

Libertarian Quasi-Paternalism

Jacob Goldin*

ABSTRACT

In many settings, people's choices vary based on seemingly arbitrary features of the choice environment. Policies that manipulate these features to improve decision-makers' well-being are paternalistic – unless one takes the unrealistic view that these features are relevant from the perspective of the choosers' preferences. In such settings, I propose that policy design can be less paternalistic if the only people assumed to be making mistakes are those whose choices are observed to vary based on the arbitrary feature of the choice environment. I discuss several characteristics of such “quasi-paternalistic” policy design and conclude by applying the principle of quasi-paternalism to the policy choice of nudges versus mandates.

I. INTRODUCTION

Suppose you are in charge of designing a retirement savings plan for some group of people (e.g., employees at your firm, residents in your state) and your goal is to make the people in that group as well-off as possible. The specific choice you face is whether individuals will be automatically enrolled in the savings plan with an option to opt out, or whether the enrollment default will be non-participation and individuals will have an option to opt in. Although the other features of the savings plan itself will be the same in either case, you know from past experience that more people will choose to participate when enrollment is opt-out than when it is opt-in.¹ Which design should you choose, and how can you make this decision in a non-paternalistic way?

Answering questions like this requires, first and foremost, a definition of paternalism, but traditional notions of paternalism are not up for the task. As the concept is usually understood, a policy counts as paternalistic if it is justified on the belief that it will make a person better off than if the person had been left to choose between the available options for him or herself. But in many settings, such as the savings plan example described above, recognizing and accounting for individual mistakes is the only reasonable way to go about

*Assistant Professor of Law, Stanford Law School. For helpful discussions, I am grateful to Daniel Deacon, Ben Eidelson, Daniel Hemel, Christine Jolls, Cass Sunstein, and to Thomas Lambert and the participants in the *Evaluating Nudge: A Decade of Libertarian Paternalism* symposium at the University of Missouri School of Law. I am also grateful to the editors of the *Missouri Law Review* for excellent editorial suggestions.

1. For evidence of this type of effect, see, for example, Brigitte C. Madrian & Dennis F. Shea, *The Power of Suggestion: Inertia in 401(k) Participation and Savings Behavior*, 116 Q.J. ECON. 1149 (2001).

designing policy. That is, when individuals' choices vary according to seemingly-arbitrary features of the choice environment (such as which option is the default), policymakers can conclude one of two things: either (1) some people are making systematic mistakes (i.e., selecting the default option when they would have been better off under the non-default), or (2) which option people prefer truly depends on the feature of the choice environment that varies (i.e., the enrollment default). Often, it is impossible to maintain the latter claim with a straight face – does anyone seriously believe that the amount most people should be saving for retirement turns on which plan they happen to have been defaulted into? In such cases, one reaches the unappealing conclusion that any sensible approach to policymaking is necessarily paternalistic.²

In this short article, I propose a distinction between two different forms of paternalism. As a starting point, we can divide decision-makers into two groups. “Inconsistent choosers” are those whose choices vary based on some arbitrary factor under the government’s control (like defaults), and “consistent choosers” are those whose choices are not sensitive to such factors. The basic idea in this article is that in some sense it is more paternalistic to make policy assuming that the consistent choosers are making a mistake than it is to make policy assuming that the inconsistent choosers are doing so. Specifically, a policy is *consistent choice paternalistic* if it is based on the belief that it improves the decisions of the consistent decision-makers – i.e., if it assumes that individuals' voluntary choices are mistaken even when those choices do not vary based on any arbitrary factor. In contrast, a policy is only *quasi-paternalistic* if it is designed to improve the choices of the inconsistent decision-makers but takes the choices of the consistent decision-makers to be correct.³

Returning to the savings plan example, quasi-paternalistic policymaking assumes that those employees who decide to enroll in the savings plan under both the opt-out and opt-in enrollment defaults are actually better off participating than not participating. And similarly, a quasi-paternalist policy assumes that those who choose not to participate under both enrollment defaults are in fact better off not participating than participating. In contrast, quasi-paternalism does not dictate anything about what policymakers should take to be the preferences over participation of those employees whose enrollment decisions vary based on whether or not participation is the default. Because of this, quasi-paternalism can accommodate the belief that the inconsistent choosers are erring under one enrollment default or the other.

