

Democratic Paternalism, Not Libertarian
Paternalism

PAUL GOWDER*

INTRODUCTION 747

I. PATERNALISTIC CHOICE INTERVENTIONS 750

 A. *What is Paternalism?* 752

 B. *Weak Interventions* 754

II. LIBERTY: A FALSE START 756

 A. *Easy Cases for Sunstein and Thaler* 758

 B. *What Counts as Unfreedom?* 759

 C. *In Defense of the Choice Architecture Argument* 761

 D. *Will Imposition* 762

 E. *Against Hayek* 764

III. THE IDEA OF DISEQUILIBRATION 766

 A. *Disequilibrium and Choice* 767

 B. *Disequilibrium and Liberty* 773

IV. PATERNALISM, RESPECT, AND EQUALITY 774

 A. *Respect for Autonomy as an Expressive Ideal* 776

 B. *When Are Paternalistic Justifications Insulting?* 778

 C. *Paternalism and Democracy* 782

INTRODUCTION

There is a striking high-level isomorphism between the debates about individual paternalism and what we might call “political” or “collective” paternalism—the various forms of elite intervention on collective agency in democratic states (of which the most salient is constitutional judicial review). Both debates, in their conventional forms, center on an alleged tension between competence and autonomy: incompetent mass decision makers, often oriented to short-term

* Associate Professor of Law, University of Iowa College of Law. I thank participants in the GISME symposium that led to this volume for helpful feedback, my colleague Tom Gallanis, and my research assistants Daniela Talmadge and David Miller. © 2016, Paul Gowder.

rather than long-term concerns, make decisions which they themselves, in more reflective modes, would reject. The debate poses this question: if elites intervene on those decisions, particularly in reflectively reversible ways (nudge-esque setting of default rules; judicial striking down of legislation subject to constitutional amendment), are they doing something objectionable with respect to the decisional autonomy of those on whom they intervene?

These debates are misguided. In both cases, the opposition between autonomy and competence is a mistake, because the kinds of paternalistic interventions proposed, either on the political level or on the personal level, are in general either neutral or indeterminate with respect to the criterion of autonomy-preservation. This is so for essentially the reason Cass Sunstein has identified: agents (individual or collective) do not make their decisions in unstructured, ex nihilo (incentive-free, status quo-free, institution-free) contexts in the first place; there is not a principled way to make general statements about what kinds of changes to those contexts have what kinds of effects on the underlying degree of freedom of choice such agents experience.

Instead, the best worry about paternalism on both levels is that it is potentially inconsistent with the respect due to equal democratic citizens.¹ To intervene on another's decision under the claim that one is better capable of serving the interests of the other than he or she is on his or her own is—under circumstances to be described—to express disrespect for that person as an equal citizen and rational agent as well as to express an inflated conception of one's own capacities and wisdom. To treat someone paternalistically is, without more, to treat him or her like a child, an inferior, not an equal.

I shall argue that paternalistic interventions, to the extent they are necessary in light of actual citizen incompetence, should be carried out in a way to minimize that inequalitarian impact. That can be achieved, *inter alia*, by leveraging, wherever possible, genuine higher-level political consent through democratic processes that respect the ability of citizens to reason collectively, with both procedural and substantive elements designed with an educational purpose, in order to remedy the reasoning defects that justify their enactment, and in a way that tends to disrupt fixed choices and thereby prefer temporally later versions of intervened-upon agents to temporally earlier versions thereof. This is so because a choice intervention is more likely to be consistent with the equal status of the intervened-upon to the extent it recruits their participation in its implementation; to the extent it assists them to develop their rational capacities; and, ultimately, to the extent it reincorporates those developed capacities into its own revision. That is, to the extent such an intervention is democratic. Accordingly, paternalism should be democratic, not libertarian.

1. A number of scholars have recognized this connection between paternalism and disrespect. *See, e.g.*, Nicolas Cornell, *A Third Theory of Paternalism*, 113 MICH. L. REV. 1295 (2015); Seana Valentine Shiffrin, *Paternalism, Unconscionability Doctrine, and Accommodation*, 29 PHIL. & PUB. AFF. 205 (2000).

The context of this argument is a symposium on “nudging”—the notion that some weak paternalistic choice interventions (where “weak” is an idea that will be developed anon) are permissible in view of their minimal or indeterminate impingement on the liberty of the person on whom the interventions are made. The Sunstein/Thaler “nudge” concept has become a leading defense of (at least a category of) paternalistic choice interventions, and examples of nudges are legion.²

Below, I begin by first motivating my merger of the debate about elite interventions onto the decisions of democratic publics and the debate about elite interventions into the decisions of individuals. I will then adapt what I see as the core of Sunstein’s autonomy-based defense of the “nudge” into a general set of objections to liberty-based critiques of a broad (although not comprehensive) set of choice interventions, on both the individual and collective level. I argue that these objections successfully defang critiques of relatively weak paternalistic interventions considered either under a liberal conception of liberty according to which it is violated by interference in agents’ choices, or under one according to which it is violated by substituting the will of some external agent for that of the agent whose liberty is under consideration.³

I next introduce an alternative idea, that of “disequilibration,” an intervention that unsettles an agent’s status quo choices and gives them the opportunity to make new choices in the future. I argue that a disequilibrating intervention prefers the choice of an agent at a later time to an agent at a former time, and may have the potential to express respect for such an agent—be that agent individual or *polis*—considered as a developing reasoner.

I then shift to discussing outright an egalitarian approach to paternalism; one that objects on the ground that paternalism insults the object of intervention. I argue that interventions that represent the intervened-upon as suffering from transitory rational impairment or external constraint are preferable on respect grounds to those that represent them as incompetent.

I close by outlining the basis of a democratic version of the egalitarian objection to paternalism as well as a preliminary description of an egalitarian democratic structure for paternalistic interventions.⁴

2. Richard H. Thaler, Cass R. Sunstein, & John P. Balz, *Choice Architecture*, in ELDAR SHAFIR, *THE BEHAVIORAL FOUNDATIONS OF PUBLIC POLICY* (2012) contains a nice collection of examples of nudges.

3. I understand the liberal idea of interference with choice roughly along the lines given by PHILIP PETTIT, *REPUBLICANISM: A THEORY OF FREEDOM AND GOVERNMENT* 53 (Oxford: Clarendon Press, 1997). A caveat: I am not primarily a participant in the normative conversation on paternalism. Rather, my professional focus lies in neighboring areas such as status egalitarianism, democratic theory, judicial review, and problems of collective action. Accordingly, if this essay makes a contribution to the literature on paternalism, it does so only by adapting insights from those neighbors. Leaving one’s wheelhouse in this way always poses the risk that one says things that are foolish from the standpoint of the scholars on whose territory one has trampled; the reader (particularly the non-expert reader) is warned that this may be a problem here. In short, I do not wholly trust the first half of this essay, and neither should you. Specialists in paternalism should take it as an external provocation.

4. Full explication will have to await some interventions into democratic theory that I am still working on ironing out.

I. PATERNALISTIC CHOICE INTERVENTIONS

First, let us take a look at the territory as a whole. Under discussion are “choice interventions” which I shall shorthand describe as “interventions.” The term stands in for a general category of interactions in which one agent or group of agents (the “intervenor”) intervenes on the choice of another agent or group of agents, either by setting the latter’s choice (reversibly or not), changing the preconditions of that choice (such as by modifying the information likely to be available to or salient to that agent), or by changing the consequences of that choice by making it more or less costly.

Interventions come in a variety of forms (only some of which have been slotted under the heading of “nudge”). The classic heavy-handed interventions, of course, are punishment and direct physical compulsion, but we also have psychological manipulation and informational manipulation (and don’t even ask me to delineate a boundary between the two) of various sorts, covering everything from mandatory disclosures, concealing information, shifting information to make various things more or less salient, controlling the language in which options are presented, and the like. There is also at least one case of what we might call a “structural” intervention, namely the setting of a default rule, and others may exist as well.⁵

While this symposium is nominally about interventions on individual decisions, I find interventions on collective decisions more interesting, and also happen to think the debates about the two domains can helpfully learn from one another, because they are more alike than different. Accordingly, I will sketch some general features of those two debates here as context for the argument that follows. While the individual and collective cases are not identical in all respects, they share enough common features that it may be fruitful to consider them together, at least in a paper like this (i.e., loosely rooted, operating orthogonal to rather than within contemporary lines of debate, consciously testing potentially indefensible ideas). In doing so, I follow in a tradition running from Plato through Jon Elster of treating together the agentic capacity of individuals and of states.⁶

Conventionally understood, paternalistic interventions like the “nudge” are attempts to thread the needle between the twin poles of autonomy and competence. There’s a decision that We (the wise, those who Know Better, or, perhaps, the political community as a whole) think They (the foolish, those who Do Not Know, or individuals within that political community) are entitled to make, as a matter of personal autonomy. But we cannot help but observe that they make bad decisions, either by our lights, or even by their own lights (in more

5. Sunstein and Thaler often talk about default rules with respect to organ donation, an example I shall repeatedly return to here. Cass R. Sunstein & Richard H. Thaler, *Libertarian Paternalism Is Not an Oxymoron*, 70 U. CHI. L. REV. 1159, 1191–93 (2003).

6. See Republic; see JON ELSTER, *ULYSSES UNBOUND: STUDIES IN RATIONALITY, PRECOMMITMENT, AND CONSTRAINTS* (2000).

reflective moments). They drink 64-ounce sodas and then bemoan weight gain, they smoke, and they refuse to vaccinate their kids. So we intervene.

The alleged autonomy-competence trade-off shows up in democratic political choices as well: we support juries because we think there ought to be democratic input in the criminal process, but we worry that juries cannot make decisions as well as judges can. Accordingly, there are longstanding debates about the extent to which elite judges are entitled to intervene in the choices of juries.⁷ Judicial review, likewise, can be interpreted as an elite intervention into the choices of legislators and mass publics. Even the entire body of administrative law is often defended with the notion that elite bureaucrats are more competent than the more democratically legitimate (that is, more reflective of popular autonomy) legislators whose decisions they replace.⁸

Another feature that the political and the individual spheres share is a longstanding conceptual uncertainty about the nature of the agent whose autonomy and competence are in question. The political sphere is composed of collective agents, and it is an open question how to construct those agents. Our normative evaluations of different kinds of political institutions may depend on those constitutive choices. For example, constitutional judicial review seems more palatable from the standpoint of democratic autonomy if we conceive of the *demos*, qua political agent, as having a continuous identity that incorporates the founding generation as well as our own; it seems less palatable if we view the actions of judges not as elements of a contemporary democratic political choice but as external impositions on it.⁹

Similarly, in the individual sphere, it is astonishing how much the classical tradition of conceiving of individuals as warring agents of different levels of rationality (ranging in levels of disreputability from Plato's tripartite soul Freud) sounds like contemporary talk about time-inconsistent agents, akrasia, and the like. And while it (hopefully) is not necessary to take a position on metaphysical questions about identity in order to come to a coherent conclusion about choice interventions, we can at least take note of the heuristic use of the concept of internally conflicting agents, particularly when understanding time inconsistency. There is doubtless a subjectively meaningful sense in which the me who eats an entire pizza is different from the me who feels sick and fat twenty minutes later, and rejects (or even resents) being saddled with the stupid choices of the earlier one. This is important: it allows us to make sense of choice interventions from one temporal slice of an agent to another, as when Ulysses

7. My favorite example is J.C. McWhorter, *Abolish the Jury*, 57 AM. L. REV. 42 (1923). For a more measured discussion and references, see Beth Z. Shaw, *Judging Juries: Evaluating Renewed Proposals for Specialized Juries from a Public Choice Perspective*, 10 UCLA J.L. & TECH 3 (2006).

8. Kathleen Bawn, *Political Control Versus Expertise: Congressional Choices About Administrative Procedures*, 89 AM. POL. SCI. REV. 62 (1995) begins with a good summary of the core ideas.

9. For an argument that the rulings of judges are elements of the choice of the *demos*, see Paul Gowder, *The Countermajoritarian Complaint*, 23 TRANSNAT'L L. & CONTEMP. PROBS. 7 (2014).

tied his later self to the mast.¹⁰ As Elster has pointed out, the same idea applies to constitution-making.¹¹

Like individual paternalistic choice interventions, a number of political choice interventions are justified not on the grounds that the decisions intervened on are bad by some external standard (immoral, harmful to others, etc.), but on the grounds that they are harmful to the interests or values (including the interest in decisional autonomy) of the agent on whose choices intervention is carried out. For example, the Supreme Court's most famous statement on the conditions under which judicial review should be exercised, footnote four of the *Carolene Products* case, singles out "legislation which restricts those political processes which can ordinarily be expected to bring about repeal of undesirable legislation"—that is, democratic self-disablement, like "restrictions upon the right to vote," and "restraints upon the dissemination of information"—as a particular target for scrutiny.¹² We might call those "paternalistic political choice interventions," for it is this property of being justified by the subjective interests or values of the intervened-upon agent that makes an intervention paternalistic. The Court says to the *demos*: "that decision you just made will undermine your own democratic autonomy, here, let us fix that for you."

