important to note that privacy and bodily integrity are constitutionally protected interests, but they do not benefit from the highest presumptive protection because they have not been acknowledged as “fundamental” interests (Gostin 2008, 135-136, 142).

---

In sum, the concept of **PLRA** in constitutional law has a highly specific function. It requires the state to be in a position to demonstrate that it has selected the least restrictive means, only in case it deems necessary to interfere with fundamental liberties and rights to implement a legitimate public health goal. **Narrow PLRA** makes no such claim with respect to nonfundamental liberties and rights.

**Widening the scope of PLRA in public health law and ethics**

Although the **PLRA** is a US constitutional principle, a least (or less) restrictive alternative clause appears elsewhere in domestic and international law. Importantly, there is a trend toward the use of **PLRA** in statutory law and regulations, beyond its original constitutional context. For instance, in the US, the *Turning Point Model State Public Health Act* (2003) advises the adoption of wide-scoped PLRA, requiring state and local public health authorities and agencies to always give priority to the use of the least restrictive means.

In parallel to the legal widening of the scope of application of **PLRA**, many public health ethicists cite this principle in contexts that do not involve the protection of constitutional freedoms (Bayer and Fairchild 2004, 489). I propose to distinguish two models of wide-scoped **PLRA**.

---

3 In the European Union, **PLRA** is often called the “necessity test” (Sykes 2003; Harbo 2010). The two locutions are sometimes used with different connotations (see Childress et al. (2002, 172); Childress and Gaare Bernheim (2008, 160)).

4 This report is the result of the efforts of representatives from five states, nine national organizations and government agencies, and public health experts (including Lawrence Gostin).
The first model widens the scope of application of PLRA to all coercive public health measures, whether they interfere with fundamental rights and liberties or not. This is what I shall call “broad PLRA.” For instance, Ross Upshur writes that,

The full force of state authority and power should be reserved for exceptional circumstances and … more coercive methods should be employed only when less coercive methods have failed. Education, facilitation, and discussion should precede interdiction, regulation or incarceration. (2002, 102)

A second model broadens even more fully the scope of PLRA. It suggests that what I shall call “ubiquitous PLRA” should be applied to all public health measures. Here PLRA must be applied even with respect to policies affecting no fundamental liberty or right. Perhaps the best illustration of this position is the Nuffield Council on Bioethics' intervention ladder (2007). According to such a ladder, regulatory instruments are classified according to the degree of “intrusiveness” of the public health measure judged according to its degree of infringement on individual freedom (2007, 42). For policy purposes, the Council recommends a graduated approach in which the smallest degree of infringement on freedom is prioritized over more restrictive or intrusive interventions.5

The intervention ladder expands the requirement of strict scrutiny that the means chosen be necessary to achieve the legitimate goals of public health policy. But instead of indexing the degree of justification required to the importance of the liberty interfered with, or its coerciveness, the Nuffield Council suggests that the burden of justification

---

5 In the public health ethics literature, “more restrictive” and “more intrusive” are often used interchangeably. The point is to include both restrictions on liberty and intrusion on privacy.
and the degree of restrictiveness or intrusiveness co-vary. More restrictive or intrusive interventions require stronger justifications.

While **narrow PLRA** is a constitutional concept offering retrospective judicial oversight, the intervention ladder is a prospective ethical tool that policymakers ought to use in determining the moral justifiability of their decisions. The intervention ladder is an ordinal ranking of interventions based on their degree of restrictiveness or intrusiveness similar to what others have called the regulatory or enforcement pyramid (Ayres and Braithwaite 2002, chapter 2), that is, a form of regulatory step-by-step escalation.\(^6\) Thus, John Coggon has recently defended the intervention ladder against the charge that it involves an ordinal ranking of interventions. He argues that,

> Taking the intervention ladder as a conceptual tool need not commit us to any pre-analytic assumptions, such as that doing nothing requires no justification, eliminating choice will always be the most controversial intervention, or that there is any sort of ordinal ranking corresponding with the position in the ladder. Rather it presents, in a concise and useful manner, the range of approaches that may be taken in regard to public health policy. (2012, 80)

Coggon misses the real function of the intervention ladder, which is explicitly mentioned in the Nuffield Council Report: “The higher the rung on the ladder at which the policy maker intervenes, the stronger the justification has to be…. In considering which ‘rung’ is appropriate for a particular public health goal, the benefits to individuals and society should be weighed against the erosion of individual freedom” (2007, 42). **Contra** Coggon, I believe that the intervention ladder is not merely a useful presentation of the range of regulatory tools available to the policymaker. It is a spatial metaphor for a mono-dimensional PLRA exclusively concerned with minimizing the impact of policy on individual liberty. The Nuffield Council also attempts to defend its ladder against a formulaic interpretation:

> “It is not the case that the option of ‘doing nothing’ requires no justification, as deciding to ‘do nothing’ is itself a value judgment and may have adverse consequences for some. For example, not regulating vehicle speed limits or blood-alcohol limits might result in deaths or injuries …. Primarily, the function of the ladder is to compare alternative approaches in terms of their intrusiveness and likely acceptability, and not a means of allowing judgments in absolute terms. The intervention ladder is, therefore, not a formulaic device, but, further to our observations on proportionality, a tool for bringing into sharper focus the issues at stake.” (2007, 42)
ubiquitous PLRA is embodied in a detailed ranking of interventions even within the set of noncoercive interventions (e.g., if provision of information is feasible, it is preferable to use nudge-like defaults, which are preferable to the use of incentives or disincentives).

The Nuffield Council's intervention ladder guides policymakers toward a fine-grained comparison of alternatives. Unless policy makers have a compelling reason for opting for a more restrictive or intrusive intervention, they ought to give priority to the use of the least restrictive or intrusive alternative.\(^7\)

**Nudges and the scope of PLRA**

What are the implications of narrow, broad, and ubiquitous PLRA for the ethics of nudging? Narrow PLRA would systematically favor the use of nudges over more restrictive interventions only when a fundamental liberty is at stake and rational persuasion is unavailable. In contrast, under broad PLRA, nudges are always preferable to coercive interventions, but all non-coercive interventions are on an equal footing, that is, nudges may be preferred to persuasive interventions. Ubiquitous PLRA offers the strongest priority rule. Nudges are always preferable to more restrictive interventions, but

---

\(^7\) Other authors hold more ambiguous positions, mentioning in passing a PLRA with a more indefinite scope. For instance, Gerald Dworkin mentions that the state can have recourse to paternalistic policies under certain conditions, including PLRA (1972, 84).
if interventions less restrictive of liberty than nudges are available, they should be used instead of nudges. Hence persuasive influences are systematically preferable to nudges.

3. The Justification for Broad or Ubiquitous PLRA Examined

Oddly enough, there is scarcely any sustained and explicit attempt at defending either broad or ubiquitous PLRA in the public health ethics literature. John Coggon's remark that “the principle of least restriction is self-explanatory” (2012, 169) best reflects the lack of perceived need to explicate PLRA and to argue for it, since it is also assumed to be self-justificatory. I will however present and critically examine six possible justifications for broad or ubiquitous PLRA (see also Dawson (2011, 10-12) for another critique of many versions of PLRA).

First justification: Reasonable acceptability

A moral and political consideration in favor of widening the scope of PLRA can be abstracted from the framework of public health ethics that James Childress and colleagues have defended (2002). This line of defense could be used to back either broad or ubiquitous PLRA. (Childress et al. are proponents of ubiquitous PLRA, a point I discuss later).

Childress et al.'s claim that “[i]n a liberal democracy, the justification of coercive policies, as well as other policies, must rest on moral reasons that the public in whose name the policies are carried out could reasonably be expected to accept” (Childress et al.
Wide-scoped PLRA is justified on the grounds that it reflects the “public philosophy of liberal, pluralistic, democracies” (Childress and Gaare Bernheim 2002, 160). They conclude that reasonable acceptability leads to the adoption of a principle that they judge “quite similar” to the requirement of strict scrutiny in US constitutional law (2002, 172, footnote 10).

This line of reasoning is misguided. The requirement that policies of liberal democracies be reasonably acceptable does not entail widening PLRA. On the face of it, there is no reason to think that citizens would always require that the means-ends relationship be strictly necessary, and not for instance, substantial, or even reasonable.

In addition, “reasonable acceptability” could be understood as a requirement to use “public reasons” to justify laws and policies (Rawls 1993). In Rawlsian terms, public reasons are contrasted with sectarian reasons rooted in a comprehensive conception of the good life. Public reasons are responsive to the fundamental interests of citizens who endorse a variety of conceptions of the good life. But the mainstream Rawlsian interpretation of reasonable acceptability does not entail wide-scoped PLRA because no conception of a means-ends relationship is tied to a comprehensive conception of the good life.

**Second justification: PLRA as a balancing approach**

Childress and colleagues provide a second argument in favor of a wide-scoped PLRA. More specifically, they claim that ubiquitous PLRA, or in their terms, “least infringement,” is one of five justificatory conditions to promote public health goals (along with effectiveness, proportionality, necessity, and public justification). Their
understanding of least infringement is that public health practitioners ought to minimize infringement on each *prima facie* relevant moral consideration. In their view, least infringement is justified since it is merely another way of presenting the balancing method in bioethics, an approach they endorse on separate grounds. They explicitly reference Beauchamp and Childress's work on balancing as comparable to least infringement (Childress et al. 2002, 172, footnote 10; Beauchamp and Childress 2005, 19-21).

There are two major problems with the defense of the particular version of ubiquitous PLRA that the authors maintain. First, Childress and colleagues view PLRA as requiring minimizing infringement on all morally relevant considerations. In contrast, the traditional focus of narrow PLRA is on liberty-interests. Childress et al.'s view does therefore not fit well with their claim that least infringement is based on the model of constitutional narrow PLRA. This is because, under constitutional strict scrutiny, fundamental liberties always have more weight than other interests. They are not *prima facie* relevant moral norms of some sort (principles, considerations, or values). They have either absolute priority or weighted priority, such that they count more on the balance of reasons than other moral considerations. Absolute priority would entail that certain liberties can never be traded-off with other interests. In contrast, weighted priority is open to certain trade-offs, for instance when a marginal gain in liberty protection would significantly decrease cost-effectiveness, be drastically less expedient, or would result in a strikingly less efficacious policy.

---

8 For further discussion of this issue, see section 4.
Childress et al.’s view is prone to another problem, even if they could address the first problem by positing that the function of PLRA is to minimize infringement solely on liberty interests, rather than infringement on each prima facie moral consideration. James Wilson and Angus Dawson rightly object that if Childress et al. were also committed to the separate claim that all morally justified decisions are “all things considered” decisions that minimize infringement on each prima facie relevant moral consideration (and they are definitely committed to this claim), then their defense of a separate liberty-centered PLRA amounts to an illicit form of double counting. The double counting charge is that Childress and colleagues would be committed to (a) minimizing the impact of an intervention on all morally relevant interests (including liberty interests), and also to (b) further minimizing its impact more specifically on liberty-interests (2010). Liberty interests would then count twice.

**Third justification: Slippery slopes**

Slippery slope arguments could also be invoked in favor of widening PLRA. The worry here is that the state might slowly go from interference with trivial liberties to interference with less trivial ones. PLRA would prevent this unfortunate outcome. The slippery slope argument has been used, for instance, to criticize state regulation of food. In this vein, David Resnik suggests that bans on trans fats are problematic (2007) because they can lead to a form of what he labels “food fascism,” that is, I suppose, the exercise of an arbitrary power on the “freedom to choose the food we eat” (2007, 28).

Why should we expect public health authorities to go down the slippery slope? Some proponents of PLRA worry that an increasing number of public health
interventions are no longer aimed at consensual goals, such as avoiding or preventing wrongful and serious harm to others, or severe and involuntary harm to self (Fitzgerald 2001; Resnik 2010). These considerations are sufficient to back food safety and quality-based regulations but not health promotion. The latter urges that any increased risk of harm has to be prevented or eliminated. But this is precisely what leads to the slippery slope. As virtually any food choice may present some risk of increasing mortality or morbidity, “there is no difference, in principle, between banning artificial trans fats and banning other unhealthy foods, such as processed meats and sugared drinks. Today, trans fats; tomorrow, hot dogs” (2007, 29).

How are we to assess this slippery slope argument? The first question to ask is whether going down the slippery slope matters. What is wrong with restricting food choices, given that even Resnik concedes that the freedom to choose the food one eats is not as important as freedom of expression or of religion? Resnik’s response is that interference with food choices matters because it undermines ethnic, cultural, or religious practices associated with food choices. But it is entirely unclear in what sense restrictions on access to fatty and sugary food undermine the free exercise of the fundamental liberties that Resnik seems to care for. The kind of slippery slope that is genuinely worrisome could in principle be easily avoided by rejecting restrictions that undermine the exercise of a protected fundamental liberty.

Perhaps Resnik and other proponents of slippery slope arguments could argue that public health practitioners necessarily go down the slope because they believe any risk-reduction justifies state interference with the liberty to produce, distribute, and consume food. Partisans of PLRA could say that the goal of reducing risk is a moving target. It
could continue indefinitely, and therefore it does not offer any clear demarcation point between liberties that may be legitimately restricted and those that cannot.

But it is a fallacy to claim that because there is a continuum of risk, it is impossible to determine whether the magnitude or frequency of risk is a sufficient reason for interference with choice. For instance, banning fatty food from schools is justifiable as a measure to mitigate health risk of significant magnitude and frequency that might affect the lives of members of a particularly vulnerable group. Compare this case with the recent trend to eliminate class recess because of the increased risk of injury during playtime (Stout, January 5, 2011). The health gains do not seem to be of such magnitude or frequency to warrant strong restrictions on playtime because playing is an activity highly vital for normal physical, cognitive and affective development. Its importance dwarfs excessive concern for physical injury.

Fourth justification: Epistemic warrant

Some may argue that, under the constitutional version of narrow PLRA, most public health policies and laws fall under rational basis review, not strict scrutiny, and therefore the state has a very light burden to demonstrate that the means it has selected are necessary to achieve its goal (Gostin 2008, 143). This objection matters to the philosophical, not legal, discussion on PLRA because those who reject wide-scoped PLRA often use constitutional law as the starting point for their own theories. In addition, any version of PLRA must contain a clause on the nature of the required relationship between means and ends.
This objection to the constitutional version of narrow PLRA is correct and has ramifications for the debate over PLRA more generally. The rational basis review may set the burden of proof too low when it requires that the state be in a position to demonstrate that there is a reasonable relationship between means and ends. But it does not follow that the only alternative is to require the state to show that the means it contemplates to use are strictly necessary for the ends of its policy. One could also argue that the requirements comparable to those set under intermediate review should be adopted for all public health measures. Public health authorities ought to be always in a position to demonstrate a substantial relationship between means and ends. This demand would address the concern for an epistemic warrant without widening the scope of PLRA.