2. See Richard H. Thaler & Cass R. Sunstein, *Libertarian Paternalism Is Not an Oxymoron*, 70 U. CHI. L. REV. 1159, 1171 (2003), for the related claim that paternalism is inevitable. As described below, my claim is slightly different: paternalism is not inevitable – it is just that the alternative is unappealing because it requires imposing unrealistic assumptions about the content of decision-makers' preferences.

3. As described below, by “correct” I mean that the chosen option furthers the decision-maker’s goals and values at least as much as would any other option that the decision-maker could select.

A key feature of the proposed distinction between quasi-paternalism and consistent choice paternalism is that it provides a way for traditional anti-paternalist concerns to inform decisions about choice architecture. Critics of paternalism argue that paternalist policies are flawed because they interfere with individuals' autonomy and because people know their own preferences better than the government does.⁴ As I argue below, quasi-paternalist nudges are as consistent with these principles as possible while simultaneously acknowledging the subset of decisions that appear likely to be mistakes.

Although it is possible to nudge in a way that is quasi-paternalistic, some nudges go beyond quasi-paternalism. Nudges are consistent choice paternalistic when the arguments used to support them are based on presumptions about the preferences of the consistent decision-makers that disregard the actual, observed choices of the people in that group. In fact, many of the arguments that have been advanced to support nudges fall into this camp. For example, in the savings context advocates for nudging have argued that decision-makers are present-biased and myopic,⁵ which could apply equally to both the consistent and inconsistent decision-makers.

In contrast to the possibility that nudges can be quasi-paternalistic, I argue that mandates justified on behavioral grounds are necessarily consistent choice paternalistic. This is not for the usual reason – i.e., that mandates restrict people's choices; paternalism, as I use the term, is about the assumptions policy-makers make about people's choices and preferences when designing policy, not the substance of the policy itself. Rather, mandates are not quasi-paternalistic because a mandate can be justified over a nudge on behavioral grounds only in settings where policymakers believe that the consistent choosers are making systematic mistakes. To illustrate, consider a mandate to participate in the retirement savings plan discussed above. Intuitively, under a nudge (i.e., setting participation to be the default), all the inconsistent choosers select the default option. Under a mandate, all the consistent and also all of the inconsistent decision-makers do so. Since the inconsistent choosers end up making the same decision under both the mandate and the nudge, a necessary condition for the mandate to achieve higher social welfare than the nudge is that it makes the *consistent* choosers better off (relative to the nudge). But for this to be true, it must be the case that the consistent choosers who choose against enrolling in the plan would actually be better off if they were forced to enroll. This assumption – that the preferences revealed by the consistent choosers' choices are mistaken – is the very condition that disqualifies a policy from being quasi-paternalistic.

There is a large and ongoing debate among legal scholars about whether paternalistic policies are an appropriate response to the findings of behavioral

4. See Alan E. Fuchs, *Autonomy, Slavery, and Mill's Critique of Paternalism*, 4 ETHICAL THEORY & MORAL PRAC. 231, 236 (2001).

5. See, e.g., RICHARD H. THALER & CASS R. SUNSTEIN, NUDGE: IMPROVING DECISIONS ABOUT HEALTH, WEALTH, AND HAPPINESS 103–17 (2008). In fairness, these authors include other rationales for nudging as well.

economics. My goal in this short article is not to persuade anti-paternalists that all forms of behaviorally motivated policy interventions are justifiable but simply that such concerns are differently implicated according to whether a policy is consistent choice paternalistic or quasi-paternalistic. And similarly, for those who are already convinced that the presence of individual mistakes makes deviations from traditional non-paternalism desirable, my goal is to argue that the concerns informing traditional non-paternalism can usefully enter the debate about which types of behaviorally informed policies to adopt.

The remainder of this article proceeds as follows. Part I briefly describes a number of background concepts and reviews the reason nudging is seen to be paternalistic. Part II then sets out the notion of quasi-paternalism and describes several of its features. Part III illustrates the concept by explaining how quasi-paternalism provides a ground for preferring nudges to other policy tools that restrict choice, such as mandates.

II. BACKGROUND AND TERMINOLOGY

Consider a group of individuals who each face an identical choice. Each must select one option from a set of available alternatives. Each has some set of priorities for what he or she cares about, reflecting the person's goals and values (whatever they may be). A person's "preferences" over the available options refer to the relative ranking of the options according to how consistent each is with the furtherance of the person's priorities.⁶ For example, if Bob prefers option *X* to option *Y*, this means that *X* is more consistent with Bob's priorities than is *Y*.