A. What is Paternalism?

The language above is somewhat imprecise. We ought to reserve the characterization "paternalistic" for the *justification* of some choice intervention rather than for its content, for the same choice intervention can be justified either with respect to the interests or values of the agent subject to intervention or the interests or values of some other agent. Note that I say "justification" and not "motivation," that is, it is irrelevant what some policymaker actually understands the reasons for the choice intervention to be.¹³ I take this to be the important distinction between paternalistic interventions and all other kinds: choice interventions are ordinarily morally suspect; the special thing about paternalism is that some people have offered the interests/values of the one intervened on as a way to answer that suspicion.¹⁴

10. See JON ELSTER, *ULYSSES AND THE SIRENS: STUDIES IN RATIONALITY AND IRRATIONALITY* (Cambridge Univ. Press rev. ed. 1984).

11. See, generally Jon Elster, *Don't Burn Your Bridge Before You Come To It: Some Ambiguities and Complexities of Precommitment*, 81 TEX. L. REV. 1751 (2003) (discussing individual and collective, particularly constitutional, precommitments).

12. *United States v. Carolene Prods. Co.*, 304 U.S. 144, 152 n.4 (1938).

13. *Contra* Daniel M. Hausman & Brynn Welch, *Debate: To Nudge or Not to Nudge*, 18 J. POL. PHIL. 123, 130 (2010), who argue that paternalistic interventions are coercive interventions that are undertaken in order to pursue the interests (or, I insist on adding, values) of the one intervened on. The proper focus ought to be on justification rather than the motivation—paternalistic interventions are interesting not because of the psychology of the intervenor but because there is a live claim that the one intervened on has no reason to complain, because her interests were served.

14. This commits me to ruling out cases like Shiffrin's, Seana Valentine Shiffrin, *Paternalism, Unconscionability Doctrine, and Accommodation*, 29 PHIL. & PUB. AFFAIRS 205, 213 (2000), when one refuses to help, where helping would be supererogatory, in order to make the one asking for help learn

I also would suggest that we ought to focus on the justification rather than the motivation in order to simplify the inquiry.¹⁵ There may be cases in which an intervention is motivated by the interests of the intervened-upon, but cannot conceivably be justified by them, for example when the intervention is carried out incompetently. For analytic purposes it seems sensible to assume such cases away and focus on situations where an intervention would be justified if and only if “this intervention serves your interests/values” counts as a justification.¹⁶

To be clear: I do not mean to argue that a policy is paternalistic for justificatory reasons.¹⁷ Instead, I mean to argue here that it is the justification itself that is paternalistic: a policy, or an intervention, cannot be paternalistic on its own (although I characterize them as such, that should be understood as shorthand for “intervention justified by paternalistic reasons”); rather, a policy can be justifiable for paternalistic reasons, or, potentially, for many other kinds of reasons. Moreover, the same policy might be justifiable by different reasons in different social contexts. I am concerned not with the characterization of a policy as paternalistic but with the sufficiency of paternalistic justifications for policy, and the worry I mean to address is that paternalistic justifications may be

how to carry out the task for himself: such a choice would not require justification, and so is not paternalistic in any interesting sense.

15. See Cornell, *supra* note 1, at 1309–15 (rejecting the motivation-based account of paternalism, for reasons that seem sound to me).

16. In Shiffrin’s case, the motivational account also leads to implausible distinctions such as the notion that it is paternalistic to refuse to give you a cigarette to protect your health, but it is not paternalistic to refuse to give you a cigarette in order that one’s self not contribute to your ill-health, Shiffrin; Cornell; see *id.* at 224. What’s the difference? Incidentally, Shiffrin has an incredibly broad definition of paternalism, one that extends to cases where one intervenes on an agent’s choice because one thinks the agent is paying insufficient attention even to the interests of *third parties*. *Id.* at 216–17. This is obviously incorrect, for it would extend the label “paternalistic” to almost all choice interventions: the law against murder, for example, could be characterized as motivated by the policymaker’s concern about the would-be-killer’s inadequate regard for the interests of others in not being done to death. Nor does Shiffrin’s limitation of paternalism to the intervened-upon’s “sphere of legitimate agency,” *id.* at 218, save the definition, for I take it that those who intervene on the choices of others to protect injury to third parties have it available to them to argue that those choices are not within the “sphere of legitimate agency” of the one intervened upon just because of that third party harm, while those who wanted to make the choices without intervention would obviously argue that those choices are within their legitimate agency. *Id.* at 218. Shiffrin’s argument seems to reduce our characterization of whether an intervention is paternalistic to that question of legitimate allocation of choice-making power.

17. Marneffe argues against Gerald Dworkin’s description of paternalism in justificatory terms, but his argument appears to be directed against characterizing *policies* (or interventions) as paternalistic for justificatory reasons. Peter De Marneffe, *Avoiding Paternalism*, 34 PHIL. & PUB. AFF. 68, 72 (2006). Thus, Marneffe’s argument does not impinge on my own.

Marneffe’s argument (which is difficult to concisely summarize) hinges on the idea that by attaching the characterization of an intervention as paternalistic to features of its justification, one finds oneself saddled with a variety of seemingly paradoxical inferences from the characterization of an intervention to its justifiability. Marneffe’s mistake is to confuse the plausibility of certain candidate justifications for some intervention (arguments that might be offered in its favor) with claims about whether the intervention is all-things-considered justified. By contrast, when I say “a paternalistic justification may be offered for X,” I merely mean “paternalistic reasons are something someone might say in X’s favor with a straight face.”

insufficient to justify some policy in view of the fact that those justifications are inconsistent with certain values (freedom, equality) about which we are rightfully concerned. In short, if an intervention may only be understood as justifiable by imputing to it paternalistic reasons, is that intervention permissible?

To clarify the preceding paragraph, consider a law prohibiting smoking. That law might be justified either on the claim that it will protect the smoker from the ill-effects of smoking, or on the claim that it will protect third parties from secondhand smoke (or both). The first justification is paternalistic, the second is not. This paper is concerned with the question of whether the first claim may be permissibly offered as a justification for the choice interventions the law creates. It is not concerned with the all-things-considered permissibility of those interventions, whether the interventions themselves are correctly characterized as paternalistic, or what motivates the lawmakers to enact the law.

Paternalistic justifications will typically rest on some implicit or explicit factual claim about why the agent in question is unable to serve his or her own interests or values without the help of some intervention. A variety of such claims may feature in paternalistic justifications, including at least the distinct claims that the agent is imperfectly rational and that the agent lacks access to relevant information.¹⁸

B. Weak Interventions

We need a little more conceptual structure to understand how the advocates of nudging can propose a justificatory account of certain types of paternalism. Both individual and collective choice interventions come in varying degrees of strength, where I understand “strength” for present purposes to mean something like an economic idea, that is, a stronger intervention is one that is more costly to override for the agent on whose choice the intervention is carried out.

This is a key idea for “nudge” advocates: nudges are supposed to be justifiable, despite the objections we ordinarily have to intervening on the choices of others, because they are relatively weak interventions, in that it’s easy for agents to override them (and because they’re interventions for agents’ own good); at the limit, for example, the nudge simply sets a default rule (like in the famous organ donation example), with the idea that it is effortless for the agent to change his or her mind if she or he takes the trouble to think about it for a moment. Likewise, in the political interventions realm, again working with the American judicial review example: a judicial decision interpreting an administrative regulation is going to be easier for elected officials (standing in,

18. Conceivably, we might also see justifications that rely on factual presuppositions about external constraints on individual agents such as collective action problems and market effects as paternalistic. For example, a policy requiring union membership might be justified by its potential to serve the interests of individual workers based on the supposition that workers would otherwise be subject to overwhelming pressure from employers not to organize, pressure that can only be overcome with an external intervention-based solution to the free-rider and commitment problems. However, nothing in my argument depends on characterizing the justification of such interventions as paternalistic.

ostensibly, for the *demos*) to change (by administrative action) than the same decision about a statute, which in turn is obviously going to be much easier to change than a decision interpreting the Constitution.¹⁹ Similarly, judicial review is going to be stronger or weaker in different countries, depending on the amount of institutional burden a legislature must take on to reverse the decisions of a constitutional court.²⁰

For analytic purposes, I will say that the strength or weakness of a choice intervention is going to be relative to the relevant *budget* of the agent intervened upon. I use “budget” not in a literal sense, but as a broad way of capturing the idea that the various resources that agents might draw upon to overturn choice interventions made on them are both limited and individually variable. On the individual level, we can conceive of an agent having, for example, a willpower budget or an attention budget, and a variety of demands on that budget.²¹ An individual at the DMV—let us suppose she’s rushing through in the midst of completing a handful of errands on the brief lunch hour she gets in her cognitively demanding job, and has attention-deficit disorder besides—is unlikely to notice that her choice about organ donation, voter registration, or whatever has been set by a default rule: it makes little practical sense to say that she will be able to overturn Cass Sunstein’s intervention into her choices with the same ease as will the diligent retired lawyer who obsessively reads every document she’s asked to sign. Put differently, the cost of a choice can be described as an opportunity cost: an agent can only make so many choices.

Similarly, as of this writing Congress is egregiously polarized: Court decisions that could theoretically be reversible by congressional action in happier circumstances are unlikely to be reversible today, when Congress’s collective action budget (bipartisanship budget?) is vanishingly small—and the combined budget of the joint legislative agent that is Congress and the President may as well be zero.²²

19. For example, the recent Supreme Court case of *Michigan v. EPA* can be seen as a weak choice intervention: it reversed an agency rule subject only to the requirement that the agency go back and re-enact the same rule but in the process pay lip service to the consideration of the costs of compliance, *Michigan v. EPA*, 135 S. Ct. 2699 (2015). Of course, we might also think of that as a very strong intervention, because the EPA can never go back in time and regulate the emissions that occur while it re-enacts the rule—that slice of regulatory time is forever foreclosed. Such an intervention can also be strong if the political environment changes between the original enactment and the intervention such that it is no longer feasible to reverse the Court’s ruling with a new enactment.

20. Cf. Stephen Gardbaum, *The New Commonwealth Model of Constitutionalism*, 49 J. COMP. L. 707 (2001) (discussing the difference between strong and weak judicial review systems).

21. This is consistent with the “ego depletion” model of self-control in psychology. Martin S. Hagger et al., *Ego Depletion and the Strength Model of Self-Control: A meta-analysis*, 135 PSYCHOL. BULL. 495 (2010).

22. To some extent a legislative collective action budget might even track its dollar budget, if, for example, more financial slack means more pork to distribute to veto players. This paragraph was written in 2015, before the unified Republican control of all three branches of government changed the institutional constraints on government action.

The real-world political consequences of this collective action budget crisis and its relationship to choice intervention can be vividly seen in the litigation that led up to the recent Supreme Court decision of *King v. Burwell*. The details of the case are unimportant for present purposes, but the gist is that opponents of the Affordable Care Act brought a challenge to some federal health insurance subsidies that could have brought the whole policy edifice crashing down if it had succeeded—and defenders of the Act successfully claimed that the challenge was based on a trivial drafting ambiguity that did not capture Congress’s policy intent. This is a kind of litigation that could not even exist were Congress not totally unable to fix its own legislation—it was only brought because opponents of the Act knew that they could not get Congress to just repeal it, and it only posed a menace to supporters of the Act because they knew that they could not get Congress to pass the simple corrective amendment that would make it moot.²³

In the face of a paralyzed Congress, the final decision as to what the statute said—and hence whether the entire “Obamacare” policy edifice would survive or fail—was left to the Court. Nominally, statutory interpretation is a weak choice intervention over Congress, but when the intervened-upon agent has no budgetary slack it becomes strong indeed.

II. LIBERTY: A FALSE START

The traditional objections to both collective and individual interventions play out in the language of liberty and autonomy (hence Sunstein’s claim to “libertarian paternalism”). However, each set of objections could be recast—and I shall do so in this paper—in terms of equality, and in particular, the notion of respect, which is deployed both in objecting to paternalism on the individual level and on the collective level, ought to be understood as an egalitarian idea. So understood, we can understand the chief kinds of objections to these sorts of interventions to be about the expression of relative status.²⁴ However, in order to contextualize and motivate this claim, I first address the more conventional discourse about paternalism in terms of liberty.

The key idea of the Sunsteinian defense of the “nudge” is that it respects the autonomy of the nudged while nonetheless putting a thumb on the scale of their choices. It operates by making small alterations to the cognitive or behavioral cost structure of those choices. By making organ donation the “default rule,” it makes it slightly more (cognitively) costly for people to not donate organs, and

23. That amendment would have merely required the insertion of three words, “or federal government,” into two places in 26 U.S.C. § 36B.

24. To foreshadow that argument here, an intervention may constitute an act of disrespect against the agent subject to it, which I understand as a blow to his or her equal status—nudging someone may express the insulting implication that she or he is not capable of making decisions independently. Likewise, the democratic public may receive an insult rooted in the disrespectful implication that it cannot make its decisions directly, either due to incompetence or inadequate understanding of its own political and legal values. Cf. JEREMY WALDRON, *LAW AND DISAGREEMENT* 221–23 (1999).

slightly less costly to do so; by putting scary warnings on the cigarette package it adds a small cognitive cost to smoking.