Fifth justification: Continuum of liberties

Critics might be dissatisfied with the constitutional criteria for demarcating fundamental liberties from other liberties. This problem also goes beyond the realm of legal studies since critics of wide-scoped PLRA need to identify the class of liberties to which their version of PLRA (if any) applies. Political philosophers like Rawls (1971; 1993), make use of distinctions between basic and nonbasic liberties that owes a lot to the constitutional distinctions between fundamental and nonfundamental liberties.

Lawrence Gostin argues that narrow PLRA does not offer adequate criteria for sharply distinguishing fundamental from nonfundamental liberties (2008, 143). As a result, certain liberties are misclassified. For instance, privacy rights (such as confidentiality) should be, according to Gostin, more strongly protected than they
currently are because of their “importance to human dignity and individual freedom” (2008, 143).

However, Gostin's legitimate and widely shared dissatisfaction with the sharpness of the constitutional criteria for fundamental liberties is not a sufficient reason to conclude that “as the intrusiveness and unfairness of the public policy grows, so would the level of scrutiny” (2008, 143). The existence of a continuum of liberties does not entail that the difference between liberties is only one of degree, not of kind (2008, 143). If all liberty interests are not equally important, then Gostin cannot exclude the possibility of meaningful thresholds along the continuum of liberty. In section 4, I explain why the tripartite division of liberties defended in chapter 3 provides a compelling alternative to constitutional and wide-scoped versions of PLRA.

Sixth justification: Preference satisfaction

A final reason that might be invoked for broadening PLRA is deference to the public's preference for liberty (Lin 2003). This is perhaps what the Nuffield Council Report stresses when it claims that the intervention ladder tracks the “likely acceptability” of an intervention.

On one possible view, the criterion for moral rightness is maximizing net preference satisfaction or desire fulfillment. As a consequence, if PLRA happens to be popular, then it ought to be adopted. But one may not need to be committed to utilitarianism to defend broad or ubiquitous PLRA on the basis of preference-satisfaction or desire fulfillment. Another route is a conception of liberal democracy as essentially requiring systematic responsiveness to popular wishes (Goodin 2003, chapter
3). Such “preference-respecting” democracy also might support the widening of PLRA's scope, assuming we have independent reasons to believe that the adoption of PLRA is the most adequate way of displaying respect for people's wishes.

The arguments from preference-satisfaction or desire fulfillment are however deeply problematic. Utilitarian preference-satisfaction or desire-fulfillment, and preference-respecting democratic theory, are open to an array of objections against the moral significance of preferences, even if we grant the unsupported empirical claim that the public may have a preference or desire for PLRA.

First, deference due to personal preferences in inter-subjective contexts does not directly transpose to a similar deference as a matter of political morality. In public health ethics, the problem is whether the state can interfere with individuals' choices for the promotion of public interests, not whether individuals are entitled to meddle with each other’s preferences or desires, even for public health reasons. There is a difference between a democratic state fluoridating the water supply and a public health Zorro performing similar actions on his own.

Second, when public interests are concerned, the mere strength of preferences has no intrinsic value. After all, the fact that some people may have a strong preference for driving the wrong way on Lexington Avenue does not plead in favor of changing the traffic rules to please them, to reuse Dworkin's example (1977, 169). Or consider the following scenario: some of us would strongly prefer not to endure the company of others in crowded public transportation during rush hour, while others do not enjoy it but do not mind it either. The strength of the first group's preference would not count as a reason for excluding those who have weaker preferences for uncrowded buses from access to public
transportation. In addition, preferences or desires may not be simply arbitrary or irrelevant; they can also be repugnant (e.g., immoral, discriminatory, irrational, or vile) (Powers 1994). Preference satisfaction is open to the objection that preference for the satisfaction of one's basic needs would count as much as the satisfaction of one's deliberately cultivated expensive tastes (Dworkin 2000). These are just a few reasons for not taking into account certain preferences. The point is broad or ubiquitous PLRA cannot be based on the view that preferences should always be satisfied.

Of course, a better defense of the preference satisfaction or the desire fulfillment objection is to claim that only a subclass of preferences or desires selected because of their source or content have moral significance. Those “filtered” preferences or desires are usually called informed or reflective (Goodin 2006). Assuming for the sake of the argument that such theories are correct and relevant to political morality, the question is whether a preference for broad or ubiquitous PLRA could be reflective. I simply note here that a preference for broad or ubiquitous PLRA that would assume a hierarchy of values systematically giving priority to liberty and autonomy interests over all other values and interests seems implausible. I doubt that such a value ranking is clearly endorsable on reflection. This leaves us with the possibility that liberty should benefit from weighted priority. Exploring this issue goes beyond the scope of this chapter and would require the careful examination of different theories of filtered preferences and informed desires.

I conclude that none of the six possible justifications I have outlined in this section warrants the move from the rejection of narrow PLRA to the endorsement of
broad or ubiquitous PLRA. In the next section, I present an alternative view meant to replace current conceptions of PLRA.

4. Nudges, PLRA, and the Tripartite Division of Liberties

An alternative conception of PLRA

Given my discussion of narrow, broad, and ubiquitous PLRA, I maintain that a plausible version of PLRA should satisfy the following desiderata:

1. Affirm that the state is entitled to infringe on important liberties only when it is pursuing a legitimate public health interest;
2. Require that the ends of an intervention and its means be substantially, rather than reasonably related;
3. Be sensitive to differences in kind and in degree among liberties.

I submit that the following conception of PLRA satisfies these desiderata. PLRA requires that when public health authorities are contemplating intervention \( X \), they need to be in a position to demonstrate that the response to the three following questions is positive:

(1) Is the state pursuing a legitimate public interest?
(2) Is the relationship between the means necessary for intervention \( X \) and its ends substantial?
(3) Has infringement on liberty interests been minimized, relative to the moral importance of that liberty?

In my view, the function of PLRA is that of a three-step decision procedure in the sense that, by itself, it does no substantive moral work. The criteria for determining the legitimate scope of state action and divisions in kind and degree among more or less important liberties are substantive moral issues that PLRA is not meant to address. It is not a moral principle.\(^9\)

I discuss the determination of the legitimate sphere of public health authorities' action in chapter 5. In this chapter, I focus on the second substantive moral issue, relying on the thesis presented in chapter 3. I maintain that the tripartite division of liberty interests based on their importance for leading a self-determining life is a compelling guide for responding to the third question. Of course, other accounts of differences in kinds and degrees among liberties may also provide guidance for the application of PLRA, for instance, Rawls's distinction between basic and nonbasic liberties (see chapter 3). In this sense, my criticism of PLRA and my reconstruction of the concept are independent of the thesis I defended in chapter 3.

In order to show how the tripartite division of liberties can guide PLRA, I start by revisiting the function of, and rationale for the tripartite division of liberties, with a focus on what it means to lead a self-determining life. I next argue that, according to this view,

---

\(^9\) My conception of PLRA has commonalities with Nancy Kass' approach. Kass proposes an ethical framework consisting of several reasoning steps. One of them is least infringement on liberties (2001). But Kass defends a version of broad PLRA, that is, her focus is on coercive interventions, regardless of the importance of the liberties they interfere with (2001, 1781). See also Dawson (2011, 11).
the application of PLRA ought be limited to first tier and second tier liberties, and be sensitive to differences in degrees among these kinds of liberties. Importantly, I maintain that PLRA does not constrain state action with respect to third tier liberties at all. I explicate the implications of this way of applying PLRA for the ethics of nudges. My claim is that this conceptual machinery guides us in determining when nudges are preferable to more controlling interventions.

The function of, and the rationale for the tripartite division of liberties, revisited

Chapter 3 outlines the rationale for a tripartite division of liberties according to the degree of presumptive protection they deserve from external controlling influences (particularly coercion). In this section, I revisit the function of and the rationale for the tripartite division of liberties.

The function of the argument for a tripartite division of liberties is to question the assumption that all liberties are on a moral par, that is, they all benefit from an equally strong presumption in their favor against third party interference, solely on the basis of the intrinsic value of liberty interests. The rejection of the moral parity assumption is, therefore, compatible with the view that non-liberty-related interests may also be sufficient grounds for an enhanced degree of protection of some choices from controlling influences. Those choices might be, for example, instrumentally necessary for the exercise of certain rights and the realization of other important dimensions of well-being. Another possibility is that all choices and preferences that do not infringe on important
rights of others ought to be respected if the liberal state is to be neutral toward competing conceptions of the good life.\textsuperscript{10}

For these reasons, the criterion for the tripartite division of liberties (the focal importance of self-determination) overlaps with, but is not identical to, the criteria for the identification of rights and the protection of all choices from certain forms of interference, especially coercion. However, the value of self-determination plays a major role in demarcating the class of the fundamental human rights that benefit from a stronger degree of protection than other rights (Griffin 2008). In other words, the tripartite division of liberties is not a theory of morality or a general account of rights. Its ambition is more circumscribed. It purports to show that the value of what we call “self-determination” bestows only certain liberties with an increased degree of presumptive protection from state inference for the sake of public welfare. As a consequence, any objection against our tripartite division of liberties based on non-liberty-related moral interests or outside the boundaries of political morality misses the point.

Let me now turn to the rationale for the tripartite division of liberties that we have borrowed from Mill. The degree of presumptive protection depends on the importance of certain individual choices for the exercise and development of capacities to lead a self-determining life. What is meant here is a life that is, in its main contours, free from the exercise of power by other individuals, and social and political institutions, so as to enable the individual to have a substantial degree of control over the broad shape of his or

\textsuperscript{10} Note, however, that respect for pluralism does not necessarily imply either state neutrality (Raz 1986), or noninterference with choices that are not consequential for the ability to form and implement a conception of a worthwhile life (Griffin 2008).
her life. Self-determination is an on-going process: lives are more often self-determining than self-determined.

A self-determining life is also a life endowed with enough material resources and opportunities to be in a position to actually elaborate and execute a “plan of life.” In other words, self-determination requires (1) non-domination, and (2) access to sufficient resources and opportunities. Having and developing certain decisional capacities is also a pre-condition for the elaboration and the execution of one’s plan of life. One needs to live in social conditions favorable to a sufficient level of development of these capacities. Indeed, the psychological capacities needed for self-determination are themselves in part dependent on social and political conditions that facilitate, encourage or frustrate the development of human decisional capacities.

In sum, the criterion for the tripartite division of liberties is meant to isolate state interference with choices that have a significant one-time or cumulative impact on self-determination’s constituents and pre-conditions from interferences that have no commensurable impact.

The Millian locution “plan of life” is a bit misleading. It might suggest, despite Mill’s quite explicit claims to the contrary in chapter 3 of *On Liberty*, a high degree of reflection over what one wants most from life, an atomistic conception of the person who conceives of his or her life goals in a social vacuum, and a quasi-managerial way to relate to oneself through a planning-oriented attitude toward life. This is not what I have in mind. A plan of life, as I conceive it, is closer to what James Griffin calls a “conception of a worthwhile life.” He means by this having some ideas about what makes “a life better or worse” (2008, 46). But Griffin’s formulation is too impersonal. What is at stake
is what makes one’s own life better or worse both reflectively and unreflectively. One’s plan of life is embodied in one’s identity, lifestyle, tastes, desires, values, and commitments, which are in turn embedded in “shared social forms for life” (Raz 1986, 312). Life-shaping choices designate both one-off choices and habitual actions that have the greatest impact on one’s ability to live one’s life.

Implications of the tripartite division of liberties for PLRA and nudges

Let me turn to the implications of the tripartite division of liberties for the scope of application of PLRA by considering interferences with first tier, second tier, and then third tier liberties. I stress the ramifications of this approach for the ethics of public health nudges.

PLRA and first tier liberties. When the state is contemplating interference with first tier liberties (i.e., vital for leading a self-determining life), it should give priority to the use of the least controlling influences. In that situation, when nudges are available, they are clearly preferable to coercive interferences. Naturally, if persuasion is feasible, then it is preferable to a nudge.

But even when vital interests might be interfered with, there are complications that may relax the requirements that a strict priority rule would impose. As stated in chapter 3, I neither deny nor endorse the Millian suggestion that some liberties ought to be so strongly shielded from external controlling interferences that they should never be interfered with for any reason. Suffice to say that if there were some liberties of this sort, they would be very limited in number, since even the liberty of expression needs to be regulated, and thereby restricted. Rawls makes a similar point when he notes that, “not
everyone can speak at once, or use the same public facility at the same time for different ends” (1993, 296). If this is correct, then the status of vital or basic liberty interest only provides additional protection by way of increased burden of justification on those who want the state to restrict it for any reason.

The point is important because trade-offs between noninterference with certain liberties and the just allocation of resources are unavoidable. This is particularly true when a liberty cannot be adequately exercised without the provision of costly social conditions. For instance, Colin Farrelly remarks, in an insightful discussion of this issue, that even the right to vote is not cost free:

The distribution of polling stations within a geographical territory and the hours of operation of a polling station, etc. will also have an impact on the opportunity citizens have to exercise the right to vote. These provisions have budgetary implications, which can run into millions of dollars. (March 12, 2010)

It is therefore safe to say that often times vital liberties may be interfered with, and restricted for the sake of other values, such as public interests. Concretely, this means that nudges are not systematically preferable to more controlling influences, but that they benefit from a strong presumption in their favor when the state considers influencing choices that involve vital interests, such as the political liberties, certain reproductive choice or end-of-life decisions. There are differences in degrees of presumptive weight even within the class of first tier liberties.

**PLRA and second tier liberties.** When the state contemplates interfering with second tier liberties (i.e., consequential for leading a self-determining life), PLRA's requirement to minimize infringement on liberty interests still holds, but its defeasibility conditions are weakened. Nudges should still be, as a rule of thumb, preferred to coercive interferences,
although sometimes fully controlling influences are also legitimate. For second tier liberties, considerations of cost, for instance, gain more weight, and indeed cannot be bypassed simply based on the application of a priority rule.

*PLRA and third tier liberties.* When the state contemplates interferences with third tier liberties (i.e., non-consequential for leading a self-determining life), *PLRA* does not state that infringement on those liberties needs to be minimized. This is a radical conclusion that other defenders of nudges and many public health ethicists would not endorse. The use of nudges is not preferable to recourse to more controlling means, even as a rule of thumb. Liberty considerations are not tiebreakers between, say, coercion, the elimination of choice, and a nudge. Of course there may be other moral, pragmatic, or prudential reasons for preferring nudges or persuasion to coercion. But public health authorities do not have to demonstrate that they have minimized infringement of third tier liberty interests.

5. Objections and Responses

Critics might reply that this theory is too permissive in the sense that it provides the state with a blank check for unrestricted interference with third tier liberties. In this section, I consider four objections to my view.

First, critics may reply that my view is highly counterintuitive. Suppose for the sake of argument that a health-affecting behavior contributes little to leading a self-
determining life, so that this liberty does not benefit from even a presumption in its favor. Intuitively, there are morally significant difference between nudging and criminalizing that my characterization of third tier liberties seems to overlook.

