American Economic Review 2021, 111(3): 787–830
https://doi.org/10.1257/aer.20191039

What Motivates Paternalism? An Experimental Study[†]

By Sandro Ambuehl, B. Douglas Bernheim, and Axel Ockenfels*

We study experimentally when, why, and how people intervene in others' choices. Choice Architects (CAs) construct opportunity sets containing bundles of time-indexed payments for Choosers. CAs frequently prevent impatient choices despite opportunities to provide advice, believing Choosers benefit. They violate common behavioral welfare criteria by removing impatient options even when all payoffs are delayed. CAs intervene not by removing options they wish they could resist when choosing for themselves (mistakes-projective paternalism), but rather as if they seek to align others' choices with their own aspirations (ideals-projective paternalism). Laboratory choices predict subjects' support for actual paternalistic policies. (*JEL* C92, D12, D15)

Normative discussions of paternalism have featured in economics, philosophy, and public policy for centuries (Locke 1764, Mill 1869, Thaler and Sunstein 2003). A wide range of government regulations, such as retirement savings mandates, restrictions on payday loans and investment products, various forms of consumer protection, the criminalization of suicide, and legal doctrines concerning undue inducement and unconscionability all address paternalistic concerns (Dworkin 1972, Zamir 1998). Paternalistically motivated social programs play a large role in the US economy (Mulligan and Philipson 2000, Moffitt 2003, Currie and Gahvari 2008), and in some instances even the business models of private companies are paternalistically

*Ambuehl: Department of Economics, University of Zurich (email: sandro.ambuehl@econ.uzh.ch); Bernheim: Department of Economics, Stanford University (email: bernheim@stanford.edu); Ockenfels: Department of Economics, University of Cologne (email: ockenfels@uni-koeln.de). Stefano DellaVigna was the coeditor for this article. This paper previously circulated under the title "Projective Paternalism." We thank three anonymous referees, Joshua Knobe, Yoram Halevy, Armin Falk, Glenn W. Harrison, Kiryl Khalmetski, Frank Schilbach, Cass Sunstein, Dmitry Taubinsky, Stefan Trautmann, and seminar and conference audiences in Berkeley, Copenhagen, Cork, Cornell, Heidelberg, Innsbruck, Nijmegen, NYU, Stanford, Toronto, and Zurich for helpful comments and discussions. Max Rainer Pascal Grossmann and Yero Samuel Ndiaye provided excellent research assistance. Ambuehl gratefully acknowledges support through a University of Toronto Connaught New Researcher Award and the University of Toronto Scarborough Wynne and Bertil Plumptre Fellowship. Parts of this paper were written while Ambuehl visited the briq Institute in Bonn. This research has been approved by the University of Toronto Research Ethics Board (protocol 10279). Ambuehl and Bernheim gratefully acknowledge funding from the Alfred P. Sloan foundation (grant G-2017-9017). Ambuehl and Ockenfels gratefully acknowledge funding from the European Research Council (ERC) under the European Union's Horizon 2020 Research and Innovation Programme (grant 741409); the results reflect only the authors' views; the ERC is not responsible for any use that may be made of the information it contains. Ockenfels gratefully acknowledges funding by the German Research Foundation (DFG) under Germany's Excellence Strategy (EXC 2126/1-390838866).

[†]Go to https://doi.org/10.1257/aer.20191039 to visit the article page for additional materials and author disclosure statements.

motivated.[1] Despite the prevalence of paternalism, a *positive* investigation of the phenomenon is largely lacking. Significantly, paternalistic decision-making often falls to voters (Faravelli, Man, and Walsh 2015), low-level government officials (Moffitt 2006), and managers of private firms, rather than to experts, omniscient benevolent planners (Besley 1988), or despotic autocrats (Glaeser 2005). A large body of research catalogues the fallibility of human judgment when people choose for themselves (Kahneman 2011). It is reasonable to assume that normative judgments that guide paternalism are also susceptible to systematic confusion or bias (Rizzo and Whitman 2019).

This paper studies experimentally when, why, and how people act paternalistically. These questions are difficult to address in naturally occurring contexts because real-world policies with paternalistic elements generally implicate non-paternalistic concerns, such as externalities, and in any event people usually disagree about their efficacy. Conducting a laboratory experiment allows us to remove extraneous factors and study paternalistic behavior in isolation. In order to demonstrate external validity, we show that our main results extend to subjects' assessments of real-world paternalistic policies, and that these assessments are directly related to subjects' decisions within the experiment.

Subjects in the role of Choice Architects construct choice sets that determine the opportunities available to others in the role of Choosers. Each choice option consists of two payments, one received the day of the experiment ("sooner") and the other received half a year later ("later"). Impatience is costly: larger earlier payments are associated with smaller total payments. We employ the intertemporal choice domain because of its central importance across all of economics, because of ex ante reasons to believe that certain choices would meet with disapproval, and because it allows us to examine normative judgments that feature prominently in applications of behavioral welfare economics, including in a substantial and rapidly expanding literature (reviewed in Bernheim and Taubinsky 2018) that argues for a wide range of paternalistic policies.

In the first part of our analysis, we show that Choice Architects frequently choose to restrict others' choice options, then we establish several fact patterns that shed light on their motivations. Significantly, subjects frequently withhold options from Choosers despite ample opportunities to provide advice. The typical intervention requires the Chooser to exercise a minimum level of patience. For instance, when the available options include receiving €0 sooner and €15 later, 15 percent of Choice Architects withhold the opportunity to receive €2 sooner and €9 later, and 31 percent withhold the opportunity to receive €4 sooner and €2 later. Furthermore, Choice Architects are more likely to withhold options when the relative price of current consumption is greater. Advice concerning the options that remain in Choosers' opportunity sets follows similar patterns. Thus, Choice Architects suppress impatience either by preventing it or by advising against it.

[1] The Vanguard Group, for instance, argues that financial advisors should attempt to help clients by "providing discipline and reason to clients who are often undisciplined and emotional" and that they "can act as emotional circuit breakers … by circumventing their clients' tendencies." See Donald G. Bennyhoff and Francis M. Kinniry, "Advisor's Alpha," available at https://advisors.vanguard.com/iwe/pdf/ICRAA.pdf.

Next, we ask whether these interventions are paternalistic, in the sense that Choice Architects believe they improve the Choosers' well-being (Dworkin 1972). According to both incentive-compatible and non-incentivized measures, Choice Architects by and large believe the restrictions they impose are helpful. Furthermore, those who impose more severe restrictions believe restrictions are more beneficial. We rule out alternative motives, such as the possibility that subjects intervene simply because they prefer to do something rather than remain idle, or because they enjoy exercising control over others.

The next step in our analysis is to investigate whether Choice Architects' interventions conform to the types of normative judgments that often appear in applications of behavioral welfare economics. One common position maintains that the availability of immediate rewards, coupled with a tendency to place "too much" weight on near-term experiences ("present bias"), generate excessive impatience (e.g., O'Donoghue and Rabin 1999, 2006; Gruber and Kőszegi 2001, 2004; DellaVigna and Malmendier 2004). The findings summarized above are potentially consistent with the hypothesis that Choice Architects adopt that perspective. Is the alleviation of perceived present bias in fact a central objective of the observed interventions? If it is, removing the lure of immediacy by delaying the receipt of all payments should substantially reduce interventions. Yet we find that people impose slightly *more* patience in otherwise equivalent decision problems with "front-end delay." Thus, their behavior indicates that they generally disapprove of impatience, not that they are concerned about present bias.

Another hypothesis of interest is that paternalists make decisions for others based exclusively on their own assessments of right and wrong without regard to the preferences or subjective experiences of the affected individuals. We find, on the contrary, that even non-libertarians respect Choosers' preferences to a degree, but are unwilling to tolerate sufficiently impatient choices. In particular, Choice Architects respond to information about Choosers' preferences in an accommodating direction. For instance, they impose substantially more patience on Choosers who describe themselves as impatient and unhappy than on those who claim to be content with their impatience.