The suppressed premise in such defenses of nudging is that the costs imposed by nudges are well within the budgetary slack of those who are nudged—so the dispreferred choices, the smoking and not donating organs, are still within the choice sets of the nudged. In a strict sense, the nudged are a little less free to engage in those choices, but they are not wholly unfree.

The second, more explicit, premise in that argument is that the loss of freedom to smoke or be buried with one's organs comes with a gain of freedom (in some potentially suspicious "positive liberty" sense) to refrain from smoking or donating organs. The nudge reduces the cognitive or behavioral costs that are imposed on those choices. At least, this is my reading of Sunstein's "nudging is inevitable" argument: the world already imposes costs on choices; nudges make some choices cheaper while making others more expensive. For that reason, a nudge is often no worse than the status quo ante in terms of the overall externally imposed cost-structure on an agent's choices, and the extent to which the agent enjoys a practical kind of freedom.²⁵

Let us generalize the first of these ideas into the notion of a weak intervention: an intervention is weak when it is cheap for the intervened-upon agent to reverse. How cheap? Sufficiently cheap that the agent, if motivated, need not sacrifice other important interests in order to do so.

But this notion of cheapness is incomplete: "if motivated" buries a key problem, for a premise of the nudge line of argument is that agents pay steep motivational costs (represented by akrasia, general fear and procrastination, the cost of devoting limited attention to a huge number of possible decisions a person can be making at any moment, etc.) for their decisions, so a cheaply reversible intervention in this sense need not be cheap at all, in terms of an agent's actual practical capacity to reverse it.

Consider, again, the organ donation example. Sunstein and Thaler have argued that the difference in default rule between default donation in Europe and default nondonation in the United States accounts for a vast difference in incidence of organ donations.²⁶ If this is true, we might hypothesize that motivational inertia plays a very strong role in people's decisions about organ donation—and thus that it is at least potentially fairly costly for people to attend to the donation choice. Arguably, were an agent to devote attention to the organ donation choice, she would have to sacrifice paying attention to something else that mattered to her. For that reason, I am not certain that shifting the organ

25. For example: the default rule used to be organ nondonation. Now the default rule is organ donation. Either way, there is a default rule, so either way the structure of the world shapes people's choices; how, Sunstein and Thaler ask, can we say that the chooser is less free in the world where someone has consciously set the default rule at donation than in the world where someone has unconsciously set it at nondonation?

26. Cass R. Sunstein & Richard H. Thaler, *Libertarian Paternalism Is Not an Oxymoron*, 70 U. CHI. L. REV. 1159, 1191–92 (2003).

donation default rule is genuinely a weak intervention, though it might be.²⁷

To help us think about that problem, let us more generally take on a distinction between internal costs and external costs; the internal costs to a choice include things like motivational and attentional hurdles for individual agents, as well as coordination problems for collective agents, and are to be distinguished from costs that are imposed by factors external to an agent (legal restrictions on the agent's choice, the deception of others that imposes information costs on the agent, etc.). Of course, this distinction is hardly unproblematic, for assigning a source of choice cost to an agent or to external constraints will often require making a value-laden judgment about the sorts of things that ought to be attributable to the agent in the first place, but let us work with it as a rough heuristic. We can suppose that an intervention is *very weak* with respect to a given agent subject to that intervention when it is cheaply reversible for that agent with respect to both internal and external costs. Such an agent need not be particularly motivated, particularly attentive, particularly wealthy, particularly literate, etc. to make the ultimate choice, even after intervention.

A. *Easy Cases for Sunstein and Thaler*

We may now identify an *easy case* for the nudger: an intervention that is predictably very weak for all or almost all members of the population, and that counteracts the effect of some external bit of choice architecture that all or almost all of the population would predictably find more costly to overcome. If Sunstein's argument can justify any paternalistic intervention, it can justify easy case interventions.

Are easy case interventions real? It seems to me that there is one obvious example: scary warnings on cigarette packages. It is at least plausible to assume that they do not pose a real barrier to people who actually want to smoke (psychological evidence might upset this claim), and that to the extent they do any work, it is in protecting people who might otherwise take up smoking on impulse from being subject to the powerful constraint of nicotine addiction. This is not a perfect example, but should the reader wish to hold an example of an easy case in his or her mind it is probably as good as any.

Since defenders of nudging need only show that there are *some* defensible nudges, not that *all* nudges are defensible, if easy case nudges are defensible (and if there are more than a trivial number of easy case nudges, which I shall assume to be true without defense), the nudger has won the day. In what follows, I shall argue that easy case nudges are defensible from the liberty standpoint, but that only a subset of easy case nudges, in both the political and

27. An alternative explanation for the strong effect of the default rule, consistent with understanding organ donation default rule setting as a weak intervention, would be that people just do not care about the choice. On that hypothesis, it is not that the cost of attending to organ donation choices is particularly high, but that the payoff, in terms of the potential to better satisfy an agent's ordinary preferences, from doing so is so low. That seems to be the implicit assumption of Sunstein and Thaler's argument.

the individual domains, are defensible from the respect/equality standpoint. I will then sketch a set of constraints to delineate the nudges that are acceptable from an equality standpoint.

B. What Counts as Unfreedom?

In defense of the nudge, Sunstein has offered one key argument, which, it seems to me, is sufficient to answer standard objections to paternalism from the standpoint of autonomy—this is the famous argument that “choice architecture” is inevitable. In this section, I will add a few words in defense of Sunstein’s argument from two classic conceptions of what it means to undermine a decision-maker’s autonomy.

To begin, however, I must explain a perhaps idiosyncratic approach that I take to the problem of the normative evaluation of situations where a person’s or community’s autonomy (or freedom, or liberty, or agency—I use these terms interchangeably) is undermined. I understand undermining of autonomy to be morally troubling independent of our ability to identify some agent who is blamable for the undermining. A person who has had bad freedom-luck, for example, because his or her capacity to make decisions has been undermined by some natural circumstance, is suffering from a morally meaningful lack of freedom, and may generate in others some reason to remedy that lack (either as a mandatory duty or as a ground for praise qua supererogation).

This leads fairly directly to the idea that any kind of acting/failing to act distinction (killing and letting die, coercing and letting be coerced) in the domain of the unfreedom of others is morally insignificant. I do not propose an extended defense of this, but from it, it will follow that if, for example, agent A interferes with a choice of agent B, but, in doing so, relieves B from some naturally-imposed interference with choice, we cannot condemn A’s act as a morally culpable infringement on B’s liberty without first asking about the all-things-considered effect of that act on it. This is a kind of consequentialism about freedom. Another way to think about it is that all questions about the moral significance of freedom of choice operate from the standpoint of the one whose choices are being considered rather than of some other who is acting on that person; if two states of affairs are equally constraining from the standpoint of a choice-maker, then those two states of affairs are equally morally significant. I have pumped some intuitions along these lines elsewhere.²⁸

That move is also a rather obvious blunt-instrument way of making Sunstein’s argument go through much easier than it otherwise might. I confess this openly. For the reader who thinks that the claim moves far too fast (especially without a serious defense), I merely say that my chief goal in this essay is not to defend Sunstein’s argument, but to advance my alternative claim, namely, that we ought instead to be evaluating paternalist interventions in the context of

28. Paul Gowder, *Market Unfreedom*, 26 CRITICAL REV. 306, 321–22 (2014).

democratic equality. While a rough *prima facie* defense of Sunstein forms one step en route to that argument, I do not lay any claim to more than that. If you want a more careful defense of Sunstein, read Judith Lichtenberg's contribution to this symposium. I have little to say to the interlocutor who would reject my rejection of the killing/letting die distinction in the domain of freedom.

I will say one thing to head off an obvious objection. This does not commit me to some kind of extreme variety of positive liberty under which an agent may be "forced to be free."

To avoid that objection, let us assume that there is a fact of the matter about whether agent A, having interfered with agent B's choices, actually advanced B's liberty, taken-as-a-whole. (I have nothing to say to those whose metaethical commitments would require denying the fact-of-the-matter claim.) Then, I suggest that we ought to defer to B's judgment about the extent to which A's interference has advanced or undermined B's liberty, for at least two reasons.

First is an epistemic reason. B actually experiences the choices open to her; she will ordinarily (at least at more reflective moments) be a much better judge of the state of her freedom than will A.

Second is a conceptual reason. It is reasonable to think that the extent to which B enjoys freedom will depend on the choices that she can, practically speaking, exercise. It is also reasonable to think that an agent who believes her freedom to be constrained will lack the motivation or information to exercise her full range of choices. It follows that B's subjective experience of constraint (at least reflectively, over a period of time) will constitute an actual constraint on her choices.

With those refinements, the conception of liberty sketched thus far has the resources to reject many objectionable "forcing to be free" cases by generating at least a *pro tanto* objection to any interventions in which the agent on whom intervention is carried out experiences his or her freedom to be constrained as a result. In fact, I think this sort of idea about liberty is generally compatible with standard liberal accounts of negative liberty (which I shall assume to be in play here unless otherwise specified), more or less as described (albeit rejected) by people like Pettit,²⁹ with the caveat that we need to take an expansive account of what counts as an interference with choice—one that does not require a deliberate or identifiable interferer.³⁰

So with that in play, and having effectively begged the question about whether naturally occurring choice architecture is or isn't on a moral par with intentional choice architecture, I turn now to a sketch of a defense of the "libertarian" credentials of nudging.

29. See PETTIT, *supra* note 3.

30. Similarly, in the collective case, these refinements license me to reject obviously unacceptable claims like the notion that it is permissible for one nation to conquer another in order to impose democratic (and hence politically free) institutions on it. The subjective experience of unfreedom, of being subject to direct and extreme coercion, matters (and clearly distinguishes something like democracy-by-conquest from something like judicial review of democratic acts).

C. In Defense of the Choice Architecture Argument

Sunstein's core argument is that no choice exists in a vacuum. An individual choice is structured by a wide array of constraints and incentives, which may be communicative (advertising, information disparity, search costs), institutional (existing regulations), cultural (socialization processes), and economic. We may say the same about political choices, which are constrained by strategic, economic, and structural constraints (for example, the incentives given by a particular aggregation method, the lessons of social choice theory which indicate that no such method can merely represent the will of those who participate, the power of veto players, the limited incentives players have to participate in aggregate decision making [such as voter abstention]).

Then, Sunstein points out, the choice interventions represented by nudges are just one additional constraint or incentive. Moreover (and most importantly), these choice interventions can release other constraints (including, importantly, constraints in the form of internal costs) as they impose new ones. If, for example, the default rule changes from "do not donate organs" to "donate organs," this imposes a cost on one choice (in order to not donate organs, the individual now must fill out a form, and accordingly must be aware of the option, must overcome motivational inertia, etc.), but identically reduces the reciprocal cost of another choice (those costs used to be imposed on the organ donating choice, now they are not). There isn't an obvious way to say which state of affairs, the pre-nudge state or the post-nudge state, is better for the one nudged in terms of his or her freedom of choice. More abstractly, any change in the choice environment of a decider makes some states of affairs more difficult for that decider to reach and some states of affairs easier to reach; it's difficult if not impossible to evaluate these changes and decide which are better, overall, for the autonomy of the decider and which are worse.

If anything, this point is even more compelling when applied to collective choices, which require some mapping rule to connect our descriptions of a collective agent with the institutional instantiation of that agent's decision: those mapping rules tend to be unstable; a change in the choice environment of such an agent can often be re-described as a change in the mapping rule.

For example, suppose that Britain adopts binding judicial review ("strong" judicial review, as in the U.S.). Does that undermine the decisional autonomy of the democratic public? Well, maybe.

There might be a case for that proposition to the extent the majority vote of the House of Commons is assimilated to that decisional autonomy. But we might re-describe the institutional instantiation of the underlying autonomous choice-making capacity of the people as an interaction of the acts of Parliament and of the judiciary. That re-description would be equivalent to pointing out, in Sunstein-esque terms, that Parliamentary process is itself a constraint on that autonomous capacity (the people can only get things passed that can find their way through the electoral process, shot through with social choice and public choice problems, as well as the institutional organization of the legislative

process), and the judicial process can release some of these constraints by giving the people another tool to wield against their elected officials.³¹

I think this argument knocks down many objections to easy case choice interventions. However, it should be clear that Sunstein's argument does not necessarily work with strong choice interventions. If an intervention shifts my choice from A to B, and the cost to shift it back is outside my budget constraint, then option A is genuinely foreclosed from me; if it was *predictably* outside my budget constraint and the budget constraint of people like me, then the general intervention that has been enacted into law genuinely took away a lot of options. It seems reasonable to say that in such a context, the social planner has unequivocally undermined my freedom: I could have A-ed before, now I cannot A. But we are concerned about the easy case for Sunstein: a choice intervention that is predictably weak.

D. Will Imposition

There's an account of freedom that might pose more trouble for this kind of argument. Not neorepublican domination, which is essentially just a theory of equality anyway,³² but an account according to which infringing someone's liberty has something to do with substituting your will for theirs.³³ On that way of thinking, it is not so much the imposing of costs on choices that poses the problem, but the way in which the agency of the one doing the imposing tramples on that of the one imposed.