Second, some may argue that I do not acknowledge the value of bodily integrity. Imagine the case of an individual who does not wish to get inoculated with a vaccine that physicians judge mildly useful to him. Suppose that unbeknownst to him, he is drugged and inoculated with the vaccine. By hypothesis, the vaccine is harmless to him and is even likely to bring him some minor benefits. Stephen Holland (2009) argues that this case raises a major problem for supporters of various conceptions of autonomy or related interests that focus on major life-affecting choices (e.g., Brännmark 2006; Archard 2007; Nys 2008), since this event does not affect the overall shape of one’s life. Yet it is intuitively objectionable.

Third, critics may point out privacy issues. Suppose that research conclusively correlates certain patterns of Internet usage to depression, so that the presence of particular styles of Internet behavior is a reliable indicator of the probability for depression (Chellappan and Kotikalapudi June 15, 2012). This information could be used by a health monitoring authority to detect Internet users who might be depressed, via software installed on the user’s electronic device, and alert the user when “patterns might signal symptoms of depression.” If the users are university students, one proposal is to install such software “on campus networks to notify counselors of students whose Internet usage patterns are indicative of depressive behavior.” To make the case entirely relevant to our topic, imagine the university is funded by the state through taxes. The state and citizens may have public health and financial interests in reducing the rate of
mental health issues by early detection and treatment among this segment of the population. This counterexample highlights issues of privacy that need to be adequately addressed.

Fourth, one might challenge the view that only life-shaping decisions should be protected from external interferences, leaving choices over more trivial matters at the mercy of others’ whims. Raz has this problem in mind when he writes that,

> It is [...] unacceptable that we should not be able to decide on trivia such as when to wash and when to comb our hair. This aspect of the requirement of adequate choice is necessary to make sure that our control extends to all aspects of our lives. This is clearly required by the basic idea of being the author of one's life. (Raz 1986, 374)

Raz concludes that the liberal state should offer an adequate range of options to people to choose from, including choices that are not life shaping.

In the next paragraphs, I respond to each objection in turn.

**Criminalization**

My view is compatible with the claim that there are significant moral differences between nudging and criminalizing.

First, criminalization of an activity changes its normative status. This in turn is crucially important for leading a self-determining life in the sense I have outlined. It is in effect impossible to lead a life, which is in its broad contours one's own life, if one is at the mercy of arbitrary criminalization of quotidian activities. Risk reduction cannot justify turning an activity into a crime. We have reasons to suspect that concern for health is not the implicit motivation behind criminalization. Rather criminalization expresses a negative moral judgment on a way of life.
Second, if the criminalization of an activity is coupled with jail sentences for offenders, a vital liberty interest, freedom of movement, is undermined without a sufficient justification. There is coercion and coercion. Raz puts the problem in the following terms:

First [coercion] violates the condition of independence and expresses a relation of domination and an attitude of disrespect for the coerced individual. Second, coercion by criminal penalties is a global and indiscriminate invasion of autonomy. (Raz 1986, p. 418)

Criminalizing an activity amounts to an attempt to fully control someone's behavior. It has either insulting expressive effects, or is prone to the “mixture problem.” The latter arises because criminal penalties often overshoot: imprisonment not only deprives one of the possibility of performing the activity it is supposed to prohibit, but it also forecloses access to a whole series of other liberties, some of which are necessary for leading a self-determining life, for instance freedom of movement.

Critics could reply that many public health interventions my view would favor do not criminalize consumers' behavior but they criminalize producers' behavior (Grill 2009). I doubt this is the case for the ban on XL soda drinks in New York, where one could be fined for selling XL sugary drinks (over 16 fluid ounces) in restaurants, fast-food franchises, sports arenas, movie theaters, food trucks, street carts, and delis (Grynbaum, May 30, 2012). But supposing that for product Y, the producer or the distributor would be sued on criminal charges, then even if access or lack of access to product Y is not consequential for liberty, banning it would lead to the criminalization of

---

11 For an illuminating discussion of this issue, see Stanton-Ife 2006.
the actions of third parties (producer and distributors). This issue introduces interesting complications with respect to product safety regulations that I cannot pursue here.

**Bodily integrity**

Bodily integrity encompasses at least four types of interests: personal security, respect, trust, and liberty. Threats to bodily integrity put one “in constant fear of physical and psychological abuse” (Powers and Faden 2006, 19). They also violate “any minimal notion of respect for persons as moral equals (Powers and Faden 2006, 19). The prospect of being drugged and then vaccinated undermines the value of trust in healthcare if physicians use deception to control the patient’s body. In addition, while health has a value of its own, in Holland’s example the condition for which the individual is inoculated is minor, hence it does not *prima facie* offer a good and sufficient reason for the intervention, even if the liberty interfered with did not require heightened protection. Moreover, the context in which this case is presented is remote from public health since the motive for the action is individual, not collective, benefits. But in this case the motive is indefensible. Individuals’ choices are due presumptive protection given that, for creatures like us, the body is an integral part of what makes us vulnerable persons whose path through life is conditioned by their interactions with others. Therefore significant invasions of bodily integrity for no good reason are rightly perceived as putting one at the mercy and control of others in nontrivial ways.

**Privacy**
The third objection concerns matters of privacy. Privacy raises particularly difficult
problems in public health and healthcare because a certain level of monitoring of
individuals is necessary to protect and promote health. Many interventions, whether
persuasive or coercive, depend on prior access to information about individuals’
behavior. Because some shield from unwelcome third parties scrutinizing one’s verbal or
nonverbal behavior is necessary for “experiments in living” to figure out how one wants
to live one’s life, privacy falls in part under self-determination. But matters are
complicated here because the scope of privacy interests strongly tied to self-
determination is uncertain (Griffin 2008, chapter 13). In addition, data collection plays an
important role in epidemiology.

Presenting a detailed theory of privacy rights is not a project I will pursue here.
However, I note that one important indicator that a privacy issue deserves heightened
self-determination-based protection is whether the information allows individuals to be
easily identified, or whether data collected is de-identified for statistical analysis of trends
across populations. In the latter case, monitoring does not expose the way of life of any
particular person to public scrutiny or censorship. Arguably, the mental health
counterexample is worrisome because it identifies individuals at risk and threatens their
unsupervised freedom of thought, expression, and access to information. It puts them in a
position to be de facto accountable for their personal behavior to mental health
professionals, and therefore may generate forward-looking self-censorship.
Lifestyle

The fourth objection focuses on the issue of interference with apparently trivial choices that are part of individuals' lifestyle. If interference were successful, it would debase the claim that many health-related liberties should not benefit from a presumption in their favor. The difficulty is to unpack what interests of moral significance are at stake in routine practices (rather than crucial one-off decisions), such as those involved in eating habits, smoking, and alcohol consumption (Wikler 1978; Wikler and Beauchamp 1995). Do these choices matter for being the author of one's life?

There are problems with the way Raz formulates this issue. State interference even with trivial liberties is illegitimate if democratic state action is not justified by public interest. State actions should not be arbitrary. Hence, part of the issue lies in whether the government is in a position to make expert decisions with respect to health-affecting behavior (Veatch 2009), and whether individuals have reasons to trust public health experts.

Authors who defend the moral significance of lifestyle choices often fail to precisely delineate what the moral concern with lifestyle is. Thus, they use interchangeably the terms “personal lifestyle,” “way of life,” “personal behavior,” “the good life,” and “comprehensive conceptions of the good life” (Korthals 2008, 446-447). But separate moral considerations already provide liberties that are expressive of individuals' moral values and intimate choices with a presumptive weight, such that the state cannot enforce a sectarian conception of the good life on individuals.

For the discussion of healthy or unhealthy lifestyles to be more precise, we need to narrow down what aspects of habits and choices lifestyle choices designate. I suggest
that healthy or unhealthy lifestyles refer to individual and collective patterns of health-affecting production and consumption routine habits and choices made among socially available valued options and normative practices (Blyton et al. 2010, Cockerham 2010, 159). Lifestyle choices are deeply constrained by social norms and socio-economic structures that influence or shape the evaluation of one's options. When is public health authorities’ intervention to influence those social factors by inculcating certain norms morally problematic?

I submit that the right way to approach this problem is holistic. Most attempts to change lifestyle, considered in isolation, seem benign. Problems emerge when their cumulative weight affects their place in a culture by contributing to the stigmatization of certain ways of living or validating a disrespectful attitude towards a specific group. Of course, the history of the population and other possible sources of cumulative disadvantage need to be taken into account.

In a nutshell, a health influence attempting to change a lifestyle needs to take into account its contribution to the creation of a culture of practices whose shape and meaning might express disrespect for people's dignity, integrity, and identity (Kukla 2005).

6. Conclusion

In sum, I have examined in this chapter whether nudges should be systematically preferred to more controlling influences in order to pay due respect to liberty interests. I have argued that the deceivingly intuitive principle of the least restrictive alternative
cannot be used to provide recourse to nudges with an absolute priority over the use of other means. Unlike other authors, I defend the view that PLRA is irrelevant when liberties interfered with deserve no presumption in their favor, and defend a distinctive model of PLRA considered as a decision procedure that is triggered just in case important liberties are at stake.
CHAPTER V

JUSTIFYING PUBLIC HEALTH NUDGES

1. Introduction

Despite the immense amount of recent literature both for and against nudges, the ethics of nudging has had trouble finding its center of gravity. In particular, not enough has been written about the delineation of the particular moral concerns that nudges give rise to, when contrasted with other forms of influences. Surely, nudges are impermissible if the aim of the nudger is morally wrong in itself or if the nudge will predictably cause serious harm to the nudges (Blumenthal-Barby and Burroughs 2012). But in this respect, the moral assessment of nudges is similar to the moral assessment of any other type of influence.¹

¹ Like most well designed policies, public health nudges might be expected to make the majority better off and a minority worse off. Even the libertarian paternalistic requirement that the nudges ought to make the recipients better off as judged by themselves does not mean that all those whose decisions might be ex ante affected by a nudge should be expected to benefit from that nudge. This is because few policies have uniform effects on the population. Those who believe that a welfare-promoting policy is morally justified only if it makes no one worse off bear the burden of demonstrating that their view would not rule out most public policies of the modern welfare state, assuming we agree that these policies are overall valuable. Of course, if the harm that affects a minority of people is either severe or a violation of their rights, then that
In order to make progress toward the clarification of what is morally at stake with the use of public health nudges, I have provided, in previous chapters, an account of the nature of nudges, in general, and of their claim to preserve freedom of choice (chapters 1 and 2). I have also argued that public health nudges are not systematically preferable to more controlling influences merely because they are substantially noncontrolling (chapters 3 and 4).

In this chapter, I examine the justificatory burdens that are unique to the use of public health nudges. I begin by distinguishing public health nudges from other population-wide health nudges (section 2). I then distinguish the minimalist and the maximalist conceptions of what it takes to justify public health interventions (section 3). In the two final sections, I analyze each conception and explain why they lend support to different sets of nudges (sections 4 and 5).

2. Public Health Nudges and Population-Wide Health Nudges:

Mapping the Terrain

Population-wide health interventions (including some nudges) affect the choices and behavior of groups of individuals, for example through the modification of the choice environment. They can be contrasted with more precisely targeted interventions, such as

policy is morally unjustified. I have benefited from a discussion with Paul Dolan and Bob Veatch on this point.
individually tailored interventions designed to improve the health of particular persons that have no direct and significant effect on the choices and behavior of third parties.

In this section, I argue that population-wide health nudges and public health nudges are not coextensive. As a consequence, some population-wide health nudges should be assessed according to the salience of certain moral concerns that are not central to public health ethics. Consider the following examples of population-wide health nudges (or possibly prods), in addition to Clinical Trial Enrollment, Deposit Contract, and Hand Washing introduced in chapter 1:

**Anesthesia Equipment.** To reduce risk of medical errors in anesthesiology, each gas nozzle is designed to be compatible with just one connector so as to minimize the need for deliberation when giving anesthetic drugs to patients (Thaler and Sunstein 2008, 89).

**Intelligent Assignment.** Medicare beneficiaries are assigned by default to a healthcare plan based on the historical data about their drug use during three months in the previous year (Thaler and Sunstein 2008, 161-176).

All of these examples may lead to public health improvements, but only the goal of Hand Washing is primarily to promote public health. Hand Washing is thus primarily justified on the basis of considerations typical of public health (see sections 3 and 4).
**Intelligent Assignment** is mainly part of health policy and should be assessed with regard to its ability to offer the public the most cost-effective access to healthcare, and whether it is sufficiently respectful of the moral significance of choice. ²

**Anesthesia Equipment** is at the border of clinical medicine and quality improvement. It can be required on the basis of the physicians' obligations to act in the best interest of each of their patients, and of hospitals' obligation to provide physicians with facilities that are the most conducive to the fulfillment of their obligations compatible with the hospitals' budgetary constraints. It is, however, a population-wide health nudge since it affects the health of all users of surgical services.

**Clinical Trial Enrollment** raises moral concerns more salient in clinical research. When the trial involves more than minimal risk, any influence that might alter the full voluntariness of the consent is prohibited. Relying on foot-in-the-door marketing techniques through the use of defaults, and on the patient’s perceived obligation to reciprocate a favor (even a phone call) may compromise the voluntariness of the individual’s choice to participate in a clinical trial. The concerns surrounding individual consent are more salient in risky clinical research than in, for example, low-risk public health inquiries where de-identified data is used to generate epidemiological data.

Contrast these cases with **Deposit Contract**. This example could be viewed as part of a health promotion program designed to promote the health of identifiable individuals, and it is tailored to their personal needs (Buchanan 2000). Here, unlike

---

² “Health policy refers to decisions, plans, and actions that are undertaken to achieve specific health care goals within a society” (World Health Organization, URL= http://www.who.int/topics/health_policy/en/, accessed October 7, 2012).
typical public health interventions, consent comes at the end of a shared decision-making process in which patient and physician set a health improvement goal. In addition, the patient needs to decide on the amount of money he or she wants to use to motivate his or her actions, relying on our predictable tendency to overreact to loss aversion. This is neither an example of population-wide health intervention, nor an example of public health policy. However, other health promotion programs might qualify as population-wide health interventions. For instance, bike-sharing systems like Capital Bikeshare in Washington, DC and Vélib' in Paris are part of the built environment. They have health promoting effects (for those who choose to use them to improve their health), as well as partly overlapping public health and environmental-friendly effects.