Following a number of recent papers,⁷ I define nudges in terms of a policymaker's choice of frame. A "frame" is a feature of the decision-making environment that affects what people choose (for at least some choosers) but that is irrelevant from the point of view of the choosers' preferences.⁸ For example,

6. The ranking of options embedded in one's preferences is based on whatever information is available to the decision-maker at the time of the decision. Note that this understanding of preferences differs from the "revealed preferences" approach traditionally employed in welfare economics – if preferences were defined to be identical to one's choices, there would be no space for individuals to make mistakes, and so any value to policies like nudges would be ruled out.

7. See, e.g., Jacob Goldin, *Which Way to Nudge? Uncovering Preferences in the Behavioral Age*, 125 YALE L.J. 227 (2015); Jean-Michel Benkert & Nick Netzer, *Informational Requirements of Nudging* (U. Zurich, Dept. Econ., Working Paper No. 190, 2016), https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2597631. Although less concise than other ways of defining a nudge, a virtue of this approach is that it makes precise which types of policy tools are being discussed and facilitates formal analysis of their characteristics. For a discussion of related issues, see Brian Galle, *What's in a Nudge*, ADMIN. L. REV. (forthcoming).

8. Yuval Salant & Ariel Rubinstein, (*A,f*): *Choices with Frames*, 75 REV. ECON. STUD. 1287, 1287 (2008) (defining a frame to be "irrelevant in the rational assessment of the alternatives"). This definition of a frame is slightly stricter than the one in

a frame might denote the order in which the available options are presented to the chooser, which option is labeled as the default, or any manner of variations in which the choice is presented (as long as none of those variations affect which menu option the chooser actually prefers).

In many settings, policymakers have flexibility over which frame to implement. For example, if the government runs a website through which people purchase health insurance,⁹ the government must decide the order in which to list the available plans and which features of the plans to make salient. Similarly, the government may choose to regulate how private companies present various choices to consumers, such as whether retailers present credit card fees as surcharges or discounts relative to paying with cash.¹⁰

A “nudge” refers to a decision to adopt one frame over another in a particular choice setting in order to promote the well-being of the people facing the choice.¹¹ For example, suppose that government employees are more likely to enroll in an employer-sponsored retirement savings plan when they are automatically enrolled into it upon beginning their employment, as compared to when participating requires them to affirmatively opt in. Assuming employees’ preferences over participating in the plan are unrelated to whether or not participation is the default, the nature of enrollment into the plan (i.e., whether it is opt-in or opt-out) represents a frame. If the government determines that employees would be better off participating in the plan than not participating, designing the plan to have opt-out enrollment would constitute a nudge.

An important feature of this definition of a nudge is the requirement that it be intended to promote the well-being of the people facing a particular choice. This excludes situations in which one adopts a particular frame to affect people’s decisions in ways that are primarily intended to benefit third parties. For example, one widely discussed policy is changing the default for organ donor status so that it is an opt-out system instead of an opt-in system.¹² The rationale for this policy is not primarily to make the potential donors themselves better off (such as by correcting their mistakes about whether to donate), but rather to improve the well-being of the potential organ recipients. Or similarly, changes in the choice architecture that promote green energy use¹³ or a

Goldin, *supra* note 7, at 238, in that it does not allow as a frame some variation that imposes transaction costs on selecting one particular option or another.

9. See, e.g., HEALTHCARE.GOV, <https://www.healthcare.gov> (last visited Aug. 15, 2017).

10. See *Expressions Hair Design v. Schneiderman*, 137 S. Ct. 1144, 1150–51 (2017). For one perspective on this issue, see Todd J. Zywicki, Geoffrey A. Manne & Kristian Stout, *Behavioral Economics Goes to Court: The Fundamental Flaws in the Behavioral Law & Economics Arguments Against No-Surcharge Laws*, 82 MO. L. REV. 769 (2017).