It seems to me that at a minimum the Sunstein argument forces an important clarification of the will-imposition argument. Is it only a restriction on one's liberty when the other (the will-imposer) actually aims his or her will at one's own, or merely when he or she exercises his or her will, and it happens to become imposed on one's own? Consider architecture: an architect may structure a physical space such that it impinges on the wills of those who find themselves occupying it, but without any such intention on the architect's part. (The architect built that staircase because it looks pretty, now it becomes an impediment to the access of those with mobility impairments to public space.) If such an aiming is not a condition of the proposition that the one imposed upon has had his or her liberty compromised, then essentially the same argument as in the previous section applies, for the external constraints on others which nudges relieve (by imposing their own) are often created by human agents—someone decided that the default rule should be organ nondonation rather than donation, for example, and thereby imposed their will on the decisions of those who would (in some imaginary wills-of-others-free world) have donated their organs. Even if they did so unintentionally.

31. See generally Gowder, *supra* note 9.

32. See generally PAUL GOWDER, *THE RULE OF LAW IN THE REAL WORLD* 58–77 (2016).

33. I have discussed this family of views in Paul Gowder, *Death and Taxes in NFIB v. Sebelius*, 27 PUB. AFF. Q. 243, 249 (2013).

This seems like an appealing position to take. For if we accept that there are many human-created constraints on the wills of ordinary people (and, of course, on collective agents as well), then the intentional version of the will-imposition argument seems to deny us any language in which to criticize them if they are not intended to be constraints. But that cannot be right: surely the architect who cluelessly builds the staircase that makes a previously-accessible public park inaccessible for people in wheelchairs is subject to criticism independent of his or her intentions toward the victims of this practice, and those in wheelchairs suffer just the same regardless of whether the terrible architecture was intended or not; the same goes for choice architecture more generally.³⁴ Even if we resist calling this an imposition on the liberty of those who cannot get into the park, we still have to be able to say that the architect did something bad to them, and that a social policy would be praiseworthy to the extent it corrected that bad thing; that claim alone seems to be enough to make the Sunstein argument go through against the will-imposition theory of liberty, if we suppose that will-imposition can exist in the form of unintentional choice-architecture.

More generally, let us call the non-will-imposing external constraints (i.e. choice architecture) on people impositions upon their “schmiberty.” Nudges might interfere with agents’ liberty, but in doing so, remove impositions on their schmiberty. As schmiberty too is morally important, this constitutes a defense of nudging. There may be legitimate objections to sacrificing liberty in the interest of schmiberty, but surely those objections must stop somewhere: there has to be some large degree of schmiberty for which it is permissible to sacrifice some tiny degree of liberty; schmibertarian paternalism is defensible just to that extent. And since the world is necessarily full of schmiberty-impairing choice architecture, simply in virtue of the fact that humans, in living their lives, create things that get in the way of other humans, there are likely to be at least some permissible nudges.

But let us go a little further. Suppose that we think that there are no realistic permissible trade-offs of liberty for schmiberty, and, furthermore, that an infringement of liberty only takes place when someone intentionally imposes his or her will on someone else. Nonetheless, the world is still full of preexisting choice architecture that infringes on liberty, and there’s still a case for nudges to reverse those. The marketeer who decides that the candy will be put at child’s-eye level in the supermarket checkout aisle, for example, is intentionally imposing his or her will on the customers in much the same way as the regulator would be, *qua* nudge, by insisting that vegetables be put there as well.

So perhaps we should simply modify the easy case by which we’re judging Sunstein’s argument. Instead of “an intervention that is predictably weak for all or almost all members of the population, with respect to internal and external costs, and that counteracts the effect of some external bit of choice architecture

34. Moreover, the most natural language for that criticism is the language of freedom: the architect has restricted the freedom of movement of those impeded by the staircase.

that all or almost all of the population would predictably find more costly to overcome,” let’s just redefine the easy case to be “an intervention that is predictably weak for all or almost all members of the population, both with respect to internal and external costs, and that counteracts the effect of some external bit of choice architecture that all or almost all of the population would predictably find more costly to overcome, and which is deliberately imposed by someone else.” Such cases probably still exist—indeed, depending on what you think about tobacco company marketing, the extent to which the addictiveness of their product factored into the decisions of corporate leaders, etc., the example I gave earlier might still apply.³⁵ Sunstein’s easy case, as modified, is still easy.

E. Against Hayek

Here’s one more objection someone might make to the discussion thus far. Even the easy case amounts to the interference of the state in people’s choices. And interferences by the state might be more troubling than interferences by private individuals, such that we have a distinct reason to complain about easy case state nudges where the preexisting choice architecture that they relieve comes from market interactions. Of course, this argument doesn’t apply to all nudges, even in the individual context (it doesn’t dent the DMV organ donation default rule, for example), and it doesn’t apply to political nudges. Nonetheless, it does apply to many of the core examples of potential easy cases, where the government might use nudges to correct for intentional or unintentional private market manipulation (tobacco, super-size sodas, candy on the eye-level shelf, etc.), so it’s worth addressing directly.

There are at least two reasons we might object to the government nudge over the private choice architecture even in the easy case category. The first is a kind of epistemic idea. The regulator imposes her will on the consumer for the consumer’s interest, while the marketer does so for his or her own; it may be supposed that the regulator is more likely to err (because, as Mill pointed out long ago,³⁶ it is much harder to know someone else’s interest than one’s own). But we can pass by this objection for present purposes, because whatever else it is, it’s not a libertarian objection. It might be a utilitarian or welfarist objection (the government’s paternalistic interventions might destroy the value of the

35. See Richard D. Hurt & Channing R. Robertson, *Prying Open the Door to the Tobacco Industry’s Secrets about Nicotine: The Minnesota Tobacco Trial*, 280 J. AM. MED. ASS’N. 1173 (1998) (discussing tobacco company exploitation of the addictiveness of their product).

On the political/collective level, the most obvious examples of easy case interventions would be interventions against policies that are enacted without democratic authorization under existing laws. For example, judicial review to strike down laws and executive acts in the U.S. which violate federalism or separation of powers principles can be represented as simply reversing an external imposition on the will of the *demos*. This requires the reader accept the notion that the autonomy of the *demos* can only be instantiated through its laws, a claim that I defended in Paul Gowder, *The Rule of Law Against Sovereign Immunity in a Democratic State*, 93 TEX. L. REV. ONLINE 247, 250-252 (2015).

36. JOHN STUART MILL, ON LIBERTY, 139–55 (David Bromwich & George Kateb eds., 2003).

choice architecture to the marketer without creating any for the consumer). It might even feature in the egalitarian objections I will discuss in a few pages. But from the standpoint of liberty, it seems bizarre to suggest that an agent is less free when the government interferes in her choices for her own interest than when a private marketer interferes in her choices for the marketer's interest.

The second kind of objection seems more serious. This is a kind of Hayekian idea: market activity itself is an expression of individual freedom, because, after all, market choices are constrained by the preferences of all the participants, even consumers. If Grocery Store A turns a profit by putting candy on the top shelf, and Grocery Store B down the road loses money by putting broccoli on the top shelf, this is because customers prefer to have candy on the top shelf; the natural (if simplified) market outcome (Grocery Store B either gets with the program and puts candy on the top shelf too, or it ultimately gets out-competed) reflects, rather than impedes the freedom of those who shop in the store. And, it follows, the government intervention to change that is an interference with choices that people value more—and, plausibly, an overall blow against their liberty for that reason.³⁷

The problem with that argument is that it begs the question against the nudger, for it depends on the assumption that consumption choices are exogenous to markets. But this is exactly what the nudger denies.³⁸ Even in an imaginary world where people who shop at Walmart were genuinely free in an external sense to shop at Whole Foods,³⁹ I take it that Sunstein would argue their choices to subject themselves to the grocery store that flings candy in their faces are rooted in part in the fact that they have developed a taste for candy, which in turn is rooted in the fact that they have consumed candy, which, going back around, is rooted in the fact that the grocery store has flung it in their faces.⁴⁰

Put differently, the Hayekian argument offers, against the argument that we ought to interfere in markets because people are induced by them to do things they don't really want (in some counterfactually free sense), nothing but the bare assertion that the induced behavior is, in fact, what people want, and the only evidence for that assertion is that people spend their money in markets that provide it. But the claim that we can read what people want off their market behavior is precisely what is at issue between behavioral economists/psychologists and Hayekians.

37. At the symposium that formed the basis for these papers, Todd Zywicki urged an argument along much these lines.

38. Cass R. Sunstein & Richard H. Thaler, *Libertarian Paternalism Is Not an Oxymoron*, 70 U. CHI. L. REV. 1163, 1163–65 (2003).

39. Julie Beaulac, Elizabeth Kristjansson & Steven Cummins, *A Systematic Review of Food Deserts, 1966–2007*, 6 PREVENTING CHRONIC DISEASES (2009).

40. This is true even if they recognize that they are being manipulated. I can satisfy a desire induced by manipulation *because I have the desire* independent of my awareness of, or even rage about, the manipulation—why cut off the candy-desiring nose to spite my manipulated face?

Of course, the Hayekian might reject the claims of psychologists and behavioral economists that people's market behavior diverges from their considered preferences. Behavioral economics is an active area of research and debate, and there are scholars who have arguments against it. If the candy store luring doesn't actually work, then an important premise of Sunstein's argument is wrong; so much the worse for nudging. But that's an open empirical issue.⁴¹

The Hayekian has one last potential line of defense that actually might be compatible with the empirical claims of behavioral economists: maybe people want to be lured, in some second-order sense. The market rewards grocers who lure their customers into candy with profits, because people want to be lured with candy—because they have preferences over their own preferences (they'd like to be sugar addicts), or third-order preferences over who has control over their preferences (they'd rather be manipulated by grocers than by the government or whoever would do the manipulating in the absence of the grocer). And if the evidence somehow manages to thwart that claim, maybe they have fourth-order preferences over who manipulates them to have preferences over who manipulates them, and so on. The magic of revealed preference methodology might make it perfectly acceptable in economics circles for the Hayekian to think this sort of thing, and thereby become completely immune to contrary evidence—maybe if the market rewards something, we are licensed to infer that people prefer it, full stop, and to not even bother to attend to any divergence between what people do and what they say they want. But if the Hayekian is forced to go that far, she or he has already lost. Whatever the disreputable status of falsificationism in contemporary philosophy of science, we ought to listen to Karl Popper at least long enough to reject methodological commitments that rule out meaningful empirical claims like “people don't always do what they want” a priori.

III. THE IDEA OF DISEQUILIBRATION

One feature of interventions is that they may actually make it easier for agents not just to make the choice to which the intervener sets the agent's decision, but other choices as well. I think this gives us an appealing framework to look at interventions—one that fits nicely with both individual and collective

41. That being said, anyone who would argue that market choice architecture does not successfully manipulate individuals will have a hard time explaining why sophisticated actors nonetheless spend money on trying to do so. It would seem that either the consumers are irrational or the executives who spend billions of dollars on marketing are. And no, to anticipate an objection, advertising and marketing and the like do not merely inform consumers of the benefits of marketeers products, absolutely or relative to competitors. Some aggressively marketed products have no benefits to consumers. For the obvious example: the CDC reports that tobacco companies spent 9.6 billion dollars on advertising in the U.S. in 2012. I defy the Hayekian or other libertarian-inclined economist to argue with a straight face that this money was spent on anything other than manipulation. *See* CENTERS FOR DISEASE CONTROL AND PREVENTION, SMOKING AND TOBACCO USE: TOBACCO INDUSTRY MARKETING, *available at* http://www.cdc.gov/tobacco/data_statistics/fact_sheets/tobacco_industry/marketing/ [<https://perma.cc/GNX4-X2CM>].

agents, is potentially relevant (at least in some cases) to libertarian objections to them, and is directly relevant to egalitarian objections to them.

I will say, loosely to start but a little more carefully in a moment, that a choice intervention is disequilibrating when the intervened-upon agent is more capable of making new choices after the intervention than she was before the intervention. Unsurprisingly, a choice intervention is equilibrating when the intervened-upon agent is less capable of making new choices after the intervention than she was beforehand. The terminology is drawn from the analysis of collective agents, and is meant to suggest the idea that one thing which a choice intervention can do is knock such agents out of a bad game theoretic equilibrium—if, for example, the members of such an agent are playing a coordination game, and have landed in an equilibrium in which they are unable to coordinate on a good outcome, or if the policy choices of a legislature are thwarted by veto players occupying pivotal locations. However, the notion applies to individual agents as well (albeit without the game theoretic scaffolding); we might imagine that such an agent is “stuck,” for example, by cognitive biases, entrenched habits, or even threatened sanctions from third parties, and that some choice interventions may get such an agent “unstuck,” and, for that reason, be disequilibrating.