The second part of our investigation asks how Choice Architects decide which options are good for others and which are bad. Their sensitivity to Choosers' inclinations and subjective experience implies that these considerations play important roles in Choice Architects' evaluations. But when their information about the Chooser is incomplete, how do they arrive at their evaluations? In *The Theory of Moral Sentiments*, Adam Smith argued that such inferences follow from self-examination: "As we have no immediate experience of what other men feel, we can form no idea of the manner in which they are affected, but by conceiving what we ourselves should feel in the like situation." In this spirit, we identify two distinct types of departures from the benchmark case of perfect knowledge concerning others' well-being and biases, distinguished according to whether paternalists reason about others based on their own mistakes, or based on their own preferences. A *mistakes-projective paternalist* assumes others tend to share her susceptibility to error. She behaves as if she tries to help others avoid choices she herself would like to reject, but chooses nevertheless. We demonstrate that this inclination generates a *negative correlation* between the choices she makes for herself and the restrictions

she imposes on others. In contrast, an *ideals-projective paternalist* behaves as if she assumes her own preferences are relevant for others, either because she thinks they tend to share her values, or because she simply believes her perspective is valid and theirs are not. Ideals-projective paternalism generates a *positive correlation* between the choices paternalists make for themselves and the restrictions they impose on others.

We find strong support for ideals-projective paternalism. More patient Choice Architects impose greater patience on others, and this pattern reflects their judgments about what the Choosers ought to do, rather than a greater inclination to intervene. They also believe more strongly that restrictions enhance the Choosers' well-being.

Next, we ask whether paternalists simply express their welfare judgments based on accurate beliefs about others, or whether paternalistic interventions are susceptible to systematic cognitive biases. To answer this question, we focus on Choice Architects' beliefs about the choices unrestricted Choosers would make. We examine how those beliefs relate not only to the restrictions Choice Architects impose, but also to their own preferences. We find that interventions are closely related to beliefs about Choosers' inclinations. Moreover, those beliefs are systematically biased: subjects assume (incorrectly) that the choices of others tend to resemble their own (a *false consensus effect*, Ross, Greene, and House 1977). As a result, despite the tendency for more patient Choice Architects to impose greater patience, patient and impatient Choice Architects expect their mandates to bind with about the same frequency. We conclude that interventions are systematically misguided, even according to the Choice Architects' aims.

The third part of our analysis explores the external validity of our findings. We elicit subjects' support for imposing sin taxes and for regulation of high-interest, short-term lending in a neighboring country. We also measure subjects' own consumption of the targeted products. We find that real-world preferences for paternalistic policies that only impact others are strongly correlated with Choice Architects' decisions in the behavioral portion of our experiment. Additionally, the relationships between policy preferences and own consumption are consistent with ideals-projective paternalism. For instance, lighter drinkers express significantly more support for an increase in alcohol taxes in another jurisdiction.

Our work is related to a small empirical literature on paternalistic behavior. The closest parallels are Uhl (2011) and Krawczyk and Wozny (2017). In both studies, subjects choose between one of two options for themselves and decide whether to eliminate one of the options for others.[2] While these studies examine the proclivity to intervene, as well as correlations between chosen interventions and the options subjects select for themselves, they do so in settings that preclude the dispensation of advice. Accordingly, it is impossible to tell whether the subjects see intervention as intrinsically desirable, or merely as the only feasible method of expressing their opinions. Furthermore, the close juxtaposition of essentially identical decisions (for

[2] In Krawczyk and Wozny (2017), the options are two lunch items, one healthy, the other unhealthy. In Uhl (2011), the options are whether to make a second choice (i.e., choosing the point in time at which to collect a rising payment) either in advance or in the moment. The empirical literature on paternalistic behavior also includes Jacobsson, Johannesson, and Borgquist (2007); Lusk, Marette, and Norwood (2013); Gangadharan, Grossman, and Jones (2015); Schroeder, Waytz, and Epley (2017); Zlatev et al. (2017); and Daniels and Zlatev (2019).

the subjects themselves and for others) may well introduce spurious correlation through anchoring and/or a demand for consistency, which undermines inferences concerning the connection between own-preferences and interventions. Setting these important design issues aside, these papers have more limited objectives than ours, and consequently do not address the other important questions that structure our analysis: whether interventions are benevolently motivated; whether paternalists care about and respond to information concerning the preferences of the impacted parties; whether they exhibit less inclination to intervene when those parties make their choices under conditions that are commonly thought to mitigate decision biases; whether paternalistic interventions implicate systematic cognitive biases; and whether paternalism in the lab relates to real-world policy preferences. Likewise, they do not address the distinction, central to this paper, between ideals-projective and mistakes-projective paternalism. In a more recent paper, Bartling et al. (2020) focus on decisions to intervene paternalistically by removing dominated options. They largely abstract from motivations involving disapproval of others' preferences, which is a central feature of the current paper.

Several other lines of work bear on the empirical analysis of paternalism.[3] One examines the tendency for professional advisors to steer their clients toward the same options they choose for themselves (Foerster et al. 2017; Linnainmaa, Melzer, and Previtero forthcoming), a possible reflection of ideals-projective paternalism. A second studies how people feel about being in situations where others can influence or constrain their choices (Fehr, Herz, and Wilkening 2013; Bartling, Fehr, and Herz 2014; Kataria, Levati, and Uhl 2014; Lübbecke and Schnedler 2018; Ackfeld and Ockenfels 2021). A third studies social disapproval of ostensibly *repugnant transactions* (Roth 2007), such as paid organ donation (Basu 2003, 2007; Leider and Roth 2010; Elías, Lacetera, and Macis 2015a, b, 2019; Ambuehl 2017; Ambuehl and Ockenfels 2017; Clemens 2018; Exley and Kessler 2017).[4] A fourth explores how people make surrogate choices for others in settings where the surrogate cannot leave the affected individual with flexibility (see Ifcher and Zarghamee 2020 for a review).

The empirical study of paternalism connects with various other branches of the literature. The literature on *libertarian paternalism* argues that authorities can use *nudges*, rather than coercion, to adjust behavior in directions the authorities (or their policy analysts) deem beneficial (Thaler and Sunstein 2003; see Loewenstein and Haisley 2007 and Benartzi et al. 2017 for reviews). Our analysis sheds light on the types of paternalistic judgments that motivate nudges, and it identifies common biases that nudge designers could in principle learn to avoid. We contribute more broadly to a literature in cognitive science on moral heuristics (see Sunstein 2005 and Gigerenzer 2008 for reviews). There is also a connection to a literature on *projection bias* (Van Boven, Dunning, and Loewenstein 2000; Loewenstein, O'Donoghue, and Rabin 2003; Madarász 2012), in that we exhibit a manifestation of this phenomenon

[3] A larger theoretical literature relating to paternalism includes Saint-Paul (2004); Carlin, Gervais, and Manso (2013); Bisin, Lizzeri, and Yariv (2015); Altmann, Falk, and Grunewald (2017); and Laibson (2018).

[4] Ambuehl, Niederle, and Roth (2015) sketch a model with agents who exhibit a form of ideals-projective paternalism that explains why the introduction of monetary incentives can cause people to judge transactions as unethical (e.g., paid kidney donation), even though they approve of the same transaction in the absence of such incentives (e.g., in-kind kidney exchange).

in the normative domain. Finally, a collection of recent empirical studies evaluate particular paternalistic policies by deploying normative principles from behavioral welfare economics; see Bernheim and Taubinsky (2018) for a review, and Bernheim and Rangel (2009) or Bernheim (2016) for foundations.

The remainder of this paper proceeds as follows. Section I outlines our experimental design. Section II demonstrates that Choice Architects frequently force others to choose patiently, and that they believe these interventions benefit the Choosers. It also explores the responsiveness of paternalistic interventions to the introduction of front-end delay, and of information about Choosers' preferences. In Section III, we investigate hypotheses about the formation of Choice Architects' views concerning Choosers. We begin by formalizing the paternalists' decision problem, introducing the concepts of ideals-projective and mistakes-projective paternalism, and deriving their implications. After documenting patterns that point to ideals-projective paternalism, we explore the prevalence and implications of mistaken beliefs about the selections Choosers would make in absence of interventions. Section IV demonstrates that laboratory choices predict support for real-world paternalistic policies, and it documents ideals-projective paternalism regarding those policies. Finally, Section V outlines directions for further research.