11. THALER & SUNSTEIN, *supra* note 5, at 5–6.

12. See, e.g., Eric J. Johnson & Daniel Goldstein, *Do Defaults Save Lives?*, 302 SCIENCE 1338 (2003).

13. See Cass R. Sunstein & Lucia A. Reisch, *Automatically Green: Behavioral Economics and Environmental Protection*, 38 HARV. ENVTL. L.R. 127 (2014).

reduction in use of plastic bags are primarily motivated by their positive environmental effects,¹⁴ rather than the goal of making the affected decision-makers better satisfy their own preferences.¹⁵

Turning from the definition of a nudge to the definition of paternalism, one is immediately struck by the large number of definitions of paternalism used in the literature.¹⁶ However, many of these definitions share common themes. As Sunstein puts it in his contribution to this volume, “the unifying theme of paternalistic approaches is that a private or public institution does not believe that people’s choices will promote their welfare, and it is taking steps to influence or alter people’s choices for their own good.”¹⁷ In other words, a policy is paternalistic if it is implemented on the belief that it causes decision-makers to make choices that better reflect their own preferences than the choices they would otherwise make. In contrast, a non-paternalistic policy assumes people’s voluntary choices reflect their preferences.

When paternalism is understood this way, nudges are paternalistic by definition.¹⁸ This conclusion follows from the key property of a frame – that it affects choice but does not affect preferences. Consider a person whose choice is affected by the frame – i.e., who selects one option (call it *X*) under one frame and a different option (call it *Y*) under a second frame. Because the person’s preference for *X* versus *Y* is, by assumption, unaffected by the change in frame, we know that the person’s most-preferred option must be the same under both frames as well. Yet because the person chooses a different option under the two different frames, we know that the person must not be selecting his or her most-preferred option under either the first frame or under the second frame. And because our definition of a nudge requires that the choice of frame be governed by the policymaker’s belief over which frame will best promote the choosers’ well-being, it satisfies the definition of paternalism. Put differently,

14. Tatiana A. Homonoff, *Can Small Incentives Have Large Effects? The Impact of Taxes Versus Bonuses on Disposable Bag Use* (Princeton U., Indus. Relations Section Working Paper No. 575, 2013).

15. An even more extreme example concerns changes in a frame that are intended to benefit third parties by causing the directly affected decision-makers to make mistakes. For examples along these lines, see Jacob Goldin, *Optimal Tax Salience*, 131 J. PUB. ECON. 115 (2015), and Benjamin R. Handel, *Adverse Selection and Inertia in Health Insurance Markets: When Nudging Hurts*, 103 AM. ECON. REV. 2643 (2013).

16. See JONATHAN QUONG, *LIBERALISM WITHOUT PERFECTION* 73–107 (2010), for an overview and critical discussion of a number of these definitions.

17. Cass R. Sunstein, *Forcing People to Choose Is Paternalistic*, 82 MO. L. REV. 643 (2017).

18. For a nice exposition of this argument, see Robert Sugden, *Why Incoherent Preferences Do Not Justify Paternalism*, 19 CONST. POL. ECON. 226, 231–33 (2008) (“If a person’s choices over given options are affected by variables that cannot credibly be claimed to have any corresponding influence on well-being, and if we subscribe to any specific consequentialist theory of well-being, we can infer that some of those choices are not promoting that person’s well-being.”) As the title of this article suggests, Sugden ultimately disagrees with my conclusion here. The source of disagreement stems from Sugden’s embracing a non-welfarist approach to policy-making.

a nudge is paternalistic because it involves selecting a frame based on the idea that at least some decision-makers would make mistakes if a different frame were to be implemented instead.

Given that nudges are inherently paternalistic, someone hoping to follow a non-paternalist approach to policymaking might reasonably ask what alternatives are possible. One possibility is for the government to choose between alternative frames without trying to maximize the choosers' well-being, or to leave the choice of frame to private parties. As others have pointed out, this approach suffers from the obvious disadvantage that it may result in frames being adopted that do not maximize decision-makers' preferences.¹⁹ For example, private firms may choose to adopt frames that maximize profits rather than total social welfare. Note that these approaches avoid paternalism not by challenging whether decision-makers are making mistakes but by challenging how (or if) the government should respond to those mistakes in the first place.

A seemingly more attractive possibility for avoiding paternalism is to simply avoid basing policy on the belief that decision-makers are making mistakes. That is, instead of assuming that the feature of the decision-making environment that affects choices represents a frame, policymakers might instead assume that the feature is relevant to the decision-makers' preferences. Thus, if some people choose differently in one choice setting than in another, one would simply assume that for those people, the variation between choice settings is something that is relevant to their preferences. Thus if more employees decide to participate in an employer savings plan when there is opt-out enrollment than when there is opt-in enrollment, one would assume those employees are better off participating than not participating in the opt-out plan and would be better off not participating than participating under the opt-in plan. For example, following the default option might allow people to avoid suffering the hassle costs of filling out the paperwork and spending the mental effort determining how much to save.