The heart of this idea comes from two sources. First is Sabel and Simon’s landmark paper on institutional reform litigation and consent decrees, which argues that they are justifiable in part because they can destabilize preexisting political blockages and allow for more stakeholder input.⁴² Second is Heather Gerken’s work on “uncooperative federalism,” “deliberating by deciding,” and related ideas, which collectively suggest that conflicting decisions at different levels of the federal structure can clear space for political action by upsetting status quo situations and thus forcing political actors to make conscious and active choices.⁴³ These ideas can be collectively generalized to the claim that sometimes there is independent value in disrupting the fixed decisions of an agent who might be unable to disrupt those decisions on his, her, or, in the case of collective agents, its own.⁴⁴

A. Disequilibrium and Choice

Enter the concept of disequilibrium with a political science example. One thing political scientists know about complex political systems is that the policy

42. Charles F. Sabel & William H. Simon, *Destabilization Rights: How Public Law Litigation Succeeds*, 117 HARV. L. REV. 1015 (2004).

43. See, e.g., Jessica Bulman-Pozen & Heather K. Gerken, *Uncooperative Federalism*, 118 YALE L.J. 1256 (2009).

44. Sunstein and Thaler have expressed less enthusiasm about what they call “required active choosing.” Sunstein & Thaler, *supra* note 26, at 1173–74. But Sunstein has recognized elsewhere that it can promote learning in those “domains where learning is important.” Cass R. Sunstein, *Deciding by Default*, 162 U. PA. L. REV. 1, 44 (2013). I shall suggest toward the end of this paper that these domains are pervasive, especially in a democratic state.

status quo tends to be sticky, even in the face of social learning or preference change, because such systems tend to contain veto players, who have the power to block policy change and who may have preferences that diverge from those of the majority.⁴⁵ Thus, there can be situations in which a veto player prevents the majority from achieving its preferences.

Consider, as an illustration, the simple spatial model given below. Spatial models, thanks to Downs,⁴⁶ are a simple illustration of policy conflict. The one below represents a one-dimensional policy space (e.g., one which can be described as “more liberal” or “more conservative,” or the like), where each vertical line highlights a point in that policy space. Each player in this space (e.g., each legislator in a legislature, each voter in a state) is understood to have an “ideal point”—a most preferred policy. It is further assumed that the players have single-peaked preferences, i.e., their preferences decline monotonically from their ideal points, and that visual distance represents preference intensity in some useful sense.

In the example below, M is the ideal point of the median voter, and VP is the ideal point of the (only) veto player, such as, for example, a powerful committee chair in Congress. Typically, we interpret the principle of majoritarianism as suggesting that in such a one-dimensional policy space the preferences of the median voter will be implemented.⁴⁷ SQ is the policy as previously enacted by the legislature (e.g., by a prior session with a different group of legislators, and thus judicial decision is made.



A Simple Spatial Model.

45. See generally GEORGE TSEBELIS, *VETO PLAYERS: HOW POLITICAL INSTITUTIONS WORK* (Princeton University Press 2002).

46. ANTHONY DOWNS, *AN ECONOMIC THEORY OF DEMOCRACY* (Harper and Row 1957).

47. See Roger D. Congleton, *The Median Voter Model*, in *THE ENCYCLOPEDIA OF PUB. CHOICE* 707, 707–12 (Springer 2004) for general discussion. The basic claim can easily be demonstrated. Imagine an n -person legislature oriented in unidimensional policy space such that there is a player whose ideal point occupies the median position in that policy space— $(n-1)/2$ players are to i 's left and $(n-1)/2$ players are to i 's right. In a competition under majority rule between any policy to the left of i 's most preferred policy and i 's most preferred policy, i 's most-preferred policy should garner at least $(n-1)/2 + 1$ votes, comprising everyone whose ideal points are to the right of i plus i him/herself, as all of those players prefer i 's most-preferred policy to anything to the left of it. At most, the competitor policy can receive $(n-1)/2$ votes, and i 's ideal point policy should win. The same with the directions reversed goes for any policy position on the right of i 's ideal point. The only way that any policy other than i 's ideal should be enacted is if the legislature does not operate by majority vote, or if something impedes i from proposing his/her ideal policy (e.g., the control of a dominant agenda-setter, an entrenched status quo effect from a prior composition of the legislature, etc.).

Note that in this model, the majority's preferences are frustrated. The median voter would prefer to have the policy set at M, however, it is stuck at SQ, because the veto player prefers SQ to M (SQ is closer to VP than M), and the veto player has the power to prevent the question from coming to the floor, e.g., by killing relevant bills in committee.

However, suppose that a judge strikes down the statute setting the policy at SQ. The policy reverts to the status quo ante (before the SQ statute was enacted), which, let us suppose, was at the point designated as JR above.

The veto player prefers M to JR. Accordingly, after the judicial decision, she will not stand in the way of the median voter, who now has the power to set a new policy at M.

This example is an ideal case of a disequilibrating policy intervention. If we suppose that the will of the legislature is best represented by the preferences of a majority of its members, then the status quo represents a frustration of that will because of the power of the veto player. By destroying the status quo, the court allowed the majority will to operate. Because of the inherent stickiness of legislative choices (they tend to remain in place until they are changed), in order to destroy the status quo, the court had to directly intervene upon the legislature's choice, resetting it from SQ to JR. By contrast, had the court instead moved the policy from JR to SQ, it would have carried out an equilibrating choice intervention rather than a disequilibrating one, in view of the fact that its change made it harder for the majority to set a new policy.

The claim that such a choice intervention is, *ceteris paribus*, a good thing is fairly intuitive. Ordinarily, in democracies, we want the will of the relevant legislative or electoral majorities to be enacted. Disequilibrating choice interventions help that along. To be sure, we often insist on deliberate impediments to those wills. Those impediments typically fall into two categories. Side-constraint impediments block choices or categories of choices that ought not to be made regardless of majority will. For example, we might give the members of a racial or religious minority a veto point in a legislature in order to protect them against majority oppression. By contrast, long-term-planning impediments block future choices in order to extend the temporal effect of past choices. For example, the U.S. constitution contains a supermajority requirement to entrench the choices reflected in it and make those choices last longer than they otherwise might, facilitating long-term planning. The latter case will be seen below to generate a serious objection to the idea of disequilibrium as freedom-facilitating. But at least as a first pass, it looks like we might have reason to support a disequilibrating intervention.

But first, let us consider a similar example to understand what disequilibrium is in the individual case. Some individual choices are harder to leave than others because, for example, they lead to addiction or otherwise distort cognitive choice-making processes. A Sunstein-style "nudge" in school cafeterias might involve placing non-sugary foods in a more visible or accessible location than sugary foods—increasing the cost of choosing sugary foods, relative to

non-sugary foods. Such a choice intervention would be disequilibrating to the extent that sugar consumption, like setting a policy which a veto player prefers to the median voter's ideal point, is "sticky."⁴⁸ This might be true if, as some researchers allege, sugar is addictive.⁴⁹ If sugar is addictive, heavy childhood consumption of sugar will make it more costly to subsequently make alternative consumption choices; by contrast, childhood consumption of non-sugary foods does not make it more costly to later on choose to consume either sugary or non-sugary foods. Blunter responses to addiction share this feature—a judge who sentences a heroin addict to undergo a treatment program, for example, attempts, by taking over that addict's choices in the short-term, to expand his opportunities for choice later on.

Or consider a disequilibrating intervention that simply makes choices salient to an individual. Before the intervention, the agent might have been stuck in what we could call a "cognitive equilibrium," with no reason to expend even the minimal cost necessary to reconsider a previously made (or default) decision, for example, because the agent wholly lacks information about the available options. A choice intervention might provide the agent with some cheap information about an easy way to change the decision, and reasons she or he might want to do so. (For example, employers might be required to provide a simple one-paragraph handout about the option to make individual pre-tax contributions to a retirement account and the advantages of doing so.)

With these examples in mind, it becomes possible to state something like an algorithm for determining if a choice intervention is equilibrating or disequilibrating. This algorithm is imperfect, for the structure of agentic choice is not as finely bounded as the kind of structure I am about to specify. But specifying a formal structure will help us understand the fuzzy neighborhood around that structure.

1. Define the normative agent. The first step in any such analysis must be to come to some position about who, in an ideal world, would make choices within the domain in question. (Difficulties already appear, as noted at the beginning of this essay, especially within collective agents.)
2. Define the time period of the choice intervention. A choice intervention will ordinarily control an agent's choice for some discrete period of time *T*. For example, when a court strikes down a statute, *T* is the period of time between when the court issues its ruling and when the legislature has the next opportunity to meet to make a new choice; in the sugar nudge

48. "Stickiness" here can be seen as a shorthand way of capturing the notion of path dependence, a way that some initial choices restrict the future choice situation of an agent. For a far more careful discussion of path dependence, the canonical work is Scott E. Page, *Path Dependence*, 1 Q. J. POL. SCI. 87–115 (2006).

49. Gearhardt et al., *Can Food Be Addictive? Public Health and Policy Implications*, 106 ADDICTION 1208–12 (2011).

example, T is the period of time between when the child is presented with the healthy option and when she notices the availability of sugary options.

3. Identify the change in the choice set of the normative agent before and after the time period T described in 2 above. Call the first choice set A and the second B.
4. Then we may say that choice set B is strictly larger than choice A if:
 - a. A is a proper subset of B, and all choices have the same costs in both sets, or
 - b. A has the same choices within it as B, but at least one choice is cheaper in B, and no choice is more expensive.⁵⁰

—The notion of a strictly larger choice set is meant to capture a category of cases in which an agent uncontroversially has more choice under one state of affairs than under another.
5. We may also say, similarly, that choice set B is strictly better from the agent's standpoint than choice set A if either:
 - a. Her most-preferred choice within AB is in B but not A, with respect to the preferences she has when she has both sets of options available, and with or without any interventions meant to move her from one set of choices to another (a proviso to rule out endogenous preference changes⁵¹), or
 - b. Her most-preferred choice within AB is in A and B, but is cheaper in A, and while other choices may be more expensive in A, they are not cumulatively more expensive than her most-preferred choice is cheaper—again with the appropriate provisos to rule out endogenous preference change.

50. Under a plausible understanding of the way in which choices are structured, these two versions of strictly larger choice sets are equivalent. For if we understand an agent's domain of choice not in terms of one-off options but in terms of the conjunctively realizable options available to her, then a more costly choice (in whatever currency, monetary or non-monetary, that the agent cares about, and including, for example, time and cognitive effort) will be one which requires an agent to sacrifice more other choices. For example, when the traffic court fines me for speeding, it does not just make my choice to speed more costly, it also reduces the scope of my choice set in an absolute sense, because there are things that I otherwise could have bought with the money that I spent on the fine. See Matthew Kramer, *Liberty and Domination*, in *REPUBLICANISM AND POLITICAL THEORY*, (Cecile Laborde & John Maynor eds., John Wiley & Sons 2009); Ian Carter, *How Are Power and Unfreedom Related*, in *REPUBLICANISM AND POLITICAL THEORY* (Cecile Laborde & John Maynor eds., 2009).

51. Without this condition, we might inappropriately describe as disequilibrating interventions that bring it about that she prefers her choices after intervention not because those choices have improved relative to her stable preferences, but because those preferences have been tampered with. Of course, there are difficulties with this condition in situations where it is difficult to say whether an agent has experienced preference change or option change. Addiction is a prominent example. An agent cured of a heroin addiction might be said to have gained access to an option she preferred all along (not consuming heroin), or we might instead say that her prior powerful preference for consuming heroin has been altered. Here I shall assume that the former description is correct, and that addictions are examples of *akrasia*, where an agent acts contrary to her preferences. If that assumption is false, it changes the evaluation of some cases, but does not, I think, change the plausibility of the overall model.

I will then say that a choice intervention is disequilibrating if B is either strictly larger or strictly better than A, and A is neither strictly larger nor strictly better than B.⁵² Similarly, a choice intervention is equilibrating if A is either strictly larger or strictly better than B, and B is neither strictly larger nor strictly better than B. Many choice interventions will be neither equilibrating nor disequilibrating.

Let us return to the running examples. The judicial review case given above is a disequilibrating intervention because, with respect to the median voter (assuming she represents our normative agent⁵³), B is strictly better than A. The median voter's most-preferred option, under stable preferences, became available only after, and as a result of, the choice intervention. As is appropriate for the ideal case of a disequilibrating intervention, B is strictly larger than A as well.⁵⁴

The sugar nudge case given above looks like a disequilibrating intervention: after a series of such nudges, the child has a higher probability of being spared the restriction of her choice set associated with addiction in the future because of the intervention on her cafeteria choices. However, the underlying expansion of the child's choice set is best expressed in probabilistic terms: it is expected that a child treated with the nudge will ultimately have a strictly larger choice set than a child not treated with the nudge; or, equivalently (and translating from Bayesian to frequentist language), given some M un-nudged children and some N nudged children, a higher proportion of children in N than in M will have larger choice sets in adulthood than they otherwise might. (However, not all of them will: some kids in N will still get addicted to sugar without the help of the school cafeteria.)

Many informational nudges could also be understood as disequilibrating to the extent they reduce information acquisition costs (thus expanding agents' choice sets with respect to information-bound choices), but only so long as they do not increase information-processing costs (e.g., by inducing confusion or indecision).