In sum, population-wide health nudges cover at least five partially overlapping territories: public health, health promotion, clinical research, clinical medicine, and health policy (see Chart 5.1). Of course, each of these domains may also include interventions that target specific individuals (e.g., most interventions in clinical medicine, or single-subject clinical trials). Although the boundaries between these domains are blurry, each of them has its own set of salient moral issues.
3. Two Conceptions of Justified Public Health Interventions

There are two main rival schools of thought of public health ethics around which the viewpoints of most theories gravitate. These are best exemplified in the works of Mark Rothstein (2002) and Marcel Verweij and Angus Dawson (2007), respectively, and are therefore not straw men or even mere ideal types. I'll call them respectively the
“minimalist” and the “maximalist” conceptions of justified public health interventions.\textsuperscript{3}

Table 5.1 summarizes their distinctive characteristics:

<table>
<thead>
<tr>
<th>Public health minimalism</th>
<th>Public health maximalism</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Justificatory considerations</strong></td>
<td>• Harm to others</td>
</tr>
<tr>
<td></td>
<td>• Collective action problems</td>
</tr>
<tr>
<td></td>
<td>• Efficiency</td>
</tr>
<tr>
<td></td>
<td>• Harm to self</td>
</tr>
<tr>
<td></td>
<td>• Collective benefits</td>
</tr>
<tr>
<td><strong>Legitimate ends of public health</strong></td>
<td>• Prevent major threats to health</td>
</tr>
<tr>
<td></td>
<td>• Bring about public goods</td>
</tr>
<tr>
<td><strong>Means in need of justification</strong></td>
<td>• Coercive</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Agent</strong></td>
<td>• Government</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 5.1: Two conceptions of justified public health interventions

\textsuperscript{3} My distinction between public health minimalism and maximalism only partially overlaps with the way Rothstein and Verweij and Dawson map the terrain. These scholars are essentially concerned with the identification of the specific object of public health. Verweij and Dawson draw a wedge between narrow and broad conceptions of public health, rather than conceptions of justified public health interventions (2007). Rothstein's distinction between “government intervention as public health” and “population health as public health” is to some extent closer to my set of distinctions. Note also that Rothstein discusses what he labels “human rights as public health” (2000). According to this view, any violation of human rights counts as a public health issue; in addition, well-being and health are coextensive. As many public health ethicists believe this third position is conceptually and normatively flawed, I will not discuss it here (Gostin 2001; Rothstein 2002; Powers and Faden 2006; and Verweij and Dawson 2007).
The minimalist conception has been developed around extremely demanding standards for the justification of coercive governmental interventions to protect and promote public health. On public health minimalism, only the protection of major threats to health and the efficient delivery of genuine public goods warrant the use of the coercive force of the state. That being said, in the next section I argue why public health minimalism may have tacit concerns with many nudges, and not just with direct forms of state coercion that have formed its central explicit topic.

Public health maximalism is the view that a much broader range of interventions might be justified on a distinctively public health basis. On this view, public health is concerned with the protection and promotion of population health. More expansive justificatory considerations lend support to a variety of coercive or noncoercive actions performed by state and nonstate agents. The maximalist framework applies therefore more directly to the justification of public health nudges than does minimalism, and is more likely to back the use of health nudges for public health reasons.

Public health minimalism and maximalism are the two ends of a spectrum, and public health ethicists may also defend positions that are close, but not identical to each polar extreme. For instance, some authors who share the minimalists' overall orientation may nonetheless favor more inclusive justificatory strategies for public health interventions, such as the appeal to certain conceptions of social justice, without espousing a full-blown maximalist view (e.g. Robert M. Veatch).

In this chapter, my aim is not to defend the merits of public health minimalism or public health maximalism. It is instead to highlight their characteristics, the challenges
they each face, and their ramifications for the justification of public health nudges. If successful, this chapter will clarify the discussion on the proper justification of nudges.

4. Public Health Minimalism and the Ethics of Nudges

For public health minimalists, the main normative task is to justify the use of the coercive powers of the state for the sake of public health (Rothstein 2002, 146). Minimalists claim that coercive governmental public health interventions may have one of two justified goals.

Health-related public goods

The first goal of justified public health is to design interventions that bring about health-related public goods that individuals, left to their own devices, would be unable to create or would create in a much less efficient manner. Minimalists maintain these interventions are justified because they solve collective action problems.

Public goods, strictly speaking, are goods that can be enjoyed by all (nonexcludable), cannot be broken up into individual goods (indivisible), and cannot be depleted by use (non-rival) (Hechter 1987). For instance, according to Dawson, “the

---

3 As George Klosko notes, “the class of indispensable public goods is perhaps small” (2005, 7), and that it would only include some core public health functions. But what “indispensable” means is a matter of dispute. Some classical liberals may think that it only covers public goods indispensable for leading an “acceptable life.” Others might argue that the aim of political societies is to create the conditions for human flourishing, not mere survival.
existence of herd protection through vaccination for at least some contagious diseases brings about an important public good” (2007, 161). Vaccination generally provides benefit to the individual who gets immunized against a disease (a private good); it also raises the general level of immunity among the relevant population, decreasing the risk of outbreaks of contagious diseases, and therefore providing herd protection to the population as a whole (a public good). Herd protection is obviously indivisible, since it cannot be divided up and distributed among individuals. It is non-rival in the sense that me benefiting from it does not deprive you of a chance to benefit from it as well. It is also non-excludable because those who are not vaccinated (free-riders; those who cannot get vaccinated for medical reasons; those who refuse to get vaccinated for other reasons) and those who show a weak immune response to the vaccine cannot be excluded from the benefits of herd protection. This public good cannot be obtained without collaborative endeavor. Examples include vaccination for polio or diphtheria (Dawson 2007, 164).

However, not all public health goals that are sometimes loosely labeled “public goods” bring about public goods in the technical sense. Examples include programs dedicated to the eradication of poverty-related diseases; subsidized antiviral therapy for suffers of HIV/AIDS in low-income countries (O'Neill 2011, 8); and immunization programs for Hepatitis B among high risk groups without offering post-vaccination blood tests to make sure low-responders are all offered extra shots (Dawson 2011, 15). These interventions might be quasi-public goods that only partially meet the conditions for genuine public goods, or shared goods that meet alternative conditions (the so-called collective goods, club goods, and private goods enjoyed by large groups). Because most healthcare services are scarce, these interventions bringing about quasi-public goods or
shared goods do not meet the excludability and non-rivalry conditions of public goods.\textsuperscript{5}

Minimalists who think that the proper use of state powers (from mandatory vaccination to taxation) is justified only if it brings about genuine public goods might be tempted to deny that quasi or shared public goods can be defended on distinctively public health grounds. Some people will have to involuntarily endure burdens without any tangible benefit. Minimalists who still want to support these programs might have to argue that all can still get indirect benefits from them. Indeed, the whole population may benefit from a healthier and potentially more productive population (Faden and Shebaya 2010). All individuals can get mutual benefits through a reciprocal scheme in which burdens and benefits are overall equally distributed, although not through each program. But some individuals may in fact, predictably, never benefit as much as others. In that case, defenders of those programs would need to appeal to a counterfactual notion of reciprocity best articulated by Philippe Van Parijs:

“I do this for you, not because you will give me something equivalent in exchange (whether certainly or probabilistically), but because I believe I could have been in your position and you in mine (since we are members of the same people), and I trust that (as a member of the same people, sharing the same conception of what counts as being ill, etc.) you would then have done for me what I now do for you.” (Van Parijs 1999, 3)

**Major health threats**

For minimalists, the second goal of public health is to avoid or mitigate the adverse impacts of major health threats. The behavior of individuals, the spread of diseases, or environmental factors can cause these major health threats. All other

\textsuperscript{5} Vaccination for tetanus only brings about a private good because it does not protect others from the infection (Dawson 2007, 165).
potential goals of public health “lack the urgency and public health effects necessary to require universal participation” because they do not entail the possibility of serious harm to others (Rothstein 2002, 147).

In sum, the thin set of reasons for public health measures minimalists would deem acceptable extends only to avoiding or mitigating harm to others and solving collective action problems. For minimalists, the state could conduct a policy aiming at the promotion of population or individual health (and not one of the two more specific goals I mentioned) only if it obtains the willful collaboration of individuals through persuasion, informed consent, or shared decision-making (e.g., Deposit Contract).

Let me illustrate the minimalists' conception of an alternative to state coercion. Suppose lifestyle choices have a significant impact on healthcare expenses because some individuals' lifestyle entails more health risks, and therefore leads to financial costs that others might have to pay for (through risk-pooling insurance schemes or loss in total productivity). Instead of coercing them to adopt another lifestyle, individuals who deliberately choose a lifestyle that does not coincide with what contemporary medicine deems a healthy lifestyle could actively select and purchase an insurance policy (or equivalent in a single-payer public plan) with a self-tailored list of medical services that would be covered (in addition to a single-list of basic, consensual, medical services) (Veatch 2009, chapter 17). This strategy of internalization of externalities is an alternative to Intelligent Assignment, because it minimizes recourse to defaults and is more respectful of people's reflective choices.

To conclude, the reason public health minimalists are conservative with respect to the goals that authorize the use of the coercive powers of the state is that whenever an
intervention is labeled “public health,” it inherits the *prima facie* moral justification attached to the legitimate use of coercion by public authorities of modern democracies, who act in the name of the people for the sake of urgent public health improvements.

**Minimalism, coercion, and nudges**

What does a minimalist conception of public health entail with respect to population-wide health nudges? At first glance, public health minimalists would simply not discuss the ethics of nudges because they are exclusively concerned with governmental coercive interventions. Because health nudges are, by definition, substantially noncontrolling, they are not coercive. In addition, nonstate agents often initiate or implement public health nudges, but minimalists do not offer guidelines for assessing the behavior of private sector or civil society, unless they implicitly deem any influence on health choices exerted by nonpublic agents illegitimate, a point they would need to justify. Within the minimalist’s framework, it is not clear what justification, if any, noncoercive interventions to promote collective health would require.

However, public health minimalists could argue that their broad philosophy grounds concern about certain health nudges, even on the assumption that they are not directly coercive. This makes minimalism relevant to the ethics of nudging. Both considerations they invoke are prudential in spirit.

The first argument is that the difficulty in telling nudges apart from behavioral prods is a sufficient reason for not giving them the benefit of the doubt. They could argue that, although prods are not coercive, substantially controlling influences raise concerns similar enough to those typical of coercion, since they also restrict freedom of choice. If
nudges and prods use the same shallow cognitive processes, and prods are substantially controlling, it is more prudent to treat candidate nudges as prods to be on the safe side.

This argument is similar in form to the discussion in research ethics about the criteria for distinguishing clinical research activities from clinical practice activities. For instance, the *Belmont Report* (1978) states that one of the necessary conditions for an activity to qualify as research is that it is “designed to develop, or contribute to, generalizable knowledge.” Practice activities are, in contrast, “designed solely to enhance the well-being of an individual patient or client.” But many activities have dual aims by design: they are expected to promote the well-being of individual patients, and to produce generalizable knowledge useful to treat future patients. Should activities with dual aims be reviewed by IRBs whose function is to secure the conformity of research proposals with the ethical norms that govern research on human subjects?

One possible response is that only activities whose main aim is to produce generalizable knowledge are research activities for the purpose of review. After debating the issue, the members of the National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research who contributed to the *Belmont Report* decided otherwise. They argued that any activity having even a subcomponent with the aim of developing generalizable knowledge should be reviewed as research by IRBs, even if its main goal is clearly to treat individual patients (Beauchamp and Saghai 2012). The appropriateness of such a prudential move depends on whether there is, on balance, sufficient reason to justify a risk-averse attitude toward activities that are not exclusively aimed at bringing about benefits to individual patients.
The risk-averse strategy may seem justified on the basis that going through IRB review is at worst an administrative burden, but it is justified because it could potentially lead to the detection of morally problematic activities. However, given that foregoing an activity or delaying it may also have adverse consequences for the well-being of individuals and populations, or even violate some of their rights (for example, to a sufficient level of health), the rationale for a risk-averse attitude is disputable (Saghai and Beauchamp in preparation).

Similarly, treating nudges as prods for prudential reasons may entail adverse health consequences, or the violation of rights, if the goal of the health nudges cannot be pursued by comparably efficacious measures. The burden of proof is on the minimalists who must demonstrate why the risk of mistakenly believing a prod to be a nudge is much worse than not implementing a real nudge because it is treated like a prod.

According to the second prudential argument for including health nudges within the minimalist framework, nudges could, when taken as a group, fall under the category of potentially coercive interventions. This is due to the fact that the cumulative effect of nudges targeting liberties that play no substantive role in self-determination could lead to a gradual shrinking of the liberty sphere in a manner that would gradually and indirectly undermine more important liberties.

For instance, the use of several nudges aimed at altering dietary choices could have additive or multiplicative effects, leading to the justified perception that one's way of life is strongly constrained or stigmatized by the state. Public health minimalists could argue that even when the agent of public health is a strongly centralized government, the coordination of nudge policies altering minute aspects of daily life is virtually impossible.
Therefore, although nudges taken in isolation would only have benign effects on overall human liberty, their cumulative effect might lead to unpredictable but tangible negative consequences on important liberties. To give a concrete example, public health minimalists who might concede that obesity causes a serious threat to public health (and not merely a question of health promotion) could then object to the use of nudges altering dietary choices without having to abandon their conception of public health and health promotion.

The epistemic argument for minimalism

Another argument public health minimalists cite in favor of their view is that it is difficult to offer a compelling epistemic justification for any intervention but core public health activities. Third parties cannot know what will be of net benefit to individuals, as opposed to the ability to know what will eliminate harm. Veatch presents an argument along these lines based on two premises: (1) there are several irreducible dimensions of medical well-being (e.g., the preservation of life, the promotion of health, the alleviation of suffering, the treatment of diseases, etc.); (2) there are several irreducible spheres of global well-being (social, psychological, economic, religious, aesthetic, legal, organic, etc.) (Veatch 2009, 100-101). Provided that rational persons care about their total well-being, there is no reason to believe they would want to maximize their health, or even assign the same value to it as the majority of the people. How much each dimension is valued is a matter of personal choice that is not transparent to others. Therefore, the goal of making medical recommendations for health promotion is illusory (Veatch 2009, 101).

There is a further and independent problem with the epistemic basis for health
promotion. Suppose contra Veatch that (1) the list of dimensions of well-being were the same for everyone, so that they would all be “shared values” in a minimal sense of the term, and (2) that third parties could know what each person values most, thanks, for instance, to revealed preferences. It would still be highly improbable that all individuals would have a single ranking of dimensions of well-being. Given deep disagreement at the level of evaluative standards, two consequences could follow.

If nudges were noncoercive, the problem with promoting a particular ranking of values through nudges is not that nudges interfere with liberty, but that they lack any epistemic basis. An intervention that will not benefit the target population is unjustified.

However, if nudges were treated as coercive (because they might be confused with prods or because of their potential cumulative effects), then “it looks as if almost any collective pursuit of values will involve some citizens being coerced into pursuing a value that is less important to them than a value they had to give up (say, because they were taxed for the collectively pursued value)” (Gaus 2009, 93). Value pluralism is incompatible with the enforcement of a one value-system on the whole population because such a value-system could not be justified to all persons. It would violate the principle of state neutrality toward comprehensive conceptions of the good life (for a discussion of this point, see Raz (1986) and Wall and Klosko (2003)).