I. Experiment Design

We begin by explaining the design of our experiment. Section IA describes the main types of decision problems we use to investigate paternalism. Section IB provides an overview of the structure of the experiment. The remaining subsections then present details concerning incentivization (Section IC), the Choosers' decisions (Section ID), and implementation (Section IE). For easier readability, this section condenses the presentation of our design. A comprehensive description appears in online Appendix Section D.1.

A. *Main Elements of the Experiment*

Each subject in our experiment is either a Choice Architect or a Chooser. Our interest is in the Choice Architects, who determine the set of options that will be available to Choosers. The experiment's main elements, described below, shed light on the following four questions. (i) How do Choice Architects construct opportunity sets for Choosers? (ii) Do Choice Architects believe that withholding options helps or hurts Choosers? (iii) What options do Choice Architects select for themselves? (iv) What do they believe Choosers would select absent restrictions? We discuss each of these elements in turn.

Construction of Opportunity Sets.—The Choice Architect constructs the Chooser's opportunity set from a menu of three options, as illustrated in panel A of Figure 1. Each option is a bundle of two monetary payments, one received the day of the experiment, the other received with a half-year delay. We design the options so that a Chooser can increase his present payment only by accepting a smaller amount overall. The Choice Architect must *actively* decide whether the opportunity set will

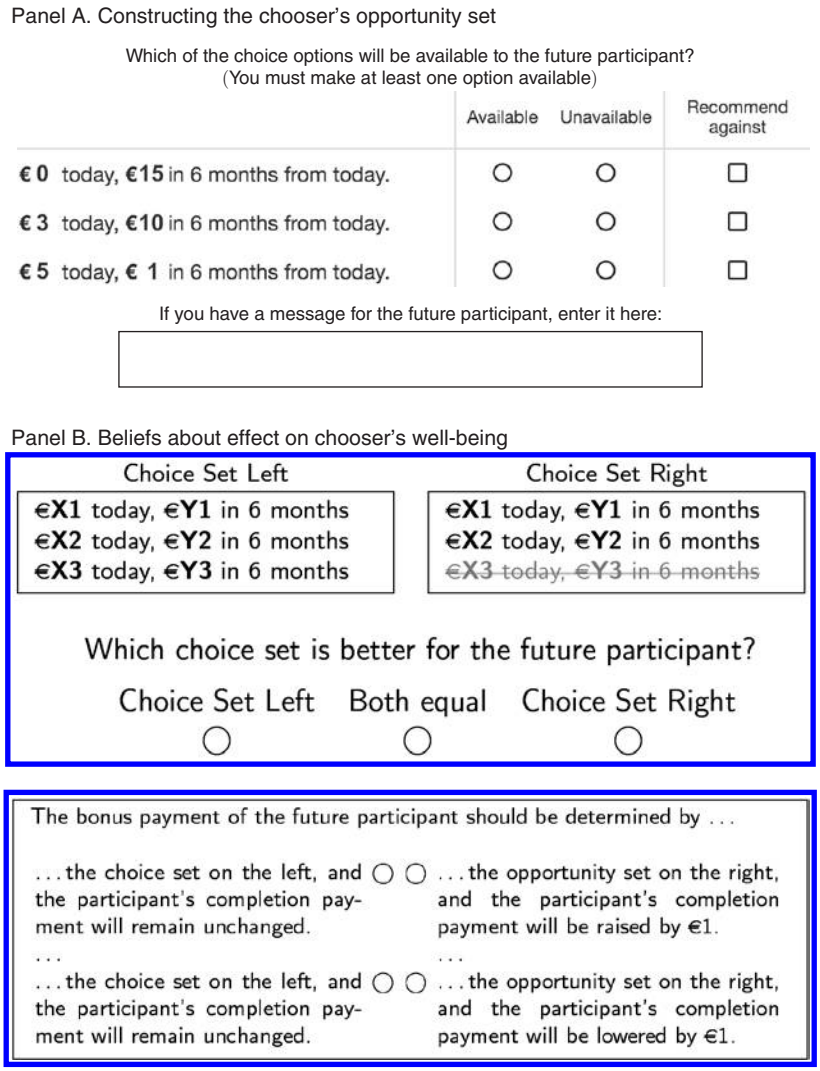

FIGURE 1. DECISION SCREENS FOR THE CHOICE ARCHITECT IN THE MAIN CONDITION

include each option; neither inclusion nor exclusion is a default. The sole restriction is that each opportunity set must include at least one option. We emphasize to subjects that there are no right or wrong answers, and that they should construct opportunity sets based on their genuine views.

We study the characteristics of the opportunity sets Choice Architects construct. Three design features limit the possible explanations for imposing restrictions. First, Choice Architects can advise Choosers. Specifically, in each round, the Choice Architect can write an unrestricted message to the Chooser, which the Chooser observes before making his decision. The Choice Architect can also convey disapproval of any option by clicking a button. In that case, the Chooser sees a red asterisk next to the corresponding option, accompanied by a statement that a previous participant advises against it. The existence of these opportunities is central to our

objective. Without them, Choice Architects might remove options simply because they have no other way to convey opinions and advice, rather than because they perceive a genuine need for restrictions. Accordingly, this feature allows us to detect situations in which Choice Architects feel they cannot trust informed Choosers to make good decisions (an expression of paternalism), rather than situations in which Choice Architects merely believe they have better information than Choosers (a state of affairs that does not necessarily implicate paternalism).

Second, we ensure that Choice Architects can only influence Choosers' outcomes, not their decision processes. In particular, the Choice Architect cannot save the Chooser time or effort, or spare him the ordeal of resisting temptation. The reason is that the Chooser ranks all three of the options that might be in his opportunity set, without knowing which are actually available or how their availability is determined. He then receives the option he has ranked most highly among those that are actually available. Choice Architects are aware of this procedure.

Third, the Choice Architect's material payoff is independent of her decisions concerning the Chooser's opportunity sets. Accordingly, the Choice Architect's only plausible motivations involve her feelings about the Chooser's autonomy or consequences. Here, our objective is to avoid confounding paternalistic motivations with conflicts of interest. This feature mirrors an important property of many paternalistic decisions. For example, members of Internal Review Boards charged with protecting human subjects are usually precluded from having personal interests in any research that is subject to their oversight.

Three additional features of this setting merit emphasis. First, our experiment focuses on *hard paternalism* (restricting opportunity sets) rather than *soft paternalism* (influencing choice without changing opportunity sets). Soft paternalism introduces other potentially confounding considerations. For example, the attractiveness of employing a nudge depends in part on beliefs about its efficacy. In our setting, efficacy is unambiguous. Second, we study paternalistic decisions by individuals, rather than groups. While many paternalistic policies result from group decision-making (e.g., through voting), the judgments of individuals are always central. Third, by using opportunity sets involving bundles of immediate and delayed monetary payments, our experiment introduces a plausible a priori rationale for paternalism: people commonly view patience as virtuous and impatience as reflecting weakness.

Elicitation of Beliefs about the Welfare Effects of Restrictions.—A Choice Architect's decision to withhold options is paternalistic only if she believes it promotes the Chooser's well-being (Dworkin 1972). We measure these beliefs in two ways on a single screen, as illustrated in panel B of Figure 1. The Choice Architect sees an opportunity set on the left, along with a subset of the same options on the right. As explained subsequently, in some rounds the subset on the right is the one the Choice Architect constructed; in other rounds it is given exogenously.[5]

Choice Architects first answer a simple non-incentivized question: *Which choice set is better for the future participant?* They select between *Choice Set*

[5] In most cases, the opportunity set on the left contains three options. However, for one-half of the rounds involving exogenous restrictions, the opportunity set on the left contains two options, and the one on the right contains only the most patient option. See Section IIE and online Appendix Section B.1 for additional details.