52. I further assume that such differences are not available without a choice intervention, i.e., that there is no way of improving the agent's choice set in the given respects without intervening on it. If this is untrue, then we might say that there is no sticky status quo for the choice set to restructure.

53. Of course, that assumption is highly debatable. But for the purposes of illustration I shall suppose both that majoritarianism with respect to the present legislative majority is the morally best legislative system to represent the collective agency of the *demos*, and that it is morally best that the preferences of the median voter get enacted under majoritarian systems.

54. Before the intervention, the median voter (and hence, *ex hypothesi*, the *demos*) had the power to set any policy no further from VP than SQ (those were the policies with respect to which the veto player would either prefer or be indifferent toward relative to the status quo). After the intervention, she has the policy to set any policy no further from VP than JR; as JR is much further from VP than SQ, she has a strictly larger set of policy options. (I make the simplifying assumption that the median voter will be able to set any position acceptable to the veto player. In most real situations, at some positions the median voter would cease to be at the median, but the simplification suffices for illustrating the concept.)

B. Disequibration and Liberty

Is a disequilibrating intervention freedom-facilitating? The answer to such a question would seem to depend on our temporal conception of the underlying agent. With respect to an agent taken before the intervention, there are many cases in which a disequilibrating intervention would actually be freedom-undermining, for an agent, individual or collective, may want to be in an equilibrium for the reasons described at great length by Jon Elster.⁵⁵

It seems quite obvious, for example, that an intervenor who swooped in as Ulysses passed the Sirens and untied him would, in doing so, make him less free, at least with respect to that temporal segment of his agency encompassed by his decision to make use of a commitment mechanism to both listen to the Sirens and to not sail to his death. Of course, it is equally obvious that such an intervenor would also make Ulysses more free with respect to the temporal slice of his agency encompassed by his desire to go and visit the Sirens right then and there.

Considering Ulysses, we know pretty well which time-slice of agency we want to privilege, but only because there is an intuitive supposition about what Ulysses understood over an entire life course would choose; we can easily identify his temporary desire to go to the Sirens as an aberration induced by an overwhelming external force, not the kind of choice that Ulysses qua agent with a stable identity would make. Similarly, we might imagine a case where a disequilibrating intervention is clearly freedom-facilitating: if I have been tricked by someone's lies into locking myself into a prison cell, then releasing me from the cell unambiguously makes me freer. But many other kinds of individual decision are more complicated. Suppose I make a bad deal, and then experience buyer's remorse. Would an intervention that releases me from my unwise contract make me more free or less? I, at least, lack a strong intuition on the question.⁵⁶ And matters are even worse when we consider a collective agent like a *polis*. If some intervenor makes it possible for a majority in Congress to overcome the resistance of an intransigent committee chair, then that might free the *polis*, conceived of as an agent represented by the present majority, to make the decision in question, but undermine the freedom of the *polis*, conceived of as a political community that extends over time, to use its existing institutional forms to make decisions that extend into the future. To make any of these kinds of decisions about how to identify the real agent as well as the relevant slice of time over which that agent's choices should count as preferable for purposes of evaluating that agent's enjoyment of freedom we would be obliged to resolve a host of other, probably intractable, normative questions.

55. JON ELSTER, *ULYSSES AND THE SIRENS: STUDIES IN RATIONALITY AND IRRATIONALITY* (Cambridge Univ. Press rev. ed. 1984).

56. Cf. Shiffrin, *supra* note 1, at 210–11 (discussing these kinds of issues with respect to the unconscionability doctrine in contract law).

For these reasons, we ought to hesitate before making any claims about the broad freedom-facilitating or impairing quality of disequilibrating interventions in general. However, we can make one important claim: disequilibrating interventions will tend to shift choice-making power from earlier-in-time agents to later-in-time versions of the same agents. Standing alone, this need not have any moral significance. Later, I shall argue that if we conceive of agency as developmental rather than static, and structure our interventions to promote the development of agentic capacity within those intervened upon, then we will have some reason to prefer interventions that shift power from earlier to later temporal slices of our agents; for that reason, disequilibrating interventions will ultimately be part of an overall package of an egalitarian democratic conception of permissible paternalism. However, some groundwork must be laid for this argument, to which I now turn.

IV. PATERNALISM, RESPECT, AND EQUALITY

A more useful objection to paternalism focuses on the relationship into which it brings the intervenor and the intervened-upon. The name, “paternalism,” highlights this relational aspect to the problem: the objection is to assuming, inappropriately, a relationship akin to fatherhood (or, more gender-neutrally, parenthood), quite literally treating the other as a child—particularly, as an incompetent. And this certainly feels like an important part of the intuitive objection to paternalism.

It is difficult to see how such a supposition can be anything other than objectionably disrespectful. It is the expression of such a disrespectful attitude, and its apparent tension with the equal status of those about whom such an attitude is expressed, that seems to me to be the heart of the objection to paternalism.

We typically think about autonomy in terms of liberty, and while many people talk about interfering with the choices of others as if to do so makes them less free, they also deploy the phrase “respect for autonomy” (which I shall assume means the same as “respect for freedom,” or, more generally, respect for the free decision-making capacity of an agent entitled to make her, his or its own decisions) as an equivalent idea.⁵⁷ I confess to a long-term confusion with that apparent equivalence. As I have argued elsewhere, to make someone unfree or to undermine their autonomy is not the same thing as showing disrespect to their autonomy—it’s perfectly possible to disrespect something while preserving it and to respect something while damaging or even

57. Including in the context of paternalism. *See, e.g.*, Hausman & Welch, *supra* note 13, at 130 (contrasting “rational persuasion,” which “respects both individual liberty and the agent’s control over her own decision-making” with “deception,” which “risks circumventing the individual’s will.” The implication seems to be that respecting liberty and circumventing will are opposite sides of a binary evaluation, and thus that disrespecting liberty is equivalent to making someone not free); Shiffrin *supra* note 1, at 207, 218, 220 (describing the problem with paternalism in terms of respect, although it is not clear whether she means to suggest the problematic equivalence noted here).

destroying it.⁵⁸

Similarly, we ought to understand preserving someone's autonomy and respecting their autonomy to be different things, which may come apart. A key case for separating the two is extremely powerful persuasion. As Nozick once pointed out, there is a model of persuasion (if one believes the philosophers who talk this way) according to which it has a coercive force to it: an ideal rational argument (call this a Nozickian argument) ought to make it impossible for the listener to disagree.⁵⁹ In that sense, offering someone a Nozickian argument might undermine their autonomy, understood as their (notional) capacity to make metaphysically free choices. Nonetheless, we ought not understand a Nozickian argument to be an instance of disrespect for the autonomy of the one convinced, for such an argument operates by recruiting the rational capacities of the listener. Since (assuming we have all been convinced by Kant at least thus far) the thing about human autonomy that makes it worthy of respect (even if not the only thing) is rational capacity, an argument that appropriately calls upon that capacity even as it undermines the freedom of choice of the one in whom it is called upon would positively respect the autonomy of that person. This seems to match everyday intuition: does anyone disagree that offering an audience a rational argument is more respectful (both of their autonomy and of them in general) than, say, lying to them, even if the rational argument is so compelling that it in some sense makes them unfree to reject it?⁶⁰

It also is possible to show disrespect for the autonomy of another while fully preserving his or her choice set. Condescension is a key example of this sort of behavior. To write a long letter full of unsolicited and patronizing advice to another person is to imply that the writer does not think the reader is capable of running his or her own life—to treat the reader like a child or a fool—and, for that reason, to insult the reader's capacity to make rational choices. Nonetheless, it does not interfere with those choices in any more than a minimal sense (so long as the writer is not one's boss, the letter may simply be thrown out).

Some interventions will both undermine or violate the autonomy of the other and show disrespect for it. If the government mandates that a device be installed on all automobiles which prevents them from exceeding the speed limit, that both takes away everyone's choice to drive faster and insults them by suggesting that they are not capable of making their own decisions about how fast to drive, understanding and responding to the incentives given to them by the law

58. See GOWDER, *supra* note 32, at 58–77.

59. ROBERT NOZICK, *PHILOSOPHICAL EXPLANATIONS* 4 (Harvard Univ. Press 1981).

60. To be sure, this vastly oversimplifies a complex discussion, which ultimately may depend on one's positions about free will in philosophy of mind. For example, one might think that the term "free choice" does not mean much except in the context of rational choice. Cf. DANIEL C. DENNETT, *ELBOW ROOM: THE VARIETIES OF FREE WILL WORTH WANTING* (MIT Press, 2d ed. 2015). But these are weeds into which I dare not enter. I assume here that there is a sense of the term autonomy that can be applied to irrational choice-making, though I do not think this is necessary for the core of the argument.

as well as their safety and the safety of others. On some influential accounts, this kind of respect for autonomy is a key virtue of law that allows people to observe the rules and decide to follow them.⁶¹

With that architecture in hand, let us return to Sunstein's easy case. It seems to me that we might be convinced that easy case interventions do not undermine the freedom or autonomy of the agents subject to the interventions for the reasons given above, but might still think that they show disrespect for that autonomy.

A. *Respect for Autonomy as an Expressive Ideal*

I take the notion of respect to be essentially a matter of value expression.⁶² To respect something is to suppose that it ought to be accorded high value and (thereby) to so accord it. To express that respect is to behave in accordance with that value (where "behave" includes linguistic behavior).⁶³ I have argued elsewhere that this is a distinctively egalitarian notion: to fail to show someone due respect is to treat them as an inferior and is objectionable for the same reasons under either description.⁶⁴

In order to understand what it means to respect autonomy, we need some account of the virtues according to which it might be seen as more or less valuable. Respect seems naturally to be dimensional in this sense. To respect the flag, for example, is partly to acknowledge its notability along the dimension "symbols of our country"; to disrespect it is to either deny that that category is a legitimate means of evaluating flags or to deny the flag's importance along that dimension.

As the ongoing discussion suggests, we have at least one dimension that is widely understood to be a virtue of autonomous choice-making capacity: rationality. To deny that someone's choices are made rationally is to cast aspersions on that person's free choice-making capacity in terms of the evaluative scale that applies to it; it is quite directly to express disrespect for that person's freedom. To suppose that one is superior to another with respect to rationality is in addition to aggrandize one's own autonomous choice-making capacity over that of the other.

This is precisely the (most appealing version of the) charge that is made against paternalism. Paternalistic justifications for interventions might express disrespect for the autonomy of the agents whose choices are thus structured, and, more importantly, might do so in ways that other forms of "choice

61. LON L. FULLER, *THE MORALITY OF LAW* 162 (Yale Univ. Press, 1964); JEREMY WALDRON, *LAW AND DISAGREEMENT* 52 (Clarendon Press, 1999).

62. See GOWDER, *supra* note 232, at 199 n.26 and accompanying text.

63. See generally Elizabeth S. Anderson & Richard H. Pildes, *Expressive Theories of Law: A General Restatement*, 148 U. PA. L. REV. 1503, 1510 (2000).

64. Paul Gowder, *The Rule of Law and Equality*, 32 L. & PHIL. 565 (2013); GOWDER, *supra* note 232, at 7–27, 58–77.

architecture” do not.⁶⁵ Moreover, to the extent we ought to interpret a given intervention as supportable by paternalistic justification (which I will explicate further in a couple of paragraphs), we would thereby have reason to see the interventions themselves as objectionably disrespectful. By forcing the intervened to make a particular choice on the grounds that it is good for him, the intervenor acts under the supposition that he is incapable of making that choice for himself. The intervenor claims to know either his preferences or the most effective mode of achieving them better than he does himself, and may even claim that the one intervened-upon lacks basic rationality as a stable character trait.⁶⁶

Pause a moment: I ought to clarify what I mean by “under the supposition” and “claims.” I have argued elsewhere that we can capture a normatively important expressive meaning of a legal act (like a government-imposed choice intervention) with rationalistic interpretation: we ought to suppose that the law expresses the kinds of beliefs about the world and the people in it which would be necessary to rationally justify it, from the standpoint of (appropriately idealized) legislators, people called upon to obey it, and the community at large.⁶⁷

Note how the justification-centered approach to determining the expressive meaning of a law coheres with the classification criteria for paternalism I

65. Cf. Hausman & Welch, *supra* note 13, at 134; Danny Scoccia, *Paternalism and Respect for Autonomy*, 100 *Ethics* 318 (1990). My view here is largely in accord with Nicholas Cornell, *A Third Theory of Paternalism*, 113 *MICH. L. REV.* 1295 (2015).