This approach, while avoiding paternalism, often does so at the cost of making unrealistic assumptions about the content of people's preferences. That is, the only way to avoid concluding that an observed choice reversal represents a frame is by assuming that the thing that has changed is relevant to the choosers' preferences. But in many cases this assumption is simply not realistic. To return to the retirement example, although it is theoretically possible that the difference in behavior in the opt-in versus opt-out settings is not a mistake, it seems exceedingly unlikely that one's priorities about how much to save for retirement (a major financial decision!) would turn on whether or not one has to spend a few minutes filling out a form.²⁰

19. See, e.g., THALER & SUNSTEIN, *supra* note 5, at 85.

20. See B. Douglas Bernheim, Andrey Fradkin & Igor Popov, *The Welfare Economics of Default Options in 401(k) Plans*, 105 AM. ECON. REV. 2798, 2799 (2015) ("In a conventional model [of 401(k) saving], unrealistically large opt-out costs (averaging thousands of dollars) are required to rationalize default effects."). In other cases it can be difficult for an outsider to determine if something is a frame or not, since doing so requires making a judgment about the decision-makers' priorities. For example,

Thus, the standard view of paternalism creates a dilemma for anti-paternalists. One is either forced to adopt an unrealistic view about the content of people's preferences (essentially assuming away the possibility of irrational behavior), or one must acknowledge that certain frames will better promote welfare than others. The dilemma is that it is difficult to accept the latter (and only reasonable) perspective, while also arguing that the government should not take these differences between frames into account when choosing between them. The remainder of this article sketches a new understanding of paternalism better suited for analyzing policy-making in such settings.

III. QUASI-PATERNALISM

This Part proposes a distinction between two different forms of paternalism and argues that the distinction is useful for analyzing behaviorally informed policy-making. The key feature of this approach is that it differentiates between the people whose choices are affected by the frame (the inconsistent choosers) and those whose choices are unaffected by the frame (the consistent choosers).²¹ In particular, a policy is *consistent choice paternalistic* if it is based on the view that the voluntary choices of the consistent choosers do not reflect their preferences. In contrast, a policy is *quasi-paternalistic* if it is based on the view that it improves the choices of the inconsistent choosers, while accepting that the choices of the consistent choosers reflect that group's preferences.

Both consistent choice paternalism and quasi-paternalism meet the definition of traditional paternalism in that they are both premised on the notion that (at least some) people's voluntary choices will not promote their own welfare. The difference between them is that quasi-paternalism limits this presumption to the people whose choices vary according to the frame. As I argue below, treating the voluntary choices of these inconsistent choosers as mistakes is, in certain senses, less paternalistic than treating the voluntary choices of the consistent choosers as mistakes.²²

someone may choose not to accept a very small risk of death from flying between two locations but be willing to accept a larger risk of death from driving between the same two locations. Whether or not such behavior represents a framing effect depends on whether the decision-maker's priorities differentiate between a risk of death from driving and an equal-size risk of death from flying.

21. See Goldin, *supra* note 7, at 242.

22. In an influential article, Bernheim and Rangel propose a measure of individual welfare in which option *A* is preferred to option *B* if the decision-maker would select *A* over *B* under each choice setting in which both are available (i.e., under every frame). See B. Douglas Bernheim & Antonio Rangel, *Beyond Revealed Preference: Choice-Theoretic Foundations for Behavioral Welfare Economics*, 124 Q.J. ECON. 51, 53 (2009). This approach is related to the one described here in that both attach additional significance to the choices of decision-makers unaffected by a frame, but unlike Bernheim-Rangel, the framework described here does not assume away the possibility that even inconsistent decision-makers have normatively relevant preferences over the

Having set out the definition of quasi-paternalism, the remainder of this Part considers four observations about the concept.

Observation 1: Quasi-paternalism can accommodate observed choice reversals.

The main advantage of relaxing the traditional definition of paternalism is to accommodate the evidence that some people's choices vary systematically in response to seemingly arbitrary factors (i.e., the evidence that frames matter). As described above, traditional conceptions of non-paternalism require making the dubious assumption that whenever people choose differently between various choice settings, those differences are the result of preference-relevant features of the choice – i.e., that frames do not exist.²³ In contrast, a quasi-paternalistic policy treats inconsistent choosers as if they are making mistakes in one of the frames. Thus, although quasi-paternalism requires making a paternalistic judgment with respect to some of the decision-makers, the payoff to doing so is the ability to avoid making unrealistic assumptions about the content of decision-makers' preferences.