To conclude, minimalists who are traditionally exclusively concerned with assessing the merits of coercive interventions will nevertheless have reasons to regard many nudges as unjustified. The state should refrain from having recourse to many population-level health nudges, at least on public health grounds. Nudges that pursue one
of the two goals minimalists deem legitimate might be *prima facie* justified.  

---

5. Public Health Maximalism and the Ethics of Nudges

Let me turn to the maximalist conception of justified public health and its implications for public health nudges. I submit that maximalists would defend a broader range of justificatory considerations for public health interventions, and also that they would characterize more broadly the goals of public health. I start with the latter point.

**The goals of public health according to maximalists**

I follow here Verweij and Dawson who write that, “The practice of public health (roughly) consists of collective interventions that aim to promote and protect the health of the public” (2007, 21). Public health is thus doubly collective since its agents (the state, the private sector, civil society) and object (population health) are collective.

For public health maximalists, the goal of public health is to promote population health. Population health is essentially concerned with (1) the aggregate or the collective health of the members of a population; (2) the way determinants of health affect groups and create patterns of systematic disadvantage among them; and (3) the underlying

---

6. There are, of course, many other reasons one might invoke to object to the use of nudges that are independent of the core commitments of minimalists, such as the lack of transparency of certain nudges.

7. One of the main differences between defenders of public health minimalism and proponents of maximalism is whether the social determinants of health, which are not the proximate cause of ill health but the main distal root causes of unequal health status, are within the purview of public health authorities'
environmental and social determinants of morbidity and premature mortality (Powers and Faden 2006; Verweij and Dawson 2007, 22-24).

The aim of public health is to improve the health of populations rather than individuals, where “populations” may have various sizes (the global population, the population of a nation-state, or simply large groups of individuals with salient commonalities). Importantly, this focus on the health of populations means that improvements in the health of individuals should be measurable at the aggregate population-level.\(^8\)

A separate commitment of maximalists is that they consider the beneficiaries of public health interventions are *unidentified* and *unidentifiable* individuals, even if they belong to a sub-group smaller than the whole population of a nation-state. It is impossible to say in advance who will benefit from a public health intervention (e.g., reducing the

---

\(^8\) For some maximalists, these improvements can also be reflected in group-level, genuinely “collective,” effects that go beyond mere aggregation of the sum of individuals' health statuses. Thus, the collective (not merely aggregative) sense of a “healthy society” refers to a society that advances certain common goods that affect the population's health, like “shared norms and behavior that create and maintain rich substantive social values, such as solidarity and trust” (Dawson 2011, 16). But maximalists need not endorse such a view.
quantity of sodium in processed food to decrease the incidence of hypertension). It is also often impossible in retrospect to determine who has benefited from a public health intervention (e.g., who would have been killed by a bioterrorist attack that has been prevented).

I explained in the previous section that, for minimalists, the aim of public health is either to prevent major threats to health or to efficiently bring about genuine health-related public goods. Here, it is important to stress that maximalists are not wedded to the latter claim. Maximalists would argue that many of the interventions I discussed in the section on quasi or shared goods might still be directly justified on distinctively public health grounds, without relying on a counterfactual understanding of reciprocity, for instance.

---

9 Hypertension is a typical problem for preventive medicine. One might choose to address it by adopting a high-risk strategy, focusing on individuals who are either already sick or are judged most likely to develop hypertension. Alternatively, a population strategy would address the problem differently. This strategy might be appropriate when segregation between individuals who are seriously at risk and those who are not is very difficult to make. For instance, compare blood pressure with the flu. The former comes in degrees, with low emerging imperceptibly into the high, while the latter is closer to an all or nothing condition. As Geoffrey Rose puts it, “we need to consider the implications of a situation in which a small risk involves a large number of people, who in the high-risk strategy would be categorized as normal” (2008, 48). Sometimes the most effective strategy may be to implement mass scale interventions. Reduction in salt consumption in the whole population, to continue with this example, may be the best strategy to tackle one risk factor for hypertension because “many people must take precautions in order to prevent illness in a few... [health benefits] may be real, but they are likely to be delayed and to come only to a few who seek them” (2008, 7).
The maximalists' justification for public health measures

Public health maximalism also modifies the justificatory conditions for public health interventions. The justifications include those endorsed by public health minimalists, but may also extend to the following list that minimalists reject: (1) increasing the efficiency of the population-health measures; (2) fulfilling the demands of social justice (mitigating the effects of social determinants of health); (3) avoiding or reducing harm to self; (4) avoiding or reducing collective incremental harm; (5) creating collective benefits, and (6) pursuing the common good (Massé 2003, 116-117; Wertheimer 2007; Faden and Shebaya 2010). Naturally, all justifications and limits are in need of specification and defense. Some maximalists may reject certain justificatory considerations listed above. For instance, they may think that interfering with the autonomous choices of competent adults to avoid harm to self is never a good reason for the justification of public health measures. They firmly oppose paternalism (see chapter 6).

Public health minimalists are critical of the enlargement of the sources of justification for public health interventions that maximalists defend. Their objection is based on the belief that many interventions that go beyond minimally securing some basic level of public health have to rely at least on a thin conception of the good (some idea of what makes life better or worse for individuals), of the common good (Beauchamp 1988), and of the acceptability of the political enforcement of a view of the good. Whereas for Veatch, value pluralism is a fatal obstacle to knowing what would benefit third parties, other scholars disagree. They provide a variety of arguments to that effect. For example, some conceive health as a primary good, that is, an all-purpose
means that people want no matter what their life projects or conceptions of the good life are (Daniels 2008). Other theorists include health as a core element of well-being within a Neo-Aristotelian approach that focuses on salient human ends that each of us wants, no matter what else we want (Powers and Faden 2006, 191). Providing any further analysis of these theories falls beyond the scope of this dissertation. I simply acknowledge that such theories are necessary to make the maximalist machinery work.

Note also that the commitment to a thin conception of the good is compatible with the rejection of the thicker thesis that health is the ultimate value that ought to be maximized (what Daniel Callahan calls “healthism” (2000)). If the goal were to make sure that all residents of a country are sufficiently (but not maximally) healthy, maximalists would not be committed to the objectionable endless pursuit of health-related risk-reduction. However, what sufficiency entails is open to debate.

Responsibilities and obligations of the private sector and of civil society

In addition, given that for maximalists, the main goal of public health interventions is the promotion of population-level health, rather than the prevention of major threats to health and the delivery of public goods stricto sensu, they have no reason to exclusively focus on questions about coercive state interventions.\(^{10}\) Indeed, population-level health can be protected and promoted through a variety of means and thanks to a number of agents.

\(^{10}\) Minimalists’ position with respect to noncoercive interventions is ambiguous. For instance, sometimes Rothstein suggests that the focus of his minimalistic position is primarily coercive state interventions that are the most in need of justifications. At other times, he suggests that true public health activities are exclusively coercive, and that when government agencies act in their noncoercive role, they improve population health but not public health. He believes that the minimalists’ exclusive focus on state coercion
Maximalists would surely grant that many traditional public health interventions require state coercion (e.g., quarantine). But they would insist that the use of persuasion and of other modes of influence is also institutionally entrenched in the history of public health (e.g., public health information campaigns). Noncoercive forms of influences may be in need of justification nonetheless. Consider the moral issues involved in the practice of embedding abstinence or anti-drug messages into the story line of TV mini-series (Faden and Shebaya 2010).

Current maximalist views acknowledge the role of the private sector and civil society (Verweij and Dawson 2007), but they do not offer a detailed analysis of their responsibilities and obligations. For maximalists, nonstate agents of public health matter because they exercise some form of power that may have a profound and pervasive impact on people's lives, even though they are not authorized to make use of the full force of the state's coercive powers.¹¹

Maximalists typically acknowledge that public health requires the cooperation or coordination between groups of individuals, such as the government, international and

¹¹ For an analysis of forms of power, including noncoercive power like the control over political agendas, the framing of questions, or the exclusion of certain topics from public debate, see De Crespigny (1968), Lukes (2005), and Grant (2012).
domestic public institutions, civil, and private institutions. Public health interventions involve societal efforts.

If the private sector and civil society were indeed agents of public health, then they each might be “pursuing their own agenda with little coordination or accountability” (Rothstein 2002, 146). A state-centered maximalist framework for public health ethics is therefore at best incomplete because it does not allocate responsibilities and obligations among agents that are important according to its own standards (Rothstein 2002, 147).

*Chart 5.2* illustrates the direct and indirect pathways for public health interventions that a maximalist conception of justified public health interventions needs to cover:

![Chart 5.2: Public health intervention pathways](chart)

*Chart 5.2: Public health intervention pathways*
The state sometimes is the direct agent of public health through four pathways:

• Police powers: e.g., it can quarantine certain individuals, coercively depriving them of the liberty of movement.

• Taxation: e.g., it can increase the cost of cigarettes to discourage smoking.\(^\text{12}\)

• Public sector policy: e.g., the state doubles the value of food stamps when they are used in farmers' markets.

• Information campaigns: e.g., CDC’s information campaign stressing risks associated with the excessive use of pesticide.

The state may also indirectly affect public health through its impact on the behavior of other nonstate agents, such as the private sector (businesses, employers, contractors, health insurers) and civil society (nongovernmental or community organizations). There are at least six mediated action pathways:

• Delegation of state powers: e.g., delegating to schools the authority to vaccinate children.

\(^{12}\) Minimalists would defend regulations against second-hand smoking on public health grounds (serious threat to the health of others), but might deny that the state can legitimately interfere with self-regarding health choices of adult smokers. It is, however, unclear to me that Rothstein's version of minimalism can be used to criticize the governments' non-coercive efforts to abate the rate of smoking, since the focus of minimalism is on the proper use of the coercive powers of the state. Rothstein could simply argue that the anti-smoking campaigns are an unwise use of scarce public resources, and deplete budget that should be allocated elsewhere. Hybrid views might provide more conceptual resources for criticizing non-coercive anti-smoking measures. For a complex discussion of the moral justification for public health anti-smoking interventions, see Ashcroft 2011).
• Regulation of professions: e.g., regulating medical research (authorizing or not Clinical Trial Enrollment) or clinical medicine (Organ Donations);

• Regulation of business activity: e.g., product safety legislation (Grill 2009);

• Partnership with the private sector or civil society: e.g., Hand Washing is part of an international public-private partnership with soap companies (http://www.globalhandwashing.org/).

• Coordination of nonstate agents' activities: e.g., the state coordinates the efforts of the private sector (clinics, pharmacies) and civil society (community organizations) to organize the efficient delivery of the flu vaccine.

• Nonbinding guidelines: e.g., The CDC may entice private businesses to adopt Cafeteria; the Occupational Safety and Health Administration (OSHA) encourages employers to adopt non-mandatory workers' heat safety measures (2011).

Although the source of these measures is ultimately the state, the private sector and civil society agents involved in them might have their own responsibilities and obligations, as well as their own interests and agenda that might sometimes conflict with their public health role.

Finally, the agent of public health might be directly civil society or the private sector without any connection with state authorities. Examples include the private Dutch independent non-profit “iNudgeyou” (http://www.inudgeyou.com/), which develops nudges and incentives to help individuals and institutions solve public health and environmental problems. Another example is Stickk.com, a website that proposes financial pre-commitment contracts similar to Deposit Contract.
Whether the private sector and civil society act as mediators of state action or independently of the state to promote public health, these agents' actions are not necessarily justified by or answerable to the same norms that govern state action. Many defenders of nudges, such as Thaler and Sunstein, maintain that public and private institutions alike are entitled to nudge individuals. This is to neglect substantial differences in the relationship in which they stand, according to a coherent maximalist framework that would assign responsibilities and obligations to all agents involved in the protection and promotion of the public's health.

6. Conclusion

In conclusion, public health maximalism offers the conceptual space for the inclusion and justification of a broader range of population-wide health nudges under the umbrella of public health than does public health minimalism. But maximalist theories do not, at this stage, provide a framework for the moral evaluation of public health nudges that are not directly conducted by the state.
CHAPTER VI

NUDGES, PATERNALISM, AND DECISIONAL CAPACITIES

1. Introduction

In this chapter, I consider two problems with the use of public health nudges. First, I explore the role that paternalism can play in justifying or criticizing the use of public health nudges, concluding that paternalism's role in this debate is less central than it is usually believed (section 2). Second, I examine the Kantian and Millian theses that nudges are either impermissible or simply distasteful because they might have adverse cumulative effects on the nudgees' capacities to make decisions (section 3). I need to pause here to note that deliberative and volitional capacities are the focus of Kantians' and Millians' attention. Thus, I focus my discussion on those capacities. However, as I showed in chapter 2, I do not believe they exhaust the kinds of capacities involved in decision-making and choice (e.g., unconscious goal-conflict recognition and resolution capacities). Finally, I argue that public health nudgers have a duty not to weaken in the long run the nudgees' decisional capacities or the exercise of those capacities (section 4),
but that they do not have a similar duty to engage those capacities or to enhance them (sections 5 and 6).  

2. The Ethics of Public Health Nudges and Paternalism

In the previous chapter, I mentioned that a maximalist conception of justified public health might count the avoidance or the reduction of harm to self as a possible justification for public health interventions. The problem I now address is whether public health nudges can be paternalistic at all. I argue two points. First, nudges are not paternalistic in the classical conception of paternalism. Second, nudges are paternalistic on expansive conceptions of paternalism, but these conceptions are misguided. I conclude that paternalism cannot play a major role in structuring the nudge debate.

My approach to the concept of paternalism is guided by two desiderata. First, the definition of paternalism should not cast so broad a net that a myriad of prima facie innocuous actions or attitudes would count as paternalistic, depleting it of its moral import. Second, however paternalism is understood, it should not exclude by definitional fiat the possibility of cases of justified paternalism, on pain of begging the question.

---

1 Section 2 is partly based on an unpublished manuscript entitled, "A twin-aim theory of social justice: Self-determination, liberty, and public health, reconsidered," co-written with Madison Powers and Ruth Faden. The problem examined in section 3 was initially formulated in that manuscript.
What's paternalistic in libertarian paternalism?

The most prominent theories offered in defense of nudges have the ambition of making policies that are in some respect distinctively paternalistic acceptable to those who traditionally oppose them. Thus, the main defenses of the use of nudges have been given on paternalistic grounds, as their labels indicate: “libertarian paternalism” (Thaler and Sunstein 2003), asymmetric paternalism (Camerer et al. 2003), light paternalism (Loewenstein and Haisley 2008), and structured paternalism (Ben-Porath 2010). As these theories share a common understanding of paternalism, my discussion of the better-known version of them, libertarian paternalism, will cover them all.