66. But see De Marneffe, *supra* note 17, at 76–94 (arguing against Shiffrin’s claim that paternalism is insulting because the government (or other intervenor) “substitutes its judgment” for the judgment of the one over whom intervention is carried out). Marneffe’s objection is that this is true of paternalistic as well as non-paternalistic interventions—the law against murder, for example, is a substitution of the government’s judgment for the judgment of the murderer that it would be all-things-considered best to kill the victim. In response, I refer the reader to my earlier argument that the expressive meaning of an intervention, like that of any law, is contingent on the meaning people in a society apply to it. See Paul Gowder, *Equal Law in an Unequal World*, 99 *IOWA L. REV.* 1021, 1047 (2014). In our society, we distinguish between ordinary practical incompetence with respect to one’s own interests or values and what we might call “moral incompetence,” but would more typically call evil. We wouldn’t call a murderer irrational or incompetent for the same reason that Hume would not call the thought process of a person who “prefer[s] the destruction of the whole world to the scratching of my finger” irrational or incompetent. Instead we would call both people wicked. To be sure, wicked is also an insult, but it is not the same kind of insult that *paternalism* can express. Moreover, applying the insult “wicked” to murders is what we *ought* to do: we rightly condemn the evil of wrongdoers. See Joel Feinberg, *The Expressive Function of Punishment*, 49 *THE MONIST* 397 (1965).

67. GOWDER, *supra* note 32, at 28–57. This kind of meaning is morally important, for several reasons. First, it captures important facts about how people can understand their social world. One way we make sense of our surroundings is by trying to figure out what other people think of us; if some beliefs are rationally implicated by the laws of one’s state (even if nobody actually holds them), one has some epistemic reason to adopt those beliefs. If those beliefs are pernicious—if they entail one’s inferiority with respect to the community’s system of value, for example—then such laws come at a cost to the self-respect of the affected individuals. Second, it captures important facts about the relationship between those who wield the power of the state and those over whom it is wielded. If the law makes it look like (is only interpretable as rational if) legislators and bureaucrats think disdainfully of ordinary citizens, that gives us some reason to worry that they actually do think that way about us, and that they are unlikely to treat us as equals ought to be treated more generally. Third, it has moral importance for its own sake: we ordinarily think that it is bad to insult people.

described near the beginning of this essay. Because paternalism is a property of justifications, and the expressive meaning of a law derives from the justifications that members of the legal community should attribute to it, interventions enacted into law that have paternalistic justifications will also typically have paternalistic social meanings.

However, rational interpretation does not stop there. For a paternalistic justification to be believable, we have to attribute (constructively, not actually⁶⁸) some beliefs to the relevant justification-holders (legislators, people called upon to obey, the community at large), particularly, those plausible beliefs that would warrant a conclusion that the law actually leads those subject to choice intervention to better achieve their values or interests than would the absence of such a law. Depending on the content of such a law and the cultural, social, and scientific background in which it is enacted, a multitude of such beliefs could serve that purpose, including, for example, beliefs about the rational capacities of those subject to choice intervention, beliefs about the diffusion of and costs for decision-relevant information, beliefs about the educational context in which people are raised, etc.

In sum, when interpreting the social meaning of a law, we find ourselves committed to an attitude about those over whom the law is to be enforced, understood as an interpretation of the attitudes those who make the law must hold in order to make the law rational. Understanding a law as paternalistically justified narrows the scope of possible attitudes that we might attribute to a rational lawmaker to those consistent with a paternalistic justification. That, in turn, has potential consequences for the respectful or disrespectful character of such laws, an issue to which I now turn.⁶⁹

B. When Are Paternalistic Justifications Insulting?

Consider a quintessentially paternalistic law, like one mandating the use of seatbelts, from this perspective. Why might we imagine a legislator would want to enact such a thing? Given that any such law requires the expenditure of public resources, diverting attention from other matters, such a law is only justifiable if it produces a concrete benefit for those regulated, that is, if they ought to be using seatbelts but are not. Since using a seatbelt is almost costless, and information about the benefits and low costs is widely available and easily understandable, the factual presupposition of the seatbelt law must be that at least some substantial group of those regulated are irrationally choosing not to use a seatbelt.

68. See discussion in Gowder, *supra* note 66, at 1043–44.

69. Note that the argument in this section does not, on its terms, apply to paternalistic interventions carried out through means other than lawmaking. While the rationalistic structure of interpretation on which it depends may extend to the interpretation of non-legal acts as well, in this paper I am only concerned with the legal ones.

This kind of approach changes the tone of the debate over paternalism even of the strong kind. From the liberty standpoint, a law mandating trivial and obviously correct behavior like seatbelt-wearing seems least objectionable just because the actions mandated are so slight and unintrusive that it is hard to imagine any serious claim of liberty in favor of not doing them. Is the freedom to die from easily avoidable stupidity really worth defending? From the respect standpoint, however, such laws are even more objectionable in virtue of their triviality and obviousness: a paternalistic law that requires someone to do something non-obviously self-preserving at least potentially reflects a realistic estimate of human rational limitations; one that requires them to do something as basic as wearing a seatbelt seems to rest on the supposition that the regulated are total nincompoops—and the regulators, mysteriously, are not.⁷⁰ To be sure, we might want to enact those laws anyway—in the balance of our all-things-considered moral judgments—it may be better to insult people than to let some percentage of them die. But we cannot ignore the moral cost of the insult.

This argument squares nicely with some critiques of judicial review. In particular, Jeremy Waldron has argued against judicial review on the grounds that it expresses a kind of distrust for the people.⁷¹ And to the (somewhat doubtful) extent that what legislatures enact genuinely represents the popular will, this critique seems forceful. It can be understood in a conventional anti-majority-tyranny conversation as the claim that powerful officials ought to trust majorities to not tyrannize minorities (and majorities, acting in their role as constitution-enactors who create or do not create judicial review institutions, ought to trust themselves or their later instantiations). But it can also be understood in a more anti-paternalistic sense, as the claim that officials ought to trust majorities not to act foolishly even with respect to their own interests, e.g., by giving up their own civil liberties in times of crisis. Writing judicial review into a constitution, on this interpretation, smacks of the same kind of elitism that paternalism on the individual level does: the sense that the constitution-drafters (not always all that democratically legitimate in the first place) know better than the ordinary people how the majority ought to rule itself, and are going to recruit judges from among their elite circle to make sure they do so.

But not all paternalistic interventions, either on the political or the individual level, require justifications that culpably suppose some kind of incompetence in

70. Moreover, to the extent officials actually hold such attitudes—as opposed to merely enacting policies that can only be justified by constructively attributing such attitudes to them—this may encourage and reinforce a kind of self-serving bias in officials: as Anderson (elaborating on Dewey) has cogently explained, the powerful systematically tend to underrate the concerns of the less-powerful. ELIZABETH ANDERSON, *SOCIAL MOVEMENTS, EXPERIMENTS IN LIVING, AND MORAL PROGRESS: CASE STUDIES FROM BRITAIN'S ABOLITION OF SLAVERY* 7–8 (2014) <http://kuscholarworks.ku.edu/handle/1808/14787> (Lindley Lecture at Univ. of Kansas). In such a cognitive environment, we ought to be particularly wary of allowing official action to be premised on the supposition that officials are smarter or more rational than ordinary citizens, for that may make it more likely in the future that officials will be able to rationalize disregarding the interests of the public in general or minorities within it.

71. Waldron, *supra* note 24, at 221–23.

the object of intervention. Consider, for example, two interventions. The first is *information-forcing*: the social planner puts some kind of disclosure in front of the agent subject to intervention. Interpreting this policy, it would be fairly natural to suppose that the law is justifiable because the one intervened upon lacks information; such a supposition does not take a position on his or her rationality or competence.⁷²

By contrast, consider a *rationality-substituting* intervention. The social planner rewords some document, not to communicate new information, but simply to rephrase the information in a slightly different way, which just happens to track some behavioral economics result. For example, the social planner rewords the handout given to new employees about the retirement options: instead of explaining voluntary tax-advantaged contributions as potential benefits, the new handout explains them as avoiding long-term losses. The most natural interpretation of such a policy is that it was aimed to take advantage of the empirical regularity of loss-aversion,⁷³ and hence that it is premised on the rational failings of those who would read the handouts. On its face, this policy would be most compatible with disrespectful, rather than respectful, paternalistic justifications.

More broadly, a choice intervention which is to be acceptable from the standpoint of respect for autonomy will be one that does not insult the one whose choice is subject to intervention—does not suggest that she lacks rational capacity. Intuitively, there are at least three categories of messages that paternalistic interventions can send consistent with the rational capacity of the regulated.

The first non-insulting message is that whatever impedes the regulated from acting rationally is external to her agency, rather than internal to it. A paternalistic intervention that can be rationally interpreted as directed at the elimination of some outside force that interferes with the intervened-upon's agency might be enacted without communicating any view about her personal qualities.

The second is that the force that impedes the intervened-upon from acting rationally is internal but transient. In such circumstances, a paternalistic intervention may rest on the assumption that the intervened-upon is rational in the long-term, but may be subject to short-term cognitive glitches (as are we all). Those short-term glitches, however, may have long-term consequences thanks to the persistent effects of some bad decisions. One important feature to note about interventions that attribute merely short-term irrationality to the intervened-upon agent is that they are consistent with a general stance of equality between the intervened-upon and the powerful officials doing the intervening. An official

72. *Contra* Hausman & Welch, *supra* note 13, at 127–30, I see information forcing as a serious choice intervention, if only because humans can process a limited amount of information and lack full control over their attention. If I shove a bunch of information in your face, I might force you to attend to it, rather than to some other information, and thereby interfere with other information-dependent choices you might want to make.

73. Daniel Kahneman, Jask L. Knetsch & Richard H. Thaler, *Anomalies: The Endowment Effect, Loss Aversion, and Status Quo Bias*, 5 J. ECON. PERSPECTIVES 193–206 (1991).

who enacts such an intervention need not be understood to suppose that she is smarter or more rational than those subject to the laws, she merely needs to suppose (and ideally that supposition should be supported by evidence) that before carrying out the intervention she has had the opportunity for reflection and external feedback (is operating via “system 2,” in the lingo of behavioral economists), while those on whom the intervention is to be applied are not.⁷⁴

The third is that the intervened-upon agent’s rational capacity is not impaired at all; instead, she is simply subject to the effects of information costs. She is making fully rational decisions with the information she has, however, the intervening official has better information.

An example of an intervention in the first category in the individual case is the regulation of an addictive substance, which may be represented as an external force that disrupts the baseline rationality of the individual agent. Admittedly, such a regulation would still suppose that the subjects of intervention are insufficiently rational to take account of the probable long-term preference changes that would result from their short-term consumption of the addictive substance. But this is, it seems to me, an implication that comes not from the legislation but from the mere classification of a substance as addictive in the first place. An addictive substance just is one that short-circuits rational choice about its use, and we generally accept that this is a limitation of human rationality. There is nothing new or unexpected about the cognitive defects that would be required to paternalistically justify legislative regulation of addictive substances; it is just a recognized part of the human condition, and hence in our society such a regulation should carry less insulting meaning about those who are subject to addiction.⁷⁵ In the political case, similar kinds of justification may apply to judicial review of legislative acts where those acts would operate to disable the public from resisting them. Statutes instituting political censorship, for example, can be understood as an external restraint on the long-term political agency of the people.⁷⁶

The second category is exemplified in the individual case by, for example, regulations forbidding operating dangerous equipment while intoxicated—even if the danger is only to oneself. A law that forbade drunk skiing, for example—which, unlike drunk driving, is probably much more likely to kill oneself than to kill anyone else—would be consistent with respect for autonomy for this

74. Daniel Kahneman, *Maps of Bounded Rationality*, 93 AM. ECON. REV. 1449, 1450–52 (2003).

75. Put differently: someone subject to addiction is not understood to fall short of commonly-accepted norms of rationality, for we typically recognize that throwing off something like heroin is an extraordinary demand, more than we may usually make of human capacity. Those who make and enforce the law are just as subject to that limitation as are those who are regulated by it. I think—though I admit that empirical evidence would be helpful here, and I do not have it—that we also recognize that the future costs of consuming addictive substances are difficult to estimate, that addiction is *insidious*.

76. Cf. *United States v. Carolene Prods. Co.*, 304 U.S. 144 152 n.4 (1938).

reason.⁷⁷ In the political case, this nicely matches the defense of the Senate offered in Federalist No. 63: That the Senate is a body of legislators that are ultimately still accountable to the people (via, at the time of the Constitution's enactment, state legislatures who are themselves elected), but who, because of their longer terms and indirect vectors of influence, are more likely to respond to the long-term desires of the people rather than their short-term desires. Observe that Madison is careful to attribute long-term rationality to the people; the Senate is "a defense to the people against their own *temporary* errors and delusions" because at "particular moments in public affairs . . . the people, stimulated by some irregular passion, or some illicit advantage, or misled by the artful misrepresentations of interested men, may call for measures which they themselves will afterwards be the most ready to lament and condemn."⁷⁸ The Senate's intervention (in the form of a power to impede the enactment of the popular will) is meant only to "suspend the blow mediated by the people against themselves, until reason, justice, and truth can regain their authority over the public mind." Regardless of whether this was sincere, or merely politically-expedient flattery to a despised public, it is suggestive that such clarifications were thought necessary.