Observation 2: Not all nudges are quasi-paternalistic.

Whether or not a nudge is paternalistic depends on the reasoning that motivates the policy rather than the content of the policy itself. Thus, the same nudge may be either quasi-paternalistic or consistent choice paternalistic depending on the reasons for which it is adopted. Specifically, if a particular nudge is justified based on the idea that it does not affect the well-being of the consistent choosers but that it causes the inconsistent choosers to make a choice more in keeping with their preferences, it will be quasi-paternalistic. In contrast, if the argument for that nudge was instead premised on the belief that the consistent choosers are making a mistake, the nudge would be consistent choice paternalistic rather than quasi-paternalistic.

To illustrate, return to the example of a government employer deciding whether enrollment into its retirement plan will be opt-in or opt-out. Suppose we know that half of the employees will choose whether to participate based on the default, and suppose further that we know that ninety percent of the

available options. Making that assumption would eliminate any scope for nudges to raise social welfare by improving the consistency of decision-makers' choices with their preferences.

23. To clarify this point, recall that a frame is defined as a feature of a choice setting that is preference-irrelevant and that affects the choices of at least some decision-makers. If a frame were to exist, it would mean that people's choices varied systematically based on some arbitrary difference between choice settings, which would in turn imply that people were choosing incorrectly in at least one of the choice settings. However, the view that people's voluntary choices are incorrect is incompatible with traditional non-paternalism.

employees who are consistent will choose not to participate. As in many similar settings, assume that we have no direct evidence about the preferences of the inconsistent choosers. Suppose that the employer decides to adopt participation as the enrollment default out of a concern that employees tend to under-save for retirement and are present biased in making savings decisions.²⁴ Note that although ninety percent of the consistent choosers make choices suggesting they prefer not participating, this reasoning suggests that many of those employees would in fact be better off if they were to participate. Hence, adopting a default of participation on this type of rationale would not be quasi-paternalistic because it would imply that even the consistent choosers were making a mistake.

In contrast, determining that a default rule of participation is welfare-enhancing would be quasi-paternalistic if the reasoning took the voluntary choices of the consistent choosers to be correct.²⁵ What this avoids is approaching the nudge as a way to correct some unrelated bias (like under-saving) on the part of the decision-maker. There are a number of approaches for setting nudges that take as their starting point the premise that consistent choosers are making the correct decision.²⁶

Observation 3: Quasi-paternalism is “inevitable,” but consistent choice paternalism is not.

One of the strongest arguments in favor of nudging is that choices about how choice architecture should be designed are inevitable.²⁷ Although decisions about choice architecture are inevitable, it is not strictly true that nudging itself is inevitable, since, as discussed above, policymakers could treat decisions about choice architecture as being about preference-*relevant* features of the decision (rather than frames). Thus, at least under the definition of a nudge used here, one can avoid nudging by making policy in a way that assumes even the inconsistent choosers are not making a mistake. But, to the extent one wishes to avoid making the often unrealistic assumption that these choice reversals reflect preference-relevant differences between the choice settings, then at least quasi-paternalism is inevitable.

However, the fact that quasi-paternalism is inevitable does not imply that other forms of paternalism are also inevitable. That is, the fact that some decision-makers are clearly making a mistake does not imply that other decision-

24. See, e.g., Shlomo Benartzi & Richard H. Thaler, *Heuristics and Biases in Retirement Savings Behavior*, 21 J. ECON. PERSP. 81 (2007).

25. In another work, a coauthor and I have referred to this starting point as “the consistency principle.” See Jacob Goldin & Daniel H. Reck, *Revealed Preference Analysis with Framing Effects* 1 (working paper 2017), https://papers.ssrn.com/sol3/Papers.cfm?abstract_id=2417709.

26. See *id.*; see also Goldin, *supra* note 7.

27. See THALER & SUNSTEIN, *supra* note 5, at 83–87.

makers – those whose choices are insensitive to the seemingly arbitrary factors – are also making mistakes.