Left, *Both equal*, and *Choice Set Right*. Second, Choice Architects complete a decision list that elicits their beliefs about the welfare effects of restrictions. Assuming they are not entirely indifferent toward Choosers, the elicitation is incentive-compatible. Each line of the list presents a binary choice of the following form: *The payment of the future participant should be determined by …* (i) *the choice set on the left*, *and the participant's completion payment will remain unchanged*, *OR* (ii) *the choice set on the right*, *and the participant's completion payment will be raised/lowered by* €p, with $p \in \{1, 0.5, 0.3, 0.1, 0.05, 0, -0.05, -0.1, -0.3, -0.5, -1\}$. A Choice Architect may believe, for instance, that enlarging an opportunity set requires positive compensation of €0.4 to the Chooser if the additional choice options create opportunities for errors. A benevolent Choice Architect will prefer the first option if $p = -0.5$, but will prefer the second if $p = -0.3$. Thus, the transfer p at which a benevolent Choice Architect switches from (i) to (ii) reveals her beliefs about the payment that compensates the Chooser for receiving the opportunity set on the right over the one on the left. We impose no restrictions on how subjects fill in these lists such as monotonicity. Because we implement (at most) one decision from one list, the Choice Architect has an incentive to choose in accordance with her genuine preferences, regardless whether she is benevolent or malevolent.[6]

Elicitation of Choice Architects' Time Preferences.—Our investigation of paternalistic mechanisms involves comparisons between the opportunity sets Choice Architects construct for Choosers and the choices they make for themselves. All Choice Architects complete six decision lists such as the one shown in Figure 2. Each line is a choice between €x_{early} the day of the experiment and €x_{late} t months later.[7]

To ensure that anchoring or a demand for consistency does not produce an artificial relationship between the decisions Choice Architects make for themselves and those that impact others, we make it difficult for them to compare options across these settings. When choosing for themselves, the amount of delay differs across the options but the monetary payments are the same. When constructing opportunity sets for Choosers, the monetary payments differ across options but the amount of delay is the same. We also limit the potential influence of confounding mechanisms by requiring Choice Architects to make choices for themselves online three to six days before the laboratory session, and by interspersing these tasks with decision lists involving risk taking to obfuscate their purpose.

The temporal separation between the online session and the laboratory session implies that the earlier payments in the online session are not immediate. This lack of immediacy could in principle cause Choice Architects to think about the options available for themselves in the online session and those available to Choosers in the laboratory session somewhat differently. Consequently, toward

[6] Our elicitation differs in two ways from the types of incentive-compatible belief elicitation techniques commonly used in experimental economics. First, we add or subtract the amount p to or from the Chooser's completion payment, not the Choice Architect's completion payment, even though the Choice Architect is the party expressing the beliefs. Second, we do not compensate the Choice Architects based on the distance between their expressed beliefs and some objective truth. Our approach more closely resembles the incentive-compatible methods used to elicit willingness-to-pay rather than those used to elicit beliefs.

[7] We use $(x_{early}, x_{late}) \in \{(2, 10), (5, 10), (8, 10), (2, 15), (7, 15), (12, 15)\}$ and t between 1 and 6 months in steps of 1 month.

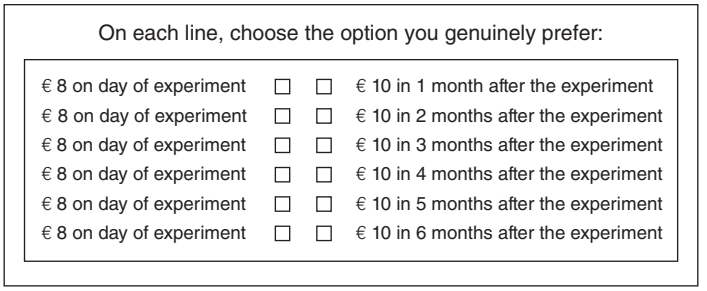

FIGURE 2. CHOICE ARCHITECTS' OWN INTERTEMPORAL CHOICES

the end of the laboratory session, Choice Architects make additional selections for themselves from each three-option set for which they previously constructed Choosers' opportunity sets. While these choices potentially raise greater concerns about anchoring or a demand for consistency than those made in the online session, we find little evidence that these confounds are in fact present in our experiment; see online Appendix Section B.2.

Elicitation of Beliefs about Choosers' Unrestricted Choices.—We elicit Choice Architects' beliefs about the distributions of unrestricted choices made by ten previous subjects for each menu encountered elsewhere in the experiment. Choice Architects drag and drop ten tags labeled "Participant" into bins representing each of the three choice options. Systematic discrepancies between these beliefs and the actual (known) distribution of unrestricted choices may imply that paternalistic interventions are predicated on mistaken perceptions of Choosers' tendencies. To ensure incentive compatibility, a Choice Architects' compensation may be based entirely on the accuracy of their reported beliefs.

B. *Structure of the Experiment and Additional Elicitations*

As noted above, the experiment consists of an online component and a laboratory component. The online component elicits Choice Architects' own intertemporal preferences. The laboratory component consists of three stages. Stage 1 includes 14 rounds of paternalistic decisions and assessments of perceived welfare effects that employ various menus of options from which Choice Architects construct choice sets. The *Main* condition comprises four rounds, each of which proceeds as described above. These rounds employ menus 1 to 4 shown in Table 1. For each subject, we randomly select the round involving either menu 1 or 2 and delay both the early and the late payment for each option in that menu by one week. As we discuss subsequently, the introduction of front-end delay allows us to test some specific ideas about the nature of paternalism. Each Choice Architect also participates in three additional conditions (ten rounds in total) for which we alter the decision problems described above to test specific hypotheses about the determinants of paternalism. These include the *Exogenous Restriction* condition (four rounds using menu 6), the *Chooser Information* condition (four rounds using menu 5), and the *Induced Chooser Preference* condition (two rounds using menus 3 and 4).

Table 1—Menus of Options from Which Choice Architects Construct Opportunity Sets

	Menu 1		Menu 2		Menu 3	
Option	Today	In 6 months	Today	In 6 months	Today	In 6 months
Most patient	\$0	\$15	\$0	\$15	\$0	\$15
Middle	\$3	\$10	\$3	\$9	\$2	\$12
Least patient	\$5	\$1	\$6	\$1	\$4	\$2
	Menu 4		Menu 5		Menu 6	
Option	Today	In 6 months	Today	In 6 months	Today	In 6 months
Most patient	\$0	\$15	\$0	\$15	\$0	\$15
Middle	\$2	\$9	\$4	\$6	\$3	\$7
Least patient	\$4	\$2	\$5	\$1	\$4	\$1

Notes: Menus 1 to 4 correspond to the Main condition. Menu 5 corresponds to the Chooser Information condition (Section IID). Menu 6 corresponds to the Exogenous Removal condition (Section IIE).

We describe these conditions when we encounter the hypotheses they address in Sections IIB, IID, and IIE, respectively.

Each subject proceeds through the rounds of Stage 1 in an individually randomized order. In each round, the Choice Architect first constructs the Chooser's opportunity set, and then reveals her beliefs about whether the three-option opportunity set or a subset thereof is better for the Chooser. In the Main condition, the comparator subset is the one the Choice Architect constructed herself.

In Stage 2, we collect additional data that shed light on aspects of the Stage 1 decisions. First, Choice Architects make surrogate choices for Choosers. These tasks are identical to those in Stage 1, except we require Choice Architects to select a single option. Assuming Choice Architects are benevolent, these decisions reveal the options they deem best for Choosers. Unlike decisions to restrict opportunity sets, surrogate choices do not implicate Choice Architects' willingness to intervene. Second, to incentivize attentiveness, we administer an eight-question test on specific features of the experiment. We tell subjects about this test in advance, and advise them that their performance on it could completely determine their earnings.[8] Third, we elicit beliefs about previous Choosers' unrestricted choices, as described above.[9] Fourth, Choice Architects select an option for themselves from each menu, also as described above. Fifth, we ask Choice Architects to adjust the completion payment of a different Chooser. They can costlessly increase that payment by €1, leave it unchanged, or decrease it by €1. We use their responses to gauge whether they are benevolent or spiteful toward Choosers.