Thaler and Sunstein write that, “The paternalistic aspect [of our theory] lies in the claim that it is legitimate for choice architects to try to influence people’s behavior in order to make their lives longer, healthier, and better” and that a paternalistic policy “tries to influence choices in a way that will make choosers better off, as judged by themselves” (2009, 5).

This characterization of paternalism is incomplete and vague. As Hausman and Welsh point out, the fact that an intervention makes someone better off according to her own view of what contributes to her well-being cannot count by itself as a reason to call an intervention “paternalistic,” rather than “benevolent” (2010, 126-127).

Furthermore, Thaler and Sunstein provide many examples of interventions that they label paternalistic without any remotely plausible reason. Many nudges they advocate are actually self-binding interventions to which the person has consented, and
which obviously stem from her own agency.\textsuperscript{2} Other interventions are clearly concerned with avoiding harm to others. Consider:

\textbf{Default Professional Immunization}. Healthcare providers working in hospitals are immunized against influenza by default, but are offered the possibility to opt-out.

Finally, some nudges are directed at protecting the vulnerable as a matter of justice, such as setting default clauses in marriage contracts in such a way that the most financially vulnerable party would benefit from a fair outcome in case of divorce.

The problem, however, runs deeper than what many philosophers take to be a simple misclassification (Frank 2008; Hausman and Welch 2010). The real issue is to identify the potential concern that paternalism raises and evaluate its relevance for the moral assessment of health nudges.

\textbf{Classical conceptions of paternalism}

The nudge debate reveals muddy problems about the very nature of paternalism. Since the publication of Gerald Dworkin's seminal 1972 article, numerous philosophers have attempted to characterize what counts as a paternalistic intervention, and to evaluate the normative significance of paternalism. However, conceptions of paternalism vary widely.

\begin{footnotesize}
\textsuperscript{2} See the recent debate on creating a smart card that would enable individuals to pre-commit to limiting their ability to purchase cigarettes beyond a certain quantity for a pre-determined duration (Chapman 2012; Colin 2012).
\end{footnotesize}
Take Dworkin's classical definition of paternalism:

By paternalism I shall understand roughly 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. (1972, 65)

According to this classical understanding of paternalism, nudges can never be paternalistic because they do not interfere with liberty in the sense Dworkin uses the term. As nudges do not restrict choice or preclude action, they are not paternalistic. Only fully or substantially controlling influences that interfere with liberty against the will of the recipient, and are intended for her good, counts as paternalistic (cf. Grill forthcoming). Dworkin makes this point explicit elsewhere:

Paternalism is the interference of a state or an individual with another person, against their will, and justified by a claim that the person interfered with will be better off or protected from harm. (2010)

**Paternalism as interference with autonomy without consent**

Many philosophers have offered much more expansive definitions of paternalism. Their aim is essentially to include cases in which interference is not limited to liberty of action, but extends to meddling with autonomous decision-making. For instance, when a physician lies to her patient, manipulates her, or withholds the truth from her for her good, he or she acts paternalistically, even though the physician does not interfere with the patient's liberty of action (Gert and Culver 1976). To account for these kinds of cases, Danny Scoccia proposes the following revised definition of paternalism:

Paternalism is the attempt to induce other people to act in ways that will benefit them or not act in ways that will harm them, which: (i) is motivated and/or justified by a beneficent concern for their welfare, and (ii) uses means other than reasoned persuasion.” (2008, 352)

---

3 The emphasis is mine.
Here, the requirement of acting against the will of the recipient is simply eliminated in favor of a broad understanding of the means the influencer uses. Those means only exclude rational persuasion. On this account, all nudges that are at least partly intended for the good of the recipient are paternalistic.

The problem is that this definition is so broad that when \( A \) offers a present to \( B \) (such as a gift card) he or she acts paternalistically. We seem to lose sight of what might be specifically troublesome with certain interferences with autonomy for the good of the influencee.

In a similar vein, other authors propose definitions of paternalism intended to cover interferences with autonomy that are not coercive. Their inclusive strategies also miss the point. For instance, Hausman and Welch, in an attempt to show that nudges are paternalistic, write that:

A policy is paternalistic if and only if it aims to advance the interests of some person \( P \) either (a) via influencing \( P \)'s choices by shaping how \( P \) chooses or limiting what \( P \) can choose or (b) some means that will take effect regardless of what \( P \) does and against \( P \)'s will. (2010, 129)

According to this disjunctive definition, an intervention need not be performed against the will of the influencee in order to be paternalistic. One objectionable implication of this view is that some nudges that the influencee welcomes are deemed paternalistic merely because they shape the way the nudgee makes a choice.

Dworkin, who once defended a classical version of paternalism, has recently offered an expansive amended definition of paternalism, which does not commit Hausman and Welch's mistake but still casts the net too broadly. He writes that, \( X \) acts paternalistically toward \( Y \) by doing (omitting) \( Z \) if and only if:
1. Z (or its omission) interferes with the liberty or autonomy of Y.
2. X does so without consent of Y.
3. X does so just because Z will improve the welfare of Y (where this includes preventing his welfare from diminishing), or in some way promote the interests, values or good of Y. (2010)

Dworkin substitutes the “against the will” condition with the much weaker requirement that the intervention be performed without the influencee's consent. Dworkin writes that, “The agent may neither consent nor not consent. He may, for example, be unaware of what is being done to him.” Consent can be read here to mean either hypothetical consent or explicit consent.

If by consent Dworkin means explicit consent, then some nudges could be paternalistic (but not self-nudges or other-regarding nudges that have been consented to). Unfortunately, this definition is so broad that it would cover many unwanted cases. Take the International Hockey Federation's mandate requiring the use of helmets in all hockey games for the sake of hockey players' health. The ban on helmet-less play was enacted without consulting each hockey player (Frank 2008). All three conditions for paternalism are satisfied since the players have no longer the liberty to wear helmets or not, they have not explicitly consented to the ban, and the ban is intended to benefit them health-wise. But it is possible that the ban is not performed against the will of the hockey players, because all, or at least most of them would prefer a ban. They believed they would be overall better off if none of them has the liberty to play without a helmet. In the absence of a ban, they know they would not be able to solve their collective action problem. If some of them were risk-takers who wanted to win at any price, they would have a strong unwanted incentive not to wear helmets in order to keep their competitive edge. As a result, they would predictably be overall worse off. I do not see what is paternalistic in such a ban, which is not performed against the will of the players, but only without their
consent. Given that there are multiple cases of this type, the moral import of paternalism so understood is rather unclear.

If by consent Dworkin means hypothetical consent, then whether an action is paternalistic or not depends on what the influencee would reasonably agree to. What reasonable persons would agree to is a disputed issue (Kuflik 2010). But at least when nudges are designed to prevent the predictable and clear-cut consequences of deficits in rationality, it might be argued that we would consent to them, and that as a consequence, they are not paternalistic.

This thesis need not be based on J.D. Trout's argument that cognitive and affective biases make actions involuntary (Trout 2005; 2007, 459). If such quotidian shortcomings were uniformly to deprive action of voluntariness, coercive interference with vital or consequential liberties as well as nonconsequential ones would be prima facie authorized. For instance, because of their severe deficits in rationality that make their actions substantially involuntary, the profoundly mentally disabled might not be authorized to exercise some liberties essential for leading a self-determining life (including reproductive liberties, and those linked to political participation). But I do not see any reason to think that the kind of cognitive and affective biases that public health nudges are usually designed to counter undermine decisional capacity to that extent. The criteria for deeming someone unable to make her own decisions because her actions are involuntary are deliberately set very high in order not to deprive others of their fundamental rights too easily (Charland 2008). The defense of nudges needs a much weaker claim: our cognitive and affective frailties sometimes justify hypothetical consent
to noncoercive interventions when nonfundamental liberties are interfered with for our own good.

**Paternalism as illegitimate substitution of judgment and agency**

Seana Shiffrin proposes yet a different expansive approach to paternalism that I reject:

I suggest that paternalism by A toward B may be characterized as behavior (whether through action or through omission) (a) aimed to have (or to avoid) an effect on B or her sphere of legitimate agency (b) that involves the substitution of A's judgment or agency for B's (c) directed at B's own interests or matters that legitimately lie within B's control (d) undertaken on the grounds that compared to B's judgment or agency with respect to those interests or other matters, A regards her judgment or agency to be (or as likely to be), in some respect, superior to B's. (2000, 218)

Shiffrin offers the following explanation of why paternalism is troublesome: it involves an insulting invasion of the influencee's sphere of legitimate judgment and agency. In short, the paternalist interferes with the influencee's liberty or autonomy through her judgment or agency without having the authority to do so. Remark that having authority to interfere need not necessarily be based on a person's explicit consent to a token intervention.4

Under Shiffrin's account of paternalism, some nudges count as paternalistic because the nudger considers his or her judgment or agency superior to that of the nudgees'. Arguably, many health nudges aim at compensating for the effects of cognitive biases, that is, deviations from patterns of optimal reasoning, and therefore imply that the nudger's judgment is superior to that of the nudgee. Some proponents of nudges acknowledge this point plainly: “In a sense, behavioral economics extends the

---

4 For a discussion of theories of authority and their connection to consent, see Raz (1986) and Shapiro (2002).
paternalistically protected category of ‘idiots' to include most people, at predictable times” (Camerer et al. 2003, 1218).

There are, however, at least four serious problems with Shiffrin's account of paternalism. First, she assumes, but does not offer, a separate theory of what constitutes someone's proper sphere of judgment and agency (Grill 2012). Shiffrin acknowledges the point in her article, when she discusses the Millian distinctions between self-regarding and other-regarding behavior (2000, 239). Given our interdependence, there is hardly any behavior that does not affect others in some way or another. Take health-related choices: even our dietary choices have effects on others. For instance, trends in weight gain are strongly correlated to whether our friends also gain weight, and weight gain spreads almost like an infectious disease in social networks (Christakis and Fowler 2007).

Second, no matter what is one's account of spheres of sovereign judgment and agency, it is hard to deny that sometimes making a judgment for others is clearly legitimate because the judgment call does not belong to that person in the first place (cf. De Marneffe 2006). For instance, in Anesthesia Equipment, hospital administrators and medical equipment companies judge that in order to minimize the risk of medical accident, it is better not to create conditions likely to require medical practitioners to deliberate over some of their decisions. A third party makes the judgment, but there is no “illegitimate substitution,” or even simple “substitution” of judgment because, in this context, it is not up to the physician to decide what is the acceptable level of risk exposure.

Third, the view that A’s judgment is superior to B's judgment might be correct. Camerer and colleagues' provocative claim is not that nudgers are entitled to always feel
superior to the nudgees. They make the more plausible claim that it is part of the human condition to think or behave occasionally, but predictably like an idiot. For instance, if $A$ is in a cool, non-emotional state and $B$ is in a hot, emotional state, $A$ might be in a better position to make a judgment about action $X$ that $B$ considers doing than $B$. My claim is not that being in an emotional state necessarily clouds one's judgment. Sometimes, the opposite is true, for example, when one should feel indignant in the face of injustice. In the arena of health-related decisions in particular, we may have trouble imagining what it feels like to be in a hot state when we are in a cool state, and we may misrepresent what might be like to be in a cool state when we are in a hot state. This phenomenon is known as the hot-cold empathy gap (Loewenstein 2005).

Fourth, even if Shiffrin could address these three problems, she fails on the second requirement I stated earlier with respect to a plausible conception of paternalism. Clearly, Shiffrin's strongly normative definition of paternalism leaves no conceptual space for the possibility of justified paternalism, since it is by definition an illegitimate substitution of judgment or agency. Taken together, these four reasons are sufficient for rejecting Shiffrin's conception of paternalism. It would be rather unhelpful for the ethics of public health nudges.

**Conclusion**

The upshot of this argument is that it is far from clear in what morally relevant sense nudges are paternalistic. This is why, in my opinion, the nudge debate has been excessively focused on paternalism, instead of discussing what the substantive issues are. In the next section, I identify one of these substantive, though neglected issues: we feel
uncomfortable with the use of nudges (whether for the good of the nudgees, the nudger, or third parties), because nudges involve some kind of meddling with our decisional capacities.

3. Decisional Capacities and the Kantian and Millian Case Against Nudges

In this section, I first examine Kantian and Millian arguments against nudges based on the view that they interfere with decisional capacities in an objectionable manner.

Some Kantian deontologists may argue that nudges express disrespect for the dignity of persons because the nudger does not present the nudgee with reasons for action. Nudges undermine the conditions for personhood.

Some Millian consequentialists may claim that nudges undermine the conditions for human flourishing by preventing the exercise and development of decisional capacities. Nudges undermine the conditions for adulthood. Even self-nudges raise concerns.

Those with these worries might claim that public health agents should abstain from the widespread use of nudges, and from encouraging others to use them.

Decisional capacities and the Kantian case against nudges

Some opponents of nudges might suggest that nudges express a form of disrespect toward the influencee because they primarily activate shallow cognitive processes. In short, by failing to engage our deliberative and volitional capacities, they treat us as beings not
worth giving reasons to, and they make us dependent on an intentionally arranged choice environment.

According to Kantians, decisional capacities are the core of what makes us agents who are able to commit themselves to projects and take responsibility. They are necessary conditions for personhood. Decisional capacities have two components. One of them is deliberation, the capacity to make judgments, to select one's ends based on reasons for action, and to determine what means are instrumental for achieving those ends through practical reasoning. The other is the will, that is, the capacity for voluntary action or the execution of the outcome of deliberation.

To make sense of the Kantian concern, we must avoid targeting a comic-strip version of Kantian ethics. According to such a version, people deserve to be respected because they are by and large fully rational and fully able to exercise self-control. As Onora O'Neill (1984) has successfully argued, such a view of persons is false given how frail our cognitive and volitional equipment is. The only thing that is really disputed is the extent of our shortcomings.

It is important to understand what respect for finite and imperfect beings capable of reason and will entails. We do not want to conclude that if a person is at some point not fully rational and self-controlled, or does not perform the optimal action, then her decisions can be disregarded and her decisional capacities need not be engaged at all. So, the version of Kantian ethics I have in mind endorses the realistic view that our decisional capacities are always limited, and at predictable times impaired to various degrees. I return to the problem of the optimal allocation of one's limited cognitive resources in the last section of this chapter.
What then does the requirement of respect for finite persons imply when it comes to their deliberative and volitional capacities? Kantians might argue that it implies abiding by several duties to others and to oneself (White 2011): a duty not to weaken in the long run decisional capacities or undermine their exercise; a duty to engage deliberative capacities when they are present, rather than bypass them; a duty to enhance decisional capacities.