Even if stronger forms of paternalism are not inevitable, they may still be appropriate if there is convincing evidence that the consistent choosers are making errors in the choice setting under consideration. Consider a retirement savings plan with automatic enrollment, for example, and suppose that the vast majority of the employees who opt out of participation come to regret their decision not to participate.²⁸ These employees are consistent with respect to the participation default²⁹ but, nonetheless, may well be making a mistake in their participation decision because of some other bias such as myopia or present bias that causes them to under-save.³⁰ Whether a welfarist should embrace consistent choice paternalism in settings of this type depends on the strength of the evidence that the choices of the consistent choosers are mistaken.

Observation 4: The distinction between quasi-paternalism and consistent choice paternalism tracks the concerns that underscore traditional arguments against paternalism.

Traditionally, non-paternalism is often advocated based on two types of arguments, one welfarist and one non-welfarist. The welfarist argument, most strongly associated with John Stuart Mill, is that individuals are more likely to know which of some set of available options is better for them than is the government.³¹ This presumption is undermined by evidence that frames affect behavior – in such cases, one must balance the empirical predictive power of this presumption against the likelihood that decision-makers' preferences really do depend on features of the choice environment that appear to be arbitrary. And although Mill's presumption is likely to be a good rule of thumb, in many cases it will not be stronger than the competing belief that the decision-maker is making a mistake. For example, when savings plan enrollment decisions turn on whether enrollment is automatic, it seems more likely that Mill's presumption is incorrect than it is that differences between the two enrollment designs reflect different preferences over how much to save.

Although settings characterized by framing effects undermine the strength of Mill's presumption for inconsistent choosers, they do not do so for consistent choosers. For the same reasons that one would normally assume that the goals of an employee who chooses to participate in a 401(k) plan would be

28. Such evidence might take the form of survey responses accompanied by a decision to opt back in to participation.

29. Assuming, as I do here, that someone who chooses not to participate under automatic enrollment would also choose not to participate if enrollment were opt-in.

30. For a number of arguments along these lines, see Ryan Bubba & Richard H. Pildes, *How Behavioral Economics Trims Its Sails and Why*, 127 HARV. L. REV. 1593, 1612–14 (2014).

31. John Stuart Mill, *On Liberty*, in ON LIBERTY 5, 7–8 (Stefan Collini ed., Cambridge Univ. Press 1989) (1859).

better served by enrolling the employee in the plan than by excluding the employee from the plan, one should also assume that an employee who consistently chooses to participate in a retirement plan (under both an opt-in and opt-out enrollment design) would, in fact, be better off participating in the plan than not participating. Put differently, evidence that a frame affects the behavior of some decision-makers does not suggest that the *other* decision-makers are making choices that diverge from their preferences. If anything, the fact that the choices of this group of decision-makers are stable across frames might make us more confident the choice reflects a considered judgment by the chooser.

Turning to the non-welfarist criticism of paternalism, here too one can make the case that quasi-paternalism is more in keeping with the underlying concerns as compared to consistent choice paternalism. Loosely speaking, the non-welfarist argument is based on the idea that there is value in treating people as if they are correctly expressing their preferences even if they are known to be making a mistake – and that the alternative to doing so is infantilizing.³² Certainly, quasi-paternalism is subject to this concern – after all, it is based on the premise that the inconsistent choosers are making mistakes in at least one of the frames in which they are observed. However, it is also the case that, having made a quasi-paternalistic judgment about which frame to adopt, the case for adopting that frame can only be strengthened by treating people as if they were not making mistakes.

To see why, return to the example of automatic enrollment in a government employer's retirement plan, discussed above. Suppose that the employer decided to adopt participation as the default option, and assume further that it made this decision in a quasi-paternalistic way. Which default would best promote social welfare depends on the preferences of both the consistent and the inconsistent choosers, and since the choice of which default to implement here was (by assumption) made in a quasi-paternalistic way, we know that the choices of the consistent choosers were assumed to reflect their preferences. Suppose further that the policymaker determined that some percentage (X) of the inconsistent choosers would be better off under participation, where X is some number less than or equal to 100. Importantly, we know that when participation is made the default, 100% of the inconsistent decision-makers would choose to participate.³³ And if we were to assume that none of those decision-makers were making a mistake, we would have to conclude that 100% of the inconsistent decision-makers preferred participation (at least under the participation default). Hence, the case for setting participation to be the default would be even stronger than under the judgment that was actually made. Thus, policies that are quasi-paternalistic can be thought of as being consistent with the

32. See, e.g., Joshua D. Wright & Douglas H. Ginsburg, *Behavioral Law and Economics: Its Origins, Fatal Flaws, and Implications for Liberty*, 106 NW. U. L. REV. 1033, 1088 (2012).