In Stage 3, subjects express opinions about four real-world paternalistic policy proposals, and then provide information about their own inclinations to engage in the affected activities. We use this information to evaluate the generalizability of our findings. The experiment ends with a brief memory check concerning choices subjects made in the online component.

[8] Online Appendix Section B.3 lists the test questions and frequencies of correct responses. Subjects do not learn anything about the content or focus of the test before making decisions concerning the Chooser. The test consists of eight questions about the stimuli the Choice Architects encountered. It does not refer to Choice Architects' own decisions.

[9] Half of our subjects, chosen at random, complete these elicitations before Stage 1. We control for this order in all of our regressions.

Online Appendix Section D.1 presents comprehensive detail about all design elements.

C. *Incentives*

Choice Architects' Decisions Concerning Choosers.—There is a 25 percent chance that we match any given Choice Architect with a Chooser. For each match, we randomly draw one of the rounds in which the Choice Architect makes a decision concerning the Chooser. With 50 percent probability, we implement the decision from the first half of that round. With the remaining 50 percent probability, we determine the Chooser's opportunity set by randomly drawing a line from the decision list in the second half of the round (which elicits the perceived benefit of a restriction) and implementing the Choice Architect's selection. Separately, with 25 percent probability, we implement the Choice Architect's chosen adjustment to the completion payment for a different Chooser who we assign at random. Choice Architects know that Choosers will participate in a subsequent laboratory session. We inform Choice Architects of the matching and implementation probabilities, and explain that, if they are matched to a Chooser, no other subject will influence the Chooser's opportunities. We also let Choice Architects know that no other subject has manipulated the choice problems that determine their own payments.

Choice Architects' Own Payment.—A Choice Architect's payment is determined either by the online component or by one of the following: the attention test, elicitation of beliefs concerning Choosers' unrestricted selections, or choices made for themselves. Each of these four alternatives is equally likely.[10] For the elicitation of beliefs concerning unrestricted Choosers' selections, the Choice Architect receives €10 minus the number of tags we must reassign to make the elicited and observed distributions of choices coincide.[11] For the attention test, the Choice Architect receives €1 for each correct answer. For the online component or the Choice Architects' decisions for herself in the laboratory component, we implement one randomly selected decision within the corresponding task block. Each Choice Architect also receives a show-up payment of €4.5 and a completion payment of €8.[12]

D. *Choosers*

Choosers participate after all Choice Architect sessions are completed. Each Chooser first selects one of four self-descriptive statements (see Section IID) and then ranks the elements of one menu. He receives the option he ranks highest among

[10] Subjects learn at the beginning of the online component that there is a 25 percent chance a single decision from that component will determine their payment, and a 75 percent chance the laboratory component will determine it, but they do not learn at the outset what the latter possibility will entail.

[11] Our elicitation procedure is the balls-in-bins method described in Delavande, Giné, and McKenzie (2011). Truthful revelation is optimal for a risk-neutral subject. Subjects understand this scheme more easily than alternatives. While risk aversion theoretically generates a tendency for measured beliefs to be overly dispersed, Choice Architects' risk preferences, elicited in the online component, predict neither the location nor the dispersion of elicited belief distributions.

[12] The completion payment in the first two sessions was €5, which we increased after feedback that subjects perceived the study payment to be low.

those his matched Choice Architect makes available. Each Chooser also receives a show-up payment of €4.5 and a completion payment of €8. If the Choice Architect task selected for implementation involves the incentive-compatible elicitation of beliefs concerning welfare effects, we raise or lower the corresponding Chooser's completion payment based on the Choice Architect's decision.

E. *Implementation*

We conducted 16 sessions with a total of 303 Choice Architects during the summer of 2018 at the Cologne Laboratory for Economic Research. Each lasted approximately 90 minutes. We recruited 100 additional subjects to study other hypotheses (see the discussion of the Choice Distribution Information condition in online Appendix Section B.4). Separately, 124 subjects participated as Choosers (see online Appendix Section B.5).[13]

The experiment is computer-based. We display instructions on-screen, and intersperse comprehension checks which subjects must complete correctly to continue. Those failing the comprehension checks can review the instructions and try again (all subjects eventually passed). The comprehension checks emphasize that there are no right or wrong answers for decisions affecting the Choosers. We process all incentive payments through PayPal and cover all transaction fees.

II. Properties of Paternalistic Interventions

A. *Do People Intervene and, If So, in What Way?*

The first step in our analysis is to ask whether people intervene into others' decisions and to characterize the nature of those interventions. We find that interventions are common, and that they generally limit impatient choices, especially when impatience is costly.

Figure 3 shows the frequencies with which Choice Architects make specific types of options unavailable, averaged across rounds in the Main condition (excluding those with front-end delay). The prevailing tendency is for Choice Architects to prevent impatient choice. They remove the least patient option 32.7 percent of the time, the middle option 11.6 percent of the time, and the most patient option 5.1 percent of the time. When we limit attention to the 86.3 percent of Choice Architects who are altruistic (in the sense that they costlessly increase a second Chooser's completion payment in Stage 2), the frequency with which they remove the most patient option drops by one-half (to 2.5 percent). The removal frequency for the middle option falls slightly (to 10.6 percent), and there is no change for the least patient option. As Table 2 shows, the pattern of withholding impatient options rather than patient ones emerges for all four menus.

Imposing a minimum degree of patience is also the modal behavioral pattern on the individual level: 44.9 percent of our subjects remove at least one option from at least one of the opportunity sets, and never remove an option without also excluding

[13] Replication data are available in Ambuehl, Bernheim, and Ockenfels (2021).

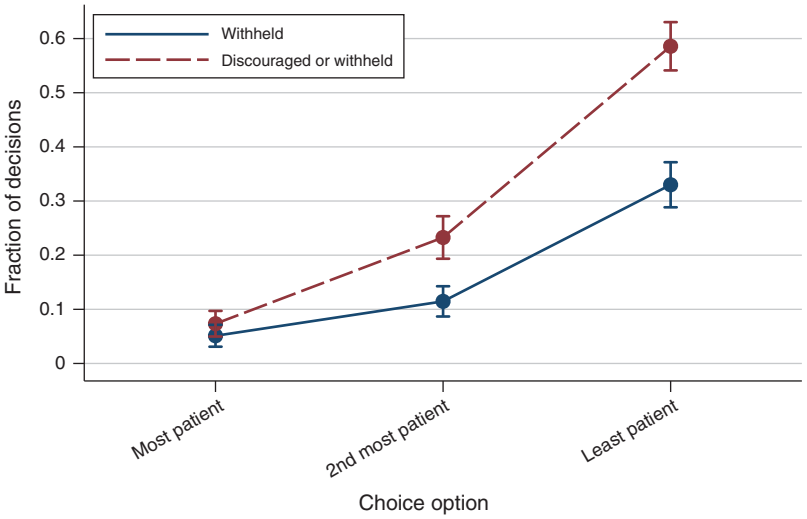

FIGURE 3. WITHHELD AND DISCOURAGED OPTIONS

Notes: The solid line displays the fraction of cases for which the Choice Architect withholds options, categorized by relative patience, from the Chooser. The vertical distance between the solid and dashed lines represents the fraction of cases in which the Choice Architect advises against the option but does not withhold it. We average the data across the menus (1 to 4) in the Main condition, excluding rounds with front-end delay. Whiskers indicate 95 percent confidence intervals based on estimates of standard errors clustered at the subject level.

TABLE 2—FRACTION OF OPTIONS WITHHELD IN MAIN CONDITION

	Percent option unavailable				
Menus	Most patient	Middle	Least patient	Observations	Subjects
1, 2, 3, 4	5.1	11.6	32.7	909	303
1	6.0	9.3	32.0	150	150
2	5.2	14.4	23.5	153	153
3	4.3	8.3	38.9	303	303
4	5.3	14.5	31.4	303	303

Note: Decisions with front-end delay are excluded.

less patient options. Choice Architects who never remove any option in the Main condition comprise the second largest category (38.3 percent). We refer to this group as Libertarians. A small fraction (4.3 percent) of Choice Architects impose an upper bound on patience.[14] The remaining 12.5 percent of Choice Architects impose non-monotonic restrictions.

Choice Architects' decisions depend not only on the rank (by patience) of an option within a menu, but also on the cost of impatience, defined as the total amount of money the Chooser forfeits by acting impatiently. In menu 3, for instance, the Chooser sacrifices a total of €1 (€15 − €14) by selecting the middle option rather than the most patient option. In menus 1, 2, and 4, the corresponding figures are €2, €3, and €4, respectively. Removal rates for the middle option increase monotonically with the forfeited amount: they are 8.3 percent for menu 3, 9.3 percent

[14] These subjects remove at least one option from at least one choice set, and never remove an option without also excluding more patient options.

for menu 1, 14.4 percent for menu 2, and 14.5 percent for menu 4 (see Table 2). A similar pattern holds for the least patient option. By selecting the least patient option from menu 2 rather than the most patient option, a Chooser forfeits €8 in total, whereas the corresponding amount is €9 for each of the remaining menus. The removal rate for the least patient option is 23.5 percent for menu 2, compared with 32 percent, 38.9 percent, and 31.4 percent for menus 1, 3, and 4, respectively. To quantify these relationships, we estimate an OLS regression relating the availability of an option to the total amount of money a Chooser would forfeit by selecting it rather than the most patient option, controlling for the rank of the option within the choice set, as well as session and order fixed effects. We find that, for each additional Euro sacrificed when selecting an impatient option, the frequency with which Choice Architects withhold that option increases by 2.9 percentage points ($p < 0.01$, see online Appendix Section B.6 for details).

The frequency with which subjects advise against options but do not withhold them exhibits similar patterns: see Figure 3. Subjects advise against impatient options substantially more often than against patient ones, just as they tend to remove the former more than the latter. Thus, patterns of hard paternalism (prohibitions) and soft paternalism (advice) mirror each other. Significantly, the pattern of advice is essentially the same, both qualitatively and quantitatively, for subjects who never withhold options. Subjects write free-form messages much less frequently (only 21.5 percent of subjects ever send a message exceeding 5 characters in length), but when they do, they typically articulate paternalistic perspectives concerning the desirability of patience. Examples include "*Patience is a virtue*," "*Stay rational. Don't weigh the present more than the future*," and "*It is important to practice patience*."

B. *Do People Believe Their Interventions Are Beneficial?*

Having shown that interventions occur in our experiment, we now ask whether they are paternalistic. A restriction on free choice is deemed paternalistic only if the party responsible for it believes that it benefits those it affects (Dworkin 1972). In this section, we show that our Choice Architects think the restrictions they impose are beneficial, and that those who impose stricter mandates believe the benefits are greater. Accordingly, we are indeed observing paternalism.

We begin by analyzing subjects' responses to the question of whether the full opportunity set or the opportunity set they have constructed is better for the Chooser. Column 1 of Table 3 (top panel) shows the distribution of responses in the Main condition (without front-end delay). In 30 percent of cases, subjects report that the choice set they have constructed is better for the Chooser than the unrestricted set, while 11 percent report that it is worse. These frequencies are artificially attenuated by the presence of libertarian subjects who never remove options, and who therefore largely report that neither set is better. For rounds in which subjects imposed restrictions, 65 percent describe the restricted set as better, while 25 percent describe it as worse. Overall, subjects are far more likely to believe their interventions are helpful than hurtful.

To evaluate the statistical significance of these findings, we regress subjects' reports about the welfare effect on indicators for the number of options removed, encoding the responses "better," "same," and "worse" as $+1$, 0, and -1, respectively

Table 3—Subjects' Beliefs about the Welfare Effects of Withholding Options

Dependent variable:	Belief smaller opportunity set better for Chooser		Negative of compensating variation	
Smaller set:	Chosen	Exogenous	Chosen	Exogenous
Menus:	1–4 without FED	6	1–4 without FED	6
	(1)	(2)	(3)	(4)
Summary statistics for dependent variables				
Distribution of beliefs				
Evaluation opportunity set is				
better	0.296	0.345		
same	0.600	0.107		
worse	0.105	0.548		
Mean negative CV			0.021	−0.224
Distribution of beliefs if options withheld				
Evaluation opportunity set is				
better	0.652	0.455		
same	0.091	0.118		
worse	0.257	0.428		
Mean negative CV			0.051	−0.122
Regression results				
1 option withheld	0.320		0.019	
	(0.068)		(0.050)	
2 options withheld	0.496		0.113	
	(0.104)		(0.079)	
Mean number of options withheld in Main condition		0.686		0.251
		(0.073)		(0.055)
Comparison set contains most patient option only		−0.182		−0.253
		(0.088)		(0.062)
Observations	909	606	866	574
Number of subjects	303	303	297	292

Notes: The top half of the table provides summary statistics for the dependent variables, and the bottom half presents regression results. FED stands for front-end delay. Unit of observation: subject-round pair. Method: Columns 1 and 2, OLS; columns 3 and 4, interval regression. Dependent variables: the dependent variable for columns 1 and 2 measures whether the Choice Architect considers the smaller opportunity set better, equally good, or worse for the Chooser than the unrestricted set, coded as 1, 0, and −1, respectively. For columns 3 and 4, it is the negative of the Choice Architect's beliefs about the compensating variation (CV) of reducing the opportunity set. For all columns 1 and 3, the smaller opportunity set is the one the Choice Architect has constructed; for columns 2 and 4, we construct it by exogenously removing either the least patient option, or the least patient and middle options. Controls: All regressions control for altruism and spite, and include session and order fixed effects. Columns 1 and 3 also control for menu fixed effects Samples: Columns employ data from different conditions, as indicated. For columns 1 and 3, the row *Distribution of beliefs if options withheld* includes all observations in which the Choice Architect withheld one or more of the options in rounds of the Main condition without front-end delay. For columns 2 and 4, that row includes observations from all rounds in the Exogenous Restriction condition for Choice Architects who withheld at least one option in the Main condition. In the bottom panel, each column corresponds to a separate regression. Columns 3 and 4 exclude subjects with multiple switches in the elicitation of compensating variations. Standard errors: clustered by subject.

(online Appendix Section B.7 reports the corresponding ordered probit regressions). We use the full sample and control for altruism versus spite on the part of the Choice Architect, i.e., whether she has chosen to costlessly increase or decrease a second Choosers' completion payment by €1. We include session, order, and menu fixed effects, and cluster standard errors by subject. On average, a Choice Architect is more likely to believe that her action benefits the Chooser when she removes one option rather than none, and the difference is highly statistically significant, as the estimates in column 1 show. When the Choice Architect removes two options, the

estimated coefficient is even larger, but the increment is only marginally significant ($p = 0.11$).[15]

One purpose of the *Exogenous Restrictions* condition, which we mentioned briefly in Section IB, is to allow us to study beliefs about the welfare effects of prespecified restrictions whether Choice Architects select them. Here we examine two of the four rounds in this condition (see Section IIE for an explanation of the other two rounds). There are two main differences between the welfare evaluations in these rounds and those in the Main condition. First, they involve comparisons between the unrestricted set and an exogenously specified subset that includes either the most patient option or the two most patient options. The number of items in the comparison set (one or two) is determined randomly for each subject and is the same in the both rounds. Second, this condition employs menu 6 (see Table 1).[16]

Because exogenously imposed restrictions need not coincide with the interventions Choice Architects deem best for Choosers, we would expect to find less optimism about their welfare effects. Indeed, as shown in top half of column 2 in Table 3, Choice Architects say that the restrictions make Choosers better off 35 percent of the time and worse off 55 percent of the time. Of course, these figures include libertarians. Non-libertarian Choice Architects (who, by definition, withhold some option in the Main treatment) say that the restrictions make Choosers better off 47 percent of the time and worse off 42 percent of the time. Moreover, even non-libertarians may view the removal of two options as excessive; indeed, in the Main treatment, they remove 0.86 options per round on average. Non-libertarian Choice Architects say that removing the most impatient option, and nothing else, makes the Chooser better off 52 percent of the time and worse off 35 percent of the time (not shown in the table). On average, these subjects clearly believe that a minimal restriction helps: the difference in frequencies is 19 percentage points ($p = 0.078$).

We report associated regression results in the bottom half of column 2 of Table 3. We use OLS as a convenient method for calculating conditional means, but do not interpret the estimates as causal effects. The regression relates welfare evaluations of exogenous restrictions, calibrated as before, to a proxy for the Choice Architect's proclivity to intervene (the average number of options removed in the Main condition), and a dummy variable indicating that the comparison set is limited to the most patient item. Once again, we estimate the regression with the full sample, control for altruism versus spite on the part of the Choice Architect, and include session and order effects. Two conclusions follow from this regression. First, Choice Architects who have demonstrated a greater proclivity to restrict Choosers in the Main condition are substantially more likely to believe that exogenously specified restrictions

[15] We also find that spiteful Choice Architects believe much less strongly that their restrictions benefit Choosers. We caution, however, that this estimate is based on only nine spiteful subjects.

[16] For the purpose of examining Choice Architects' beliefs about welfare effects, the two rounds are essentially identical, so it is especially important to report standard errors clustered at the subject level reported in Table 3. The distinction between these rounds is that we provide the Choice Architect with a different default when asking her to construct the Chooser's opportunity set: in one round, the default consists of all three options, and the Choice Architect can remove either the least patient option or the least patient and middle options; in the other round, the default consists of the most patient option, and the Choice Architect can add either the middle option or the middle and least patient options. We examine Choice Architects' selected opportunity sets for these rounds, as well as for the other two rounds of the Exogenous Restrictions condition, in Section IIE.

are beneficial. Second, Choice Architects are more likely to believe that a restriction is beneficial if it removes one item rather than two.

Columns 3 and 4 replicate columns 1 and 2 using dependent variables based on our second measure of welfare effects: beliefs about compensating variations, elicited through an incentive-compatible procedure. Here we estimate interval regressions, including only those subjects whose choices in the multiple-decision lists are consistent with a preference for increasing the Chooser's payment.[17] Results for evaluations of exogenous restrictions corroborate our findings based on unincentivized evaluations (compare column 4 to column 2). Naturally, magnitudes differ because the scales of the dependent variables are not comparable. Results for evaluations of chosen restrictions are consistent with our findings based on unincentivized evaluations (compare column 3 to column 1), but the coefficients are not statistically distinguishable from zero at conventional levels of confidence. This loss of statistical precision may reflect the complexity of the incentive compatible elicitation, which could introduce noise. While this evidence is correlational, Section IID shows that exogenously varying information about Choosers alters beliefs about welfare effects. Moreover, Section IIE provides evidence against the hypothesis that causality flows from choices to beliefs through a rationalization mechanism.

Subjects' own explanations for the restrictions they impose are consistent with our interpretation. At the end of the survey, we asked subjects why they constructed choice sets as they did. Subjects who withheld options typically cited paternalistic motives. Examples include, "*I wanted to force the future experiment participant to make the decision that is right for him*," "*A discount factor of 1/5 is very extreme. Other options offer a justifiable discount factor*," and "*I wanted to protect him from himself*." Subjects who did not withhold options, in contrast, often described libertarian reasoning: "*The more options*, *the better*," and "*I will not dictate anyone's choice*, *even though I have a strong opinion*."

Overall, we conclude that, for the most part, Choice Architects believe the restrictions they impose are beneficial, and therefore these interventions are consistent with paternalism. However, it does not necessarily follow that Choice Architects intervene *because* of this belief. Additional evidence, provided in subsequent sections, helps to bridge this gap. First, we show in Section IID that interventions are sensitive to the provision of information that changes beliefs about the benefits of restrictions. Second, we provide evidence in Section IIE that causality does not flow from interventions to beliefs through a rationalization mechanism.

C. *Do People Intervene to Correct for "Present Bias"?*

So far, we have seen that many people intervene with benevolent intentions to prevent impatient choices that favor immediate payments over delayed rewards. These findings are consistent with the objectives that behavioral economists who believe decision-makers suffer from *present bias*, defined as the tendency to favor immediate gratification excessively, commonly attribute to benevolent planners.

[17] Data from the multiple decision lists are sometimes censored at the smallest and largest amounts specified in the lists. In column 3 of Table 3, 61 (7.04 percent) observations are censored above and 49 (5.66 percent) are censored below. In column 4, these figures are 25 (4.36 percent) and 99 (17.25 percent), respectively.

Consequently, the next step in our analysis is to ask whether the alleviation of perceived present bias is in fact a central objective of the observed interventions.

As is well known, introducing front-end delay (that is, delaying all payments by a fixed amount of time) induces people to choose more patiently for themselves (Frederick, Loewenstein, and O'Donoghue 2002). Under the hypothesis that people suffer from present bias, this effect occurs because front-end delay removes the lure of immediacy that is responsible for excessively impatient choice. Thus, if people believe others suffer from present bias, front-end delay reduces the perceived benefits of paternalistic intervention. Indeed, under the particular positive and normative assumptions behavioral economists commonly invoke, front-end delay eliminates those benefits entirely.[18] Thus, if Choice Architects share the perspectives that behavioral economists often attribute to benevolent planners, if they indeed seek to correct for perceived present bias, the introduction of front-end delay should reduce or eliminate interventions. As we will see, our data reject this hypothesis in favor of the view that Choice Architects simply disapprove of impatient choices.

Focusing on the decisions Choice Architects make for themselves in Stage 2 of the laboratory session, we begin by replicating the usual finding that front-end delay increases patience. Column 1 of Table 4 presents a regression of the early payment the Choice Architect selects for herself on an indicator for front-end delay. The regression employs all of the decisions for menus 1 and 2 in the Main condition. Recall that each Choice Architect makes one of these two decisions, selected at random, with front-end delay. We include session, order, and menu fixed effects, and cluster standard errors by subject. As expected, the estimates show that the addition of front-end delay yields a statistically significant decrease in the selected early payment (€0.31, which corresponds to 5.6 percent of the greatest possible treatment effect).

Next, we ask whether Choice Architects anticipate Choosers' responses to front-end delay. Recall that, in Stage 2 of the laboratory component, we elicit Choice Architects' beliefs about the selections Choosers would make from all unrestricted menus. To generate the results in column 2 of Table 4, we regress the mean of the early payment for this elicited distribution on the same set of variables used in column 1. The results show that Choice Architects expect unrestricted Choosers to exhibit more patience with front-end delay, although the effect is slightly smaller than the impact on their own choices (€0.23).

How does front-end delay affect mandates? Figure 4 shows the frequency with which Choice Architects withhold each type of choice, averaged over all rounds in the Main condition that involve menus 1 and 2 (half of which include front-end delay). If anything, the interventions are slightly *more* common with front-end delay, not less. Column 3 of Table 4 makes the same point in a regression format. For this purpose, we define a Choice Architect's *mandate* as the largest early payment she allows the Chooser to receive. For example, if a Choice Architect offers the most

[18] Most of the pertinent literature assumes that people have quasihyperbolic preferences, which take the following form: $U_t(x_t, x_{t+1}, \dots) = u(x_t) + \beta \sum_{k=1}^{\infty} \delta^k u(x_{t+k})$, with $\beta, \delta < 1$. A common normative interpretation of this model is that δ captures true intertemporal preferences, while any deviation of β from unity constitutes a bias. In that case, a benevolent planner would try to maximize $\sum_{t=0}^{\infty} \delta^t u(x_t)$ (the long-run criterion). Bernheim and Taubinsky (2018) provide a critical evaluation of the foundations for this criterion.

Table 4—Effects of Front-End Delay and Information about the Chooser

Dependent variables:	€ Choice Architect takes early	Belief € Chooser takes early	Max € Chooser takes early	Max € Chooser takes early	Belief € Chooser takes early
	(1)	(2)	(3)	(4)	(5)
Front-end delay	−0.312	−0.228	−0.104		
	(0.088)	(0.049)	(0.093)		
Chooser information					
Patient, happy				−0.454	−1.480
				(0.084)	(0.061)
Patient, unhappy				−0.049	−1.057
				(0.072)	(0.054)
Impatient, unhappy				−0.312	−0.223
				(0.076)	(0.043)
Mean of dependent variable	0.965	1.929	4.548	3.246	1.521
	(0.098)	(0.062)	(0.082)	(0.058)	(0.045)
Observations	606	606	606	1,212	1,212
Number of subjects	303	303	303	303	303

Notes: Method: OLS. Unit of observation: subject-round pair. Dependent variables: *€ Choice Architect takes early* is the amount of euros the Choice Architect chooses to receive early when choosing for herself in stage 2 of the laboratory component. *Belief € Chooser takes early* is the Choice Architect's belief about the mean amount a Chooser will receive early if allowed to choose without restrictions. *Max € Chooser takes early* is the maximal amount of euros the Choice Architect permits the Chooser to receive early. Controls: *Front-end delay* is an indicator for whether all payments in a menu are delayed by a week. *Chooser information* is a set of dummies that indicate the statement a Chooser selects from Table 5 to describe himself. For columns 4 and 5, the omitted category is *impatient, happy*. Other controls include session, order, and (for columns 1–3) menu fixed effects. Samples: columns 1–3 use menus 1 and 2 from the Main condition; columns 4 and 5 use all rounds of the Chooser Information condition. Standard errors: clustered by subject.

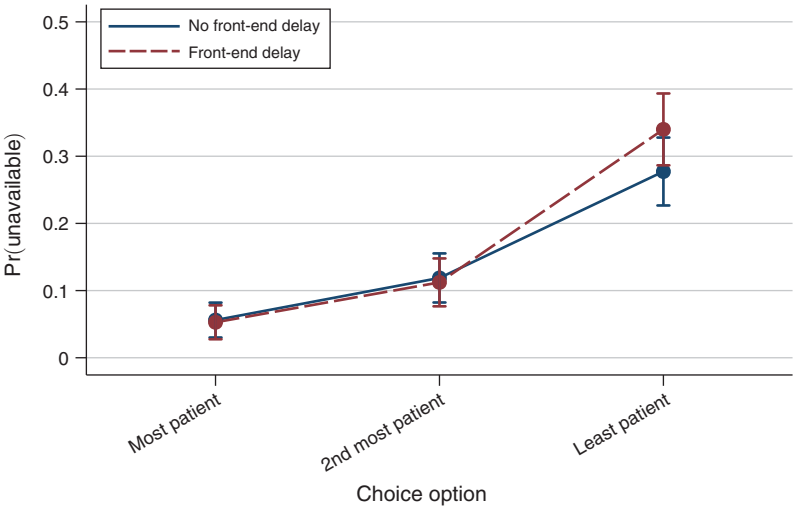

Figure 4. Effect of Front-End Delay

Notes: Whiskers display 95 percent confidence intervals with estimates of standard errors clustered at the subject level. Graph based on all decisions in the Main condition for menus 1 and 2.

patient and middle alternatives from menu 1 of Table 1, the mandate is €3.[19] Here we have regressed mandates for the same rounds on the same set of variables as in columns 1 and 2. The estimated effect of front-end delay is negative, meaning that mandates with front-end delay are more restrictive, not less, although the effect is not statistically significant. Further, Choice Architects are significantly *more* likely to think that their chosen restrictions are beneficial in the presence of front-end delay than without it, as online Appendix Section B.7 shows. Thus, we find no evidence that Choice Architects intervene to counter present bias. Rather, they appear to regard patience as generally virtuous, and intervene to enforce it.

D. *Are Interventions Sensitive to Information about the Affected Party's Inclinations and Subjective Experiences?*

Having ruled out the possibility that our Choice Architects act on the types of objectives and perspectives that behavioral economists often attribute to benevolent planners, we consider a second, simpler hypothesis: paternalists make decisions for others based exclusively on their own conceptions of right and wrong without regard to the preferences or subjective experiences of the affected individuals. As we explain in this section, our data strongly reject this hypothesis. This finding leads naturally to the question we address in Section IIIB: in light of the fact that paternalists' beliefs about others' inclinations and subjective experiences affect their interventions, how do they form those beliefs?

The analysis of this section employs data from the *Chooser Information* condition, which we mentioned briefly in Section IB. It resembles the Main condition, except that we provide Choice Architects with information about their Choosers. Specifically, before making the decisions described in Section ID, each Chooser endorses one of the four statements listed in Table 5, which describe them as either patient or impatient, and as either generally happy or unhappy with their intertemporal choices. The condition consists of four rounds, one for each of the four statements. The Choice Architect knows her Chooser will fall into one of these categories, but does not know which one. Moreover, Choice Architects believe that a substantial fraction of Choosers falls into each of the four categories.[20] Consequently, they have reason to treat their decisions in each round as consequential. All four rounds involve menu 5 (see Table 1).[21]

Figure 5, which displays the frequency of withheld options for each Chooser statement, establishes that Choice Architects respond to this information. Panel A shows interventions affecting Choosers who describe themselves as impatient. Choice Architects impose more patience if an impatient Chooser claims to be unhappy with his impatience than if he claims to be happy with it. Thus, Choice

[19] All our results are robust with respect to alternative definitions of this variable, such as the smallest amount of money the Choice Architect forces the Chooser to receive with delay.

[20] On average, Choice Architects believe that 33.1 percent of Choosers will classify themselves as patient and happy, 21.3 percent as impatient and happy, 16.5 percent as patient and unhappy, and 29.1 percent as impatient and unhappy. We elicited these beliefs either directly before or directly after eliciting beliefs about the distribution of choices by ten unrestricted subjects (randomized on the individual level).

[21] We ask subjects in the role of Choosers to select the statement that describes them best at the start of their session, before they receive any other information. If the Chooser Information condition is selected for implementation, the Chooser receives the choice set the Choice Architect constructed for the corresponding statement.

Table 5—Statements for the Chooser Information Condition

- I am a patient person. I am happy with this. (I often forego things in the present with regard to the future).
- I am an impatient person. I am happy with this. (I rarely forego things in the present with regard to the future).
- I am a patient person. I often regret my decisions. (Perhaps too often, I forego things in the present with regard to the future).
- I am an impatient person. I often regret my decisions. (Perhaps too rarely, I forego things in the present with regard to the future).

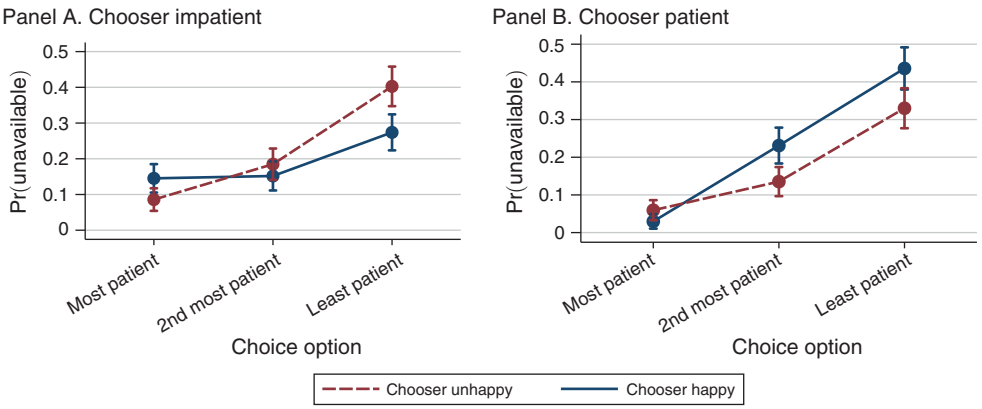

Figure 5. Effect of Information about the Chooser