These duties express a commitment to an aspirational ideal that ought to govern individual and institutional interactions. Kantians might argue against harnessing the power of cognitive biases because they judge that bypassing deliberative capacities is manipulative and expresses contempt for the dignity of persons.5

Decisional capacities and the Millian case against nudges

Millian consequentialists also value decisional capacities, but for a distinct set of reasons. For them, in order to flourish, humans need to exercise and develop their unique many-

5 There is little consensus on what manipulation is, and whether it is always wrong (Rudinow 1978; Greenspan 2003; Buss 2005; Handelman 2009). But Blumenthal-Barby's and Burroughs' definition of manipulation shows why manipulation matters for Kantians: “Manipulation occurs when one influences another by bypassing their capacity for reason, either by exploiting nonrational elements of psychological makeup or by influencing choices in a way that is not obvious to the subject” (2012, 5). The question of manipulation is connected to the issue of the transparency of nudges, in particular, in Cafeteria. For a discussion of this issue, see Bovens (2008), Beauchamp and Childress (2009, 211-212), and Blumenthal-Barby and Burroughs (2012). If my understanding of nudges is correct, then the question of transparency looses some (but certainly not all) of its urgency. I have indeed argued that, under some conditions, we can be in control of our behavior and choices even if the nudge is covert, thanks to our goal-conflict recognition and resolution capacities.
sided capacities. While contemporary Millians may not agree with every detail of Mill's view, their views are in the spirit of Mill's conception of human capacities (Powers and Faden 2006).

What matters to Mill is that people exercise and develop their unique faculties that are conducive to higher pleasures. Higher pleasures are identified through a test, and they are tightly connected to the exercise and development of certain faculties that are the most conducive to experiencing those qualitatively superior pleasures. In brief, A is a higher pleasure than B if people who have experienced both, and had access to an appropriate range of options, after deliberation, consistently choose A. A higher pleasure is such that an expert “would not resign it for any quantity of other pleasures which their nature is capable of” (Mill 1991c, 139, 140-141). Higher pleasures are associated with the exercise and development of higher faculties typical of humans: the intellect (that is, theoretical and practical reasoning), feelings (the emotional life), imagination (creativity), and moral sentiments (feeling of sympathy toward sentient beings) (Mill 1991c, 138; Mill 199a, 65).

For Mill, decisional capacities matter because they are a necessary component of individuality, which is itself one of the core components of well-being. More precisely, what Mill calls “individuality” covers both autonomy and authenticity.

Roughly speaking, for Mill, I am autonomous when two conditions are met. First, my theoretical and practical deliberation (using faculties such as reasoning, imagination, and feelings) results in beliefs and actions that are not controlled by others, and are not copied from others without independent reflection. Second, I am able to act on my
decision (I do not suffer from weakness of the will). So, deliberative and volitional capacities both matter.

The authenticity condition requires having desires, projects and a way of life of one’s own. My desires are not copied from others, even after reflection, but reflect my own nature that I need to discover. Part of the Millian conception of individuality involves self-knowledge. Mill uses the romantic language of “character,” and “spontaneity” to point out what we would now call authenticity.

Given this background, even if contemporary Millian consequentialists would not espouse Mill's romantic inclinations, they would still value self-determination for similar reasons (see chapter 3). From this perspective, nudges might seem threatening because they may deprive individuals of the opportunity to properly exercise the faculty of judgment, discernment, and will. By harnessing mental shortcuts, they may also weaken our reasoning capacities or undermine their exercise.

On a more societal level, nudges might be objectionable from a Millian perspective because they fail to encourage citizens' active participation in solving problems through an open discussion in the public space about the ends and means of policies (Burgess 2012).

Finally, nudges that are designed to address problems of weakness of the will might aggravate them instead, by making us overly dependent on our decisional and technological environment. They do not help cultivate independence and nurture an authentic life in which the actions I perform are truly mine. Nudges are at best stopgaps; they do not transform a weak person into a resolute chooser.

---

6 See Selinger (2012) for a clear statement of this objection, although he does not endorse it.
My response to Kantians and Millians who may have these worries is two-fold. First, I agree with them that agents indeed have a duty not to weaken decisional capacities or undermine their exercise, which rules out the use of certain public health nudges. Second, I argue against Kantians and Millians that there is no perfect duty to engage or enhance decisional capacities whenever one can. Therefore, I do not believe public health nudges that do not engage or enhance decisional capacities are necessarily wrong or distasteful.

4. Nudges and the Duty Not to Weaken Decisional Capacities or Undermine their Exercise

Kantians and Millians are, I believe, right when they suggest that we have a duty toward others not to weaken their decisional capacities or undermine the exercise of those capacities. I am much less comfortable with the view that we have a similar self-directed duty (that a Kantian libertarian like White (2011) defends), a disputed notion of which Mill is also critical, and that I will not address (Ogien 2007, chapter 2; Dillon 2010).7 Nudges may weaken our decisional capacities, or undermine their exercise, in four ways. First, nudgers might unwisely stir the influencees' emotions when they activate certain nondeliberative cognitive processes. Take Hand Washing: in certain contexts,

7 “I am the last person to undervalue the self-regarding virtues; they are only second in importance, if even second, to the social. It is equally the business of education to cultivate both. But even education works by conviction and persuasion as well as by compulsion, and it is by the former only that, when the period of education is past, the self-regarding virtues should be inculcated. Human beings owe to each other help to distinguish the better from the worse, and encouragement to choose the former and avoid the latter. They should be forever stimulating each other to increased exercise of their higher faculties, and increased direction of their feelings and aims towards wise instead of foolish, elevating instead of degrading, objects and contemplations. But neither one person, nor any number of persons, is warranted in saying to another human creature of ripe years, that he shall not do with his life for his own benefit what he chooses to do with it” (On Liberty, chapter 4, 84) [my emphasis].
deliberately activating the negative emotions of disgust and fear of contamination may overshoot. For instance, it could generate social panic or the stigmatization of certain social groups scapegoated as the primary contaminators, creating or reinforcing long-term biased judgments.

Second, certain interventions create or reinforce bad or unreliable modes of reasoning. Some candidate nudges are susceptible to have such effects, for instance, **HIV-Test Cash Transfer**, which triggers the *ex ante* valuation of immediate benefit of the small cash transfer in order to induce her to modify her perception of the value of long-term benefits that she undervalues (see chapter 2). If it turns out that such strategies have unwanted long-term effects (an empirical question I cannot answer), such as reinforcing a present-day bias, they should be avoided. This is an important conclusion because it is usually assumed that nudges that activate nondeliberative cognitive processes are the most problematic type of nudges with respect to their effect on decisional capacities. My claim is that the account of nudging mechanisms sketched in chapter 1 distinguishing nudges relying on nondeliberative cognitive processes from those relying on incompletely deliberative cognitive processes is useful because it enables us to identify moral problems specifically attached to the activation of incompletely cognitive deliberative processes. So, nudges relying on an incomplete form of deliberation might sometimes weaken our decisional capacities or undermine their

---

8 In Ghana hand washing with soap after using the toilet was originally as low as 3%. The **Hand Washing** TV ads resulted in a 13% increase in hand washing (better than usual argument-based campaigns designed to change health-related beliefs). This can be explained by reliance on the impulsive system (a conditioned stimulus (toilet) is associated to an unconditioned stimulus (impulsive feeling of disgust when exposed to contamination)), and on the affective system (Vlaev and Darz 2012).
proper exercise in a distinctive manner. Note that although Millians are right to believe that decisional capacities are like muscles and need to be exercised, it does not follow that they should be exerted to their full potential on each single opportunity. Over-exertion of limited capacities might be counter-productive.

Third, the systematic recourse to nudges in order to avoid public debate about nonconsensual issues might have a negative impact on collective deliberation. This is the essence of one of the Millian worries I mentioned earlier. Here, the problem is not the weakening of individuals' decisional capacities but of our capacity and inclination to collectively deliberate about the ends of social life (John et al. 2011).

There might be, for instance, strong social disagreement about the following intervention:

**Default DNR.** By default, hospitals assume that end-stage demented patients who have cardiac arrest should not be resuscitated (DNR order) (Volandes and Abbo 2007).

The problem is that if these types of critical rules are not the result of open public deliberation, in which citizens, patients, or clients actively participate, our collective capacity to discuss the ends we want to pursue together might be progressively nibbled down. We might simply get used to having important questions taken off our hands. These decisions are not technical issues that could be handled by institutions that do not benefit from any strong democratic legitimacy. Certain decisions need to be made
through some form of deliberative and participative democratic procedure, and can only later be implemented by the use of nudges that stakeholders judge acceptable.

My position is therefore substantially different from the view John et al. defend (2011). They maintain that deliberative democracy is an alternative to the use of nudges, and believe nudges are impermissible. I claim that the goal of some nudges has to be the object of public debate, but that the use of nudges is not antithetic with a deliberative democratic procedural requirement to determine their ends.

Fourth, habituation to social scaffolding might also weaken our decisional capacities or undermine their proper exercise (Selinger 2012). For instance, White (2011) vigorously opposes the view that problems of weakness of the will can be addressed by removing the object of temptation from the person's choice environment or by artificially reducing temptation through the alteration of how one perceives a choice situation. Although I agree that nudges may undermine the exercise of volitional capacities, I disagree with White's argument and his assessment of the extent to which this is a reason to think nudges are impermissible.

White's reasoning is based on an empirical assumption that he never defends, namely that nudges, prods, or incentives can only have superficial effects on the will and cannot change our preference structure for good. In some cases, preference change might be resilient even in the absence of social scaffolding through habituation. Whether the will can be reinforced through social scaffolding is an empirical issue that is still open to debate. Of course, it is also possible that the use of a myriad of non-coordinated public health nudges might indeed weaken the will. We should then avoid that kind of nudge. However, the dispersal of sources of nudging (or encouragement to self-nudge) makes
setting limits to recourse to nudges a difficult task (Bovens 2012). I intend to address this issue in future research.

5. Nudges and the Duty to Engage Decisional Capacities

While I have argued that there is indeed a duty not to weaken in the long run decisional capacities or undermine their proper exercise, I deny that there is a similar duty to always directly engage decisional capacities.

Many behavioral triggering interventions (nudges or prods) fail to engage deliberative capacities because they bypass conscious deliberative capacities altogether, but I don't see why recourse to these interventions would violate a duty. A good example is:

**White Stripes.** To prompt drivers to take a dangerous S curve more cautiously, a series of white stripes are painted on the road. “When the stripes first appear, they are evenly spaced, but as drivers reach the most dangerous portion of the curve, the stripes get closer together, giving the sensation that driving speed is increasing. One’s natural instinct is to slow down” (Thaler and Sunstein 2008, 37-39).

This example shows that some interventions help us navigate our environment in the most effective way through the use of affordances, that is, salient features of the
environment or objects that suggest how they can be used by agents with a particular set of capabilities and goals (Norman 1988). They virtually eliminate conscious deliberation to allow the smoothest interaction between the situated agent and her environment. Deliberative capacities are neither weakened nor engaged, but cognitive load is alleviated to leave room for more demanding tasks for which it is important that we slow down and consciously deliberate.

More generally, maintaining that there is a duty to engage decisional capacities whenever one can would entail that authorities should not facilitate safe and smooth driving, but should instead create conditions for the maximal exertion of the conscious capacities involved in driving (making certain choices). Society ought not be entirely organized around the systematic cultivation of the disposition for “the voluntary attempt to overcome unnecessary obstacles,” to quote Bernard Suits's definition of playing a game (2005, 55).

White Stripes is an example of a behavioral triggering intervention that bypasses conscious deliberation altogether. But sometimes nudges do not engage deliberation directly, but nonetheless affect deliberation positively. Consider,

**Food Traffic Lights.** Food manufacturers or distributors use easily recognizable “traffic light” intuitive and salient symbols (e.g., green for low, yellow for medium, red for high) to indicate how calorific every item of processed food is.

---

9 However, as Evan Selinger and Kyle Powys Whyte argue, they do not bypass the agent's nonverbal interpretative capacities, since even unconsciously recognizing an affordance depends on our ability to mobilize a culturally acquired know-how (2010; 2011).
This is a case in which a nudge works by nondeliberative cognitive processes (here, our affective systems (Vlaev and Darz 2012), eliciting emotions typically attached to colors, such as red meaning “stop,” and green meaning “pass”), but its function is one of cuing. The consumer's attention is attracted to a salient feature of a product that is material for her decision, and that she is likely to neglect. Food traffic lights are designed to serve as a starting point for the consumer's deliberation over her choices. That input facilitates decision-making because using nondeliberative cognitive processes is significantly less cognitive-resource hungry than either incompletely or completely deliberative cognitive processes, for instance adding up the actual number of calories per serving.

To take another example, Anti-Suicide Bridge Flyers is one of the smartest examples of subtle cuing that rightly avoids direct deliberative engagement. Suicidal individuals are usually not responsive to reasons (especially without human interaction, through a flyer). They are more likely to be positively affected if their attention is directed toward their body and noncognitive internal states, instead of more direct deliberative engagement that may feed in their pessimistic ruminations.

These examples show why we should reject the Kantian or Millian claim that we have a duty to engage decisional capacities parallel to the duty not to weaken in the long run decisional capacities or undermine their proper exercise. A duty to engage decisional capacities can at best be an imperfect, rather than a perfect duty. Imperfect duties are not correlative to rights of identifiable right-holders (Tasioulas forthcoming). The duty-bearer
may perform the action only occasionally. The violation of imperfect duties warrants no sanction. Fulfilling a praiseworthy imperfect duty is an act of generosity or benevolence.

It is difficult to imagine that one would have a perfect duty to engage others' decisional capacities whenever one can do so, and how such duty could be implemented (if ought implies can). If there is such a perfect duty, then Cafeteria, Generic Medication, White Stripes, and Food Traffic Lights are duty-violations, a claim that seems extremely counterintuitive. Even the weaker claim that it would be, all things considered, more desirable if we engaged others' capacities is dubious. We do not live an inauthentic life if we fail to deliberate over trivial matters or are not put by others in a situation that makes deliberation more likely. What I claim is that we should not give the activities of theoretical and practical reason an undeserved privilege. But nudges are not the only alternative to persuasion, or the use of sticks and carrots. We're more complex creatures. As Jodi Halpern and Maggie Little argue, at the core of motivation is the normative activity of coping, that is, maintaining a “sense of self and world as meaningful, stable, and secure” (2009, 144). This type of activity may require a subtle engagement with our nondeliberative capacities.

6. Nudges and the Duty to Enhance Decisional Capacities

In this section, I argue against the view that we have a perfect duty to enhance decisional capacities whenever we can. Such a duty would lead to implausible claims that I examine in turn. First, public health authorities would have to maximize the opportunity to give
explicit consent to as many token actions as possible. Second, they would have a duty to use mandated choice instead of defaults. Third, they would have a duty to teach the public metacognitive methods to avoid cognitive or affective biases, and thereby enhance their decisional capacities.

First, the very idea of requiring explicit consent (or other forms of active decision-making) for as many actions as possible faces a theoretical difficulty. O'Neill notes that,

All consent is consent to some proposed action or project under certain descriptions. When we consent to an action or project we often do not consent even to its logical implications or to its likely results (let alone its actual results), or to its unavoidable corollaries and presuppositions. Put more technically, consenting (like other propositional attitudes) is opaque. When we consent we do not necessarily “see through” to the implications of what we consent to and consent to these also. (1984, 174-175)

Because of the opacity of consenting, it is impossible to require that people offer others an exhaustive list of the clauses implied when they request consent, and it is impossible to expect the person who gives her consent to deliberate about all possible ramifications of a choice. The expansion of situations that requires explicit consent would fail to enhance others' decisional capacities. Rather, given our cognitive biases and cognitive load, if too many options are presented to us, we may not be able to sort them, and our deliberative capacities may be so overwhelmed that we may predictably not be in a position to exercise them. Cluttering our mind with trivialities would only undermine the exercise of practical reason (Harman 1986, chapter 2). This is why some arrangement of the choice and informational environment that facilitates decision making, such as grouping options (e.g., surgery versus medication) instead of presenting the chooser with
an exhaustive list of possible options (e.g. 5 types of surgeries and 9 names of drugs), is necessary for any meaningful deliberation.\textsuperscript{10}

Second, recourse to mandated choice, rather than defaults, cannot possibly be a duty, because we often implicitly converge on what is the goal of certain choices, and we prefer to facilitate acting upon those choices by the use of defaults. Systematic recourse to mandated choice by public health authorities is likely to increase the risk that many of us would make choices we would regret because we would not take the time to carefully compare different options (\textit{Intelligent Assignment}).

Third, teaching metacognitive de-biasing methods may help enhance decisional capacities, and public health authorities may have some duty to enhance individuals' deliberative capacities, for instance by exposing them to the norms of reasoning in epidemiology. What I reject is that they have a duty as stringent as the duty of physicians who are required to do everything possible to repair a patient's diminished decisional capacities before overriding her consent for major, risky surgery (Charland 2011). In that context, the physicians' duty is based on the fact that the individuals' bodily integrity is protected from external assault, and that they need good reasons for infringing on bodily integrity without consent.

I do not deny that there are some parallel cases in public health, but they are restricted to a small territory of coercive public health action that impact fundamental liberties. Examples include the use of the police powers of the state to isolate an individual who suffers from a contagious disease. Here, every effort should be made to

\textsuperscript{10} For the relevant literature in the psychology of medical decision-making, see Ubel (2002; 2012; forthcoming) and Ubel and Loewenstein (1997).
repair the person's decisional capacities if they are temporarily impaired (by the disease or a psychological problem) and explain why the individual does not take adequate measures to avoid exposing others, as long as these efforts are compatible with the adequate and timely protection of others (Parmet 2008).

7. Concluding Thoughts

In this chapter, I have argued that paternalism should not play a central role in the ethics of public health nudges. I also have maintained that the overall cumulative effect of non-coordinated nudges might, in some cases, weaken our decisional capacities or undermine their proper exercise. This is a reason to limit recourse to certain nudges. I have shown, contra certain Kantians and Millians, that we have no perfect duty to engage or enhance decisional capacities.

Given the problems with the idea of a perfect duty to either engage or enhance individuals' decisional capacities, I suspect that those among Kantians and Millians who defend duties of this kind are pursuing a perfectionistic agenda. Having a comprehensive view of what the good life is, which includes the decisional capacities and their exercise as a constitutive element, they believe public health authorities should design policies with the aim of promoting their conception of the good life. Ultimately, there is, I believe, some form of callousness in a perfectionistic ethics that puts so much value in increasing the opportunity for conscious deliberation, irrespective of its known effects on the average *homo sapiens'* ability to allocate their limited cognitive resources so that they
get a chance to live a life that is overall shaped according to their own lights. Extremism in defense of non-consequential liberties is no virtue (to paraphrase a famous political speech). This is an issue worth pursuing in future research since the charge of perfectionism is usually addressed to those who defend the use of health nudges and might impose on others their own comprehensive conception of the good life, namely, the healthy life.

More generally, the main goal of this dissertation is to reframe the nudge debate. We need to move beyond the dispute between anti-paternalists who defend freedom of choice and those who believe the use of public health nudges is unproblematic because the state merely intervenes to “level the playing field” between private companies that exploit people's cognitive vulnerabilities and public health practitioners.

In response, I have argued that nudges are not incompatible with freedom of choice and that paternalism is not central to the nudge debate.

I also reject the view of those who defend public health nudges primarily on a leveling-the-playing-field basis. This is because the scope of problems that public health nudges attempt to tackle goes way beyond the food and alcohol consumption areas that they have in mind. Granted, in Cafeteria, Food Traffic Lights, Less Than You Think, and Smaller Plates, public health agents use creative means to decrease the rate of obesity and alcohol abuse, and therefore strive to counterbalance the health effects of certain private companies' interest in maximizing profit.

However, interventions that come under the label “public health nudges” are often not designed to counterbalance the power of the private sector at all and are not concerned with dietary choices. Consider, for instance: Anti-Suicide Bridge Flyers,
BSE Loss-Aversion Campaign, Default Professional Immunization, Hand Washing, HIV-Test Cash Transfer, Intelligent assignment, Negative Surgery Framing, Organ Donations Opt-Out, or White Stripes. The leveling-the-playing-field argument can therefore not be used to short-circuit the discussion about the ethics of nudging.

Moreover, the driving force behind the nudge debate is a dispute about the responsibilities and obligations of the state, the private sector, and civil society. Many defenders and opponents of nudges alike assume that public health authorities need to shoulder a heavy justificatory burden whenever they interfere with liberty interests, but not the private sector and civil society. I have argued against this common assumption about the role of public health authorities. In future research, I intend to evaluate the assumption about the role of nonstate public health agents, namely, the private sector and civil society.
APPENDIX A: LIST OF EXAMPLES

Anesthesia Equipment. To reduce risk of medical errors in anesthesiology, each gas nozzle is designed to be compatible with just one connector so as to minimize the need for deliberation when giving anesthetic drugs to patients (Thaler and Sunstein 2008, 89).

Anti-Suicide Bridge Flyers. On a bridge where frequent suicide attempts occur, the Seoul City government puts flyers that read: “Did you eat anything yet?” (see, URL = http://www.theworld.org/2012/09/bridge-signs-used-in-south-korea-anti-suicide-efforts/).

Asparagus-Lovers. An investigator suggests to research participants that they liked or loved asparagus during childhood the first time they tried it, creating a false memory, and a false belief about the taste of asparagus. Subsequently, participants report increased general liking of asparagus, greater desire to eat it, and willingness to pay more for it (Balz May 14, 2009).

BSE Loss-Aversion Campaign. A health campaign attempts to create a positive attitude toward breast self-examination (BSE) by using a loss-aversion message: “Don’t let the chance of detecting breast cancer go!”

BSE Prevention Campaign. A health campaign attempts to create a positive attitude toward breast self-examination (BSE) by using a persuasive message: “By performing
BSE you are able to detect breast cancer in an earlier and therefore more treatable stage” (Ruiter et al. 2001).

Cafeteria. A cafeteria manager places healthy food at eye-level at the beginning of the food queue. Unhealthy food comes last and is least visible. The customer is then more likely to purchase healthy food (Thaler and Sunstein 2008, 1-3; Just and Wansink 2009).

Clinical Trial Enrollment. All research participants who have completed a clinical trial are, by default, contacted repeatedly to participate in future research, unless they actively signal their unwillingness to participate. This opt-out approach to recruitment to clinical trials is expected to increase the rate of participation in medical research, and thereby promote health. It relies on a foot-in-the-door technique (Junghans et al. 2005).

Deceptive Guards. Guards bring a prisoner to falsely believe his cell door is securely locked (by mimicking door-closing motions), whereas it is in fact unlocked, and he could walk out of the cell any time (Dworkin 1988, 105).

Default DNR. By default, hospitals assume that end-stage demented patients who have cardiac arrest should not be resuscitated (DNR order) (Volandes and Abbo 2007).

Deposit Contract. All primary care physicians of a healthcare system offer their patients the possibility to voluntarily deposit an agreed-upon sum of money with the physicians. The physicians will then return the money to the patients in small installments if the latter
meet certain agreed-upon health objectives to improve their health (e.g., losing weight, exercising, quitting smoking) (Thaler and Sunstein 2008, 232).

**Food Traffic Lights.** Food manufacturers or distributors use easily recognizable “traffic light” intuitive and salient symbols (e.g., green for low, yellow for medium, red for high) to indicate how calorific every item of processed food is.

**Generic Medication.** Medicare beneficiaries are given generic medication by default but are offered the option of getting the brand-name drug (Thaler and Sunstein 2008, 169).

**Gym Bonus.** A company offers bonus salary ($5,000 a year) to employees if and only if they go to the gym on a regular basis.

**Hand Washing.** To increase the rate of hand washing after toilet use in Ghana, a graphic TV campaign induces disgust and fear of contamination if people do not wash their hands (Curtis, Garbrah-Aidoo, and Scott 2007; Vlaev and Darz 2012).

**HIV-Test Cash Transfer.** In Malawi, residents who pick up their HIV-test results receive 10% of their daily wage in cash (Balz March 16, 2011).

**Intelligent Assignment.** Medicare beneficiaries are assigned by default to a healthcare plan based on the historical data about their drug use during three months in the previous year (Thaler and Sunstein 2008, 161-176).
**Less Than You Think.** University campuses in Montana organize an alcohol consumption-reduction campaign accurately stating that 81% of Montana college students have four or fewer alcoholic drinks each week. The campaign underlines the fact that the majority of students binge-drink less often than what most students assume (Thaler and Sunstein 2008, 68).

**Medical Abandonment.** Physician-investigator $A$ threatens to abandon patient-subject $B$ if $B$ refuses to enroll in $A$’s clinical trial.

**Nazi Ballots.** In 1938, the Nazis organized a referendum asking German voters whether they supported (1) the reunification of Austria and the German Reich, (2) and the party of “our leader, Adolf Hitler.” To say “Yes,” voters had to check a large circle right in the middle of the ballot; to say “No,” they had to check a small circle on the right side of the ballot (Balz, August 11, 2010).

**Negative Surgery Framing.** A patient is told that surgery comes with a 30% risk of death. He or she thinks: “If I get surgery, I have a 30% risk of death. Gosh, that’s really bad! I don’t want to die… I guess I should not go ahead with surgery…”

**Organ Donations Opt-Out.** The state makes organ donor status the default option with the possibility of opting out (Halpern, Ubel, and Asch 2007; Thaler and Sunstein 2008; Whyte et al. 2012).
**Paternal Competition.** In an Indian village, health professionals post children’s medical test results in a public place, creating competition among fathers to improve their children’s health (Balz January 27, 2009).

**Pharmaceutical TV Ad.** In a TV ad for medication, the list of side effects is read in a monotonous voice while displaying images of butterflies and happy people.

**Phase 1 Clinical Trial.** A, an investigator, does not provide B, a research participant, with information about the risks of enrolling in a phase 1 randomized clinical trial.

**Positive Surgery Framing.** A patient is told that surgery comes with a 70% chance of survival, he or she thinks: “If I get surgery, I have a 70% chance of survival. This is pretty high! I guess I should go ahead with surgery…”

**Quadruple Bypass Burgers.** The Heart Attack Grill offers free “Quadruple Bypass Burgers” (8,000 calories) to people weighing 350 pounds or more (Balz June 2, 2010).

**Smaller Plates.** Using 10 inches instead of 12 inches plates results in a 22% reduction of calories per serving (Wansink 2006).

**Tobacco Doubt Inoculation.** Tobacco company A designs a marketing campaign inoculating public B with doubt about the scientific validity of research on the health risk of their products by cherry-picking incomplete scientific results (Boush et al. 2009, 72).
**Toothpaste Choice.** A reduces from 30 to 20 the number of toothpaste brands B may choose from.

**University Choice.** A reduces from 30 to 20 the number of universities B can decide to attend.

**White Stripes.** To prompt drivers to take a dangerous S curve more cautiously, a series of white stripes are painted on the road. “When the stripes first appear, they are evenly spaced, but as drivers reach the most dangerous portion of the curve, the stripes get closer together, giving the sensation that driving speed is increasing. One’s natural instinct is to slow down” (Thaler and Sunstein 2008, 37-39).
12th December, 2012

To Whom It May Concern:

I write as joint Editor-in-Chief of the journal Public Health Ethics to confirm that any of the authors are free to reprint their article after seeking permission to do so. The following paper was published in our journal earlier this year.


Yours truly,

Professor Angus Dawson

Head of Medicine, Ethics, Society & History (MESH),
90, Vincent Drive,
School of Health & Population Sciences,
College of Medical and Dental Sciences,
University of Birmingham,
Edgbaston,
Birmingham,
B15 2TT.
UK.

Email: a.j.dawson@bham.ac.uk
Tel: +44 (0) 121 414 2957
REFERENCES


____. Forthcoming. Nudging forward: Comments on “Salvaging the concept of nudge.” *Journal of Medical Ethics*.


____. Forthcoming. A commentary on “Salvaging the concept of nudge.” *Journal of Medical Ethics*. 


Curtis, Valerie, Nana Garbrah-Aidoo, and Ben Scott. 2007. Masters of marketing:


Dawson, Angus, and Marcel Verweij. 2008. The steward of the Millian state. *Public


____. Forthcoming. A commentary on “Salvaging the concept of nudge.” *Journal of Medical Ethics*.


Evans, Jonathan St. B. T. 2009. How many dual-process theories do we need? One, two, or many? In *In Two minds: Dual processes and beyond*, Jonathan St. B.T. Evans and


Norman, Donald. 1988. The psychology of everyday things. London: MIT.


*Planned Parenthood of Southeastern Pennsylvania, 505 U.S. at 851.*


Saghai, Yashar. Forthcominga. Salvaging the concept of nudge. *Journal of Medical Ethics*.

____. Forthcomingb. The concept of nudge and its moral significance: A reply to Ashcroft, Bovens, Dworkin, Welch, and Wertheimer. *Journal of Medical Ethics*.

Saghai, Yashar, and Tom L. Beauchamp. In preparation. The mysterious goal of generalizable knowledge in research.


Ubel, Peter A. 2002. Is information always a good thing? Helping patients make “good” decisions. Medical Care 40(9): v39-v44.
_____ January 26, 2009. Scientocracy: Why psychological science is relevant to policy debates—and what policies might look like if they were better aligned with human nature. Psychology Today. URL= http://www.psychologytoday.com/blog/scientocracy/200901/scientocracy-policy-making-


Welch, Brynn. Forthcoming. A commentary on “Salvaging the concept of nudge.” *Journal of Medical Ethics*.


____. Forthcoming. A commentary on “Salvaging the concept of nudge.” *Journal of Medical Ethics*.

Wheeler, David C., Konrad M. Szymanski, Amanda Black, and David E. Nelson. 2011. Applying strategies from libertarian paternalism to decision making for prostate specific