Examples of the third category include, in the individual case, things like warnings on dangerous products. They also include efforts to reduce the costs of information that has already been provided, such as by making government forms simpler and easier to understand. In the collective case, examples include the judicial resolution of legal questions (including constitutional questions) that are genuinely disputed among the public, particularly where there is reason to believe that there is popular support for following the constitutional course, whatever that happens to be. To the extent we can safely suppose that judges have better legal knowledge than the public, this evaluation takes advantage of the idea that there is no insult in being told that an expert knows more than one in the area of his or her expertise.⁷⁹

C. *Paternalism and Democracy*

Now let us consider the final step in the dialectic. The defender of unrestricted paternalistic intervention might point out "but it is true that this person/*polis* is incompetent (and not necessarily transiently so)." Juries ignore the evidence. Voters routinely deliver themselves into the teeth of thugs, maniacs, fools and demagogues (and about half the population is convinced this has happened in every election, so it is not clear how the voters themselves can even assert their own collective stable competence). People smoke, drive around

77. A drunk person is only temporarily impaired, and we are all impaired while drunk; such a law communicates no insulting message about the people who happen to find themselves a few drinks in on any given day.

78. THE FEDERALIST No. 63 (James Madison) (emphasis added).

79. That is, the people want a ruling according to law. The judges know the law. Their intervention merely communicates to the people useful information about what the law is.

without their seatbelts, and decline to vaccinate their children against deadly diseases on the basis of nonsense about autism. Or, back to Sunstein, they find themselves manipulated by others, their choices structured by their cognitive environments, and their rational capacities undermined by external stimuli. Is this a problem for those who would ground objections to paternalism in a demand for respectful treatment?

It seems to me that the notion of democracy must supply an upper-bound on the extent to which such claims are admissible. For both political and individual paternalistic interventions ultimately, in a democracy, rest on some conception of public will, assent, or conviction. Thus, the defense of judicial review has to come down to the notion that the constitutional principles the courts enforce are those of the people—the very same people who are supposed to be incompetent. The law-making organ donating a default or taking decisions on scientific questions away from citizen juries too must be attributable, however indirectly, to the reasoned assent of those very same citizens. (Not even Sunstein, I imagine, would support guardianship in the Platonic sense, no matter how many organs would get put in the banks thereby.)

The truly contemptuous democratic paternalist (D.P.) must be caught up in what Habermas calls a “performative contradiction”—the mode of justification for the political act in which D.P. is engaged is inconsistent with the mode of justification for the procedure through which D.P. acts.⁸⁰

To resolve the performative contradiction, two clear options present themselves. The first is simply to deny that an appeal to democratic sovereignty requires a respectful attitude toward public competence. It may be that autonomy requires we give people a say in the decisions that affect them, even if we have no faith that this say will be exercised rationally or intelligently.

That is a pretty dreadful view, however. Any actual democratic policymaker who took it would probably trouble us greatly—it represents a vicious kind of grudging accession to the demands of others—“those idiots get a vote, so I guess I have to cater to them,” and with it a disposition to do things like deceive the public in the interests of expediency. It also represents an unrealistic self-regard on behalf of those who are allegedly more competent—the policymaker, after all, is human too, and if the *demos* cannot be counted upon to competently determine which paternalistic policies are worth adopting and which are not, then D.P. seems to be stuck with the claim that she or he is just smarter, better trained, etc. than the rest of us (and that it would not be worth the public expense to teach the rest of us to keep up).⁸¹

80. JÜRGEN HABERMAS, *MORAL CONSCIOUSNESS AND COMMUNICATIVE ACTION* 90 (1990).

81. Moreover, in actual democracies we typically think the people had a say in the selection of those who rule them, and the supposition that those who so rule are more competent than the people might require presuppositions about, at a minimum, the competence of the people to carry out that selection. What kind of people can be trusted to elect people to rule with tanks and guns but cannot be trusted to wear their seatbelts?

The second solution to the performative contradiction is to rest on a kind of second-order competence: democratic publics, structured properly, are competent in a way that individual people need not be. This could just be Condorcet again, or deliberative democracy, more complicated arguments about collective rationality, information, etc.⁸² Taking such a route, a democrat can suppose that the democratic *polis* can manifest a kind of competence that individual citizens lack.

The second-order competence thesis might help in a defense of individual paternalism, but it does little good in defending political paternalistic intervention over a *demos* itself. Moreover, it is worryingly controversial. It might be that in modern democracies (unlike in Athens), citizens are sufficiently incompetent, and political institutions are sufficiently dysfunctional, that we cannot depend on second-order competence to substitute for citizens' rationality. In particular, the administrative agencies that are likely to be pressed into service to administer nudges may have few rationality advantages over individual citizens because of the combined impact of bureaucratic pathologies that make it difficult for top-level policy initiatives to be accurately implemented,⁸³ as well as the pernicious effects of agency capture.⁸⁴

However, I think democracies have a different answer to the respect problem, one that draws an insight from the one site of unambiguously permissible paternalism in our moral world, that of actual parents. In the parent-child relationship, the incompetence of the child is represented as a temporary condition, and, moreover, one paternalism is aimed at remedying.

In democratic theory, we already have a strain of thought that focuses on an educative function of democracy as well as a democratic function of education, represented classically by, for example, John Stuart Mill.⁸⁵ At the same time, we have strong cases for the possibility that education itself can be conducted consistent with democratic respect.⁸⁶

I wish to offer a sketch of what we might call a neo-developmental theory of democracy and education, one influenced (very loosely) by the likes of Mill, Dewey, Freire, and Hegel, which focuses on the development of a moralized conception of reason in and through a political community. I cannot defend that theory here (this is probably a book-length project, although I am making a stab at it in a couple of initial papers elsewhere), and I have not come to a full

82. See JOSIAH OBER, *DEMOCRACY AND KNOWLEDGE: INNOVATION AND LEARNING IN CLASSICAL ATHENS* (2008); Elizabeth Anderson, *The Epistemology of Democracy*, 3 *EPISTEME* 8 (2006).

83. See e.g., MICHAEL LIPSKY, *STREET-LEVEL BUREAUCRACY: DILEMMAS OF THE INDIVIDUAL IN PUBLIC SERVICE* (30th anniversary expanded ed., 2010).

84. Ernesto D. Dal Bó, *Regulatory Capture: A Review*, 22 *OXFORD REV. ECON. POL'Y* 203 (2006).

85. JOHN STUART MILL, *CONSIDERATIONS ON REPRESENTATIVE GOVERNMENT* 45–69, 55–80 (Prometheus Books 1991) (1861); C. B. MACPHERSON, *THE LIFE AND TIMES OF LIBERAL DEMOCRACY* 44–76 (Oxford Univ. Press 44–69 (2011–2012) (1977); WENDY DONNER & RICHARD FUMERTON, *MILL* 33–55, 76–89 (2009).

86. See e.g., PAULO FREIRE, *PEDAGOGY OF THE OPPRESSED* (Myra Bergman Ramos trans., Continuum 1982) (1968).

understanding of what I (ought to) mean by the idea yet, but its key ideas as they now appear to me are as follows:

1. Neither individuals nor polities start off their existence as fully autonomous agents with causal power over their environments. To the contrary, individuals begin as children with no understanding of their worlds; polities are created by external agents (founders, lawgivers, stationary bandits) for their own reasons and with institutions designed to entrench their interests.
2. Individuals become more autonomous over time by coming to understand their environments and their places in them, as well as the moral significance of their acts; likewise, polities become autonomous over time by coming to understand the latent moral value of their institutions and learning to operate them in order to better fulfill that normative promise.⁸⁷
3. A political state ought to promote the development of autonomous individuals, through education, as well as autonomous polities, through promoting the autonomy of the individuals within those polities as well as the capacity for reasoned collective action of their polities.

These ideas give us a way to distinguish between acceptable and unacceptable paternalistic justification. Acceptable justifications, in such a democracy, will express respect for the rational capacities of citizens and polities (even if they are presently imperfect) by advancing those capacities.⁸⁸ Such policies will arise from a recognition that citizens' and polities' nascent capacities for rational choice and control over their environments are to be valued, and promoted because of their value.

Accordingly, a paternalistic intervention might be consistent with the respect due equal democratic citizens if it is carried out, *inter alia*:

1. With the maximum public participation in the decisional process leading to its enactment (consistent with the demands of competence)—where this participation includes formal decision making power, discussion and debate, and *ex post* revisability. This strongly counts against deceptive tactics

87. This is a more-or-less Hegelian idea rooted in functionalist ideas about states that I have expressed in a different context. See GOWDER, *supra* note 32, at 143–67. Loosely speaking, the argument I hope to make in the present context is that political institutions that last extended periods of time are likely to do so in part because they have something to offer to their people; democratic institutions can offer both moral value as well as the capacity for reshaping those institutions; and the development of that capacity and the recognition of that value are likely to be further associated with stable institutions over time—this suggests that democracy is likely to have the capacity to promote both moral and practical education in its own institutions.

88. Cf. Cornell, *supra* note 1, at 1334 (arguing that government paternalism in democracies may be less problematic because citizens are imposing the paternalistic interventions on themselves). This seems to me to be not quite right, in view of the possibility of persistent anti-paternalistic minorities (like libertarians) as well as insufficiently accountable bureaucracies and the like: many of us may not experience ourselves as the authors of democratically enacted paternalism.

or against the enactment of paternalistic interventions that are invisible to, and thus insulated from the scrutiny of, individual citizens,⁸⁹

2. Where the processes through which the intervention is both enacted and carried out are directed at educating citizens, by giving them the procedural tools to make more rational decisions, relative to the ends at which they aim,
3. In a substantively educative fashion, aimed at revealing to the objects of intervention the behavioral implications of interests they claim or values to which they are committed,⁹⁰ and
4. In a substantively disequilibrating fashion, one that prefers later instantiations of both collective as well as individual agents over earlier instantiations, and thus shows due regard for the developing reasoning capacity and continued reflection of such agents.

89. This criterion also should exclude the conscious manipulation of cognitive biases. For example, the example given a few pages back of framing tax-advantaged retirement contributions as loss-avoidance rather than benefit-seeking seems inappropriately manipulative in view of the fact that it consciously takes advantage of a known human cognitive defect. I think this is an example of a case in which democratic paternalism is more restrictive than libertarian paternalism: it's hard to see how Sunstein and Thaler can object to such a nudge.

Democratic paternalistic choice interventions that respect the autonomy of those over whom interventions are undertaken will instead be aimed at *reducing* the effect of known human cognitive limitations. Coercive addiction treatment, as noted above, seems like a quintessential example of such an intervention.

Contrast the retirement contribution nudge to a nudge like choosing to put the vegetables, and not the candy, on the top shelf. This kind of nudge does not seem to me to carry the same kind of insulting implication: it isn't necessarily *irrational* to choose the product on the top shelf (it might simply represent sensible conservation of search costs). Accordingly, a rational lawmaker can choose to require vegetables to be on the top shelf without requiring us to attribute to that lawmaker the pernicious belief that people are too dumb to look down below.

Generally, in situations where an intervener has a choice about which option is cheapest to an agent (another example is the organ donation/default rule case), the intervener may attempt to influence agents' behaviors by setting the cheap option without being chargeable with disrespectful manipulation—so long as the relative cheapness of options is not creditable *only* to the incompetence of the agents subject to intervention.

Here, Sunstein and Thaler might object that some choice in framing the retirement contributions must also be made. However, when we know that humans are subject to a specific cognitive limitation, at least sometimes we may choose to be neutral with respect to that limitation. In the case of the retirement contributions, for example, we might choose wording that describes retirement income *both* as loss-avoidance and as benefit-seeking.

That being said, there may be some situations in which we genuinely cannot avoid manipulating cognitive biases one way or another. We must accept such manipulation in those cases, and at least exercise it for ends and with means that are maximally consistent with the equality of those who are unavoidably manipulated.

90. I think we can find practical examples of political choice interventions that play a role in this kind of educative process in U.S. Supreme Court decisions like *Brown v. Board of Education* (and possibly *Obergefell*) that aim to articulate and advance a fuller conception of the central values to which the political institutions of the United States already express a (constructive) popular commitment. That is, judicial review can play a *discursive* as well as a *coercive* role, making arguments to people about what their values entail and offering them an opportunity to intellectually engage with those arguments.

Where these criteria are satisfied, even an otherwise disrespectful paternalistic intervention might be permissible because of its satisfaction of higher-level criteria of democratic respect. For example, consider the example of the seatbelt law discussed above, which I said appears worryingly disrespectful in view of the fact that it most naturally rests on the supposition that the general public is full of fools. It might nonetheless be permissible as a form of democratic paternalism where it is not imposed by technocratic administrators, but instead is enacted by representative legislators after a full public debate. Such a law may also be more respectful with a sunset provision, to express the idea that the people are capable of learning, after some forced experience, that wearing a seatbelt is beneficial to them.

Doing so seems to express a developmental rather than a dictatorial conception of paternalistic intervention: we conduct paternalistic interventions in order to make better deciders, individually or collectively, in the long-run. It also seems to reflect a kind of popular maturity in the form of mass metacognition: the people of a democracy, both individually and collectively, recognize their own ongoing process of cognitive and agentic development, as well as the fact that their decisions are and ought to be tentative and subject to further reflection, so they make appropriate choices in order to structure their own decisions accordingly. This seems like a pretty good kind of paternalism to me.