33. This argument assumes that there are no inconsistent choosers who specifically choose whichever option is not the default. The argument can be extended to take this possibility into account.

autonomy concerns behind non-paternalism because treating everyone as making a correct decision would simply bolster the case for whichever frame was thought to maximize social welfare absent that assumption. In other words, if a nudge is adopted on a quasi-paternalistic basis, the (welfarist) case for it can only be made stronger by taking the inconsistent choosers' preferences at face value.

IV. IMPLICATIONS FOR THE CHOICE OF NUDGE VS. MANDATE

The notion of quasi-paternalism provides a way of distinguishing between various types of behaviorally informed policies in terms of the extent to which they implicate paternalist concerns. Specifically, the framework suggests that there is a sense in which mandates are more paternalistic than nudges. This is because although nudges can be quasi-paternalistic, mandates can generally only be justified on grounds of consistent choice paternalism.³⁴

To understand why the arguments for mandates must go beyond quasi-paternalism, we can modify our running example to consider an employer that is choosing between nudging its employees to enroll in a retirement savings plan (by setting participation to be the default) and requiring its employees to participate in the plan (say, as a condition of their employment).

As summarized in Table 1, below, the key difference between the two policies has to do with the change in behavior of those employees who would consistently choose not to participate (even when participation is the default), when given the option. Employees in this group would end up participating under the mandate but would not participate under the nudge. In contrast, those employees who would consistently choose to participate under both defaults would end up participating under both the nudge and the mandate, as would the inconsistent decision-makers (whose choices simply mirror the default). Thus, whether the nudge or the mandate yields higher social welfare will generally turn on whether the employees who end up with a different option under the two policies are better off under one or the other. Hence, for the mandate to be better than the nudge for social welfare, it must be the case that forcing those decision-makers who consistently choose non-participation to participate would actually make those decision-makers better off.³⁵ Although this is theoretically plausible (the employees who consistently prefer non-participation might be present-biased and be better off choosing to save more), it is the type of thinking that violates quasi-paternalism; i.e., it is based on the idea that the consistent choosers are mistaken.

34. For present purposes, I will set aside other types of situation-specific considerations that might justify a mandate over a nudge such as third-party effects or administrative costs. The argument developed in this section expands on an argument sketched in Jacob Goldin & Nicholas Lawson, *Defaults, Mandates, and Taxes: Policy Design with Active and Passive Decision-Makers*, 18 AM. L. & ECON. REV. 438 (2016).

35. See Goldin and Lawson, *supra* note 34. For similar analyses in the context of 401(k) plans, see Bubb & Pildes, *supra* note 30, at 1625, and Cass R. Sunstein, *Nudges vs. Shoves*, 127 HARV. L. REV. F. 210, 212 (2014).

Thus, there is a sense in which mandates are not only less liberal than nudges in the sense of preserving choice, they are also more paternalistic, in that they are appealing only if one believes the consistent choosers are making a mistake.

Table 1: Choices Under Mandates vs. Nudges		
	Participation Nudge	Participation Mandate
Inconsistent Choosers	Participate	Participate
Consistent Choosers Who Choose Participation	Participate	Participate
Consistent Choosers Who Choose Non-Participation	Don't Participate	Participate

This conclusion that nudges can be quasi-paternalistic but mandates cannot helps shed light on a recent debate in the behavioral law and economics literature. Bubb and Pildes have claimed that the arguments used to support nudges suggest the desirability of mandates as well.³⁶ As they put it, “[I]t would be surprising if the main policy implication of the mounting evidence documenting the failure of individual choice was a turn toward regulatory instruments that preserve individual choice.”³⁷ But the discussion in this section shows the conclusion that some individuals are making mistakes does not in itself imply that mandates are desirable – the fact that frames affect the behavior of the inconsistent choosers does not imply anything about the quality of the choices made by the consistent choosers. Hence, Mill’s epistemic argument against paternalism applies here in full force: a nudge will be more desirable than a mandate unless policymakers have good reason to think that they know which option would be best for the consistent choosers and are confident enough in that determination to override the beliefs of the choosers themselves. For this reason, even a committed welfarist should generally be reluctant to favor a mandate over a nudge in settings where frames affect behavior.

36. Bubb & Pildes, *supra* note 30.

37. *Id.* at 1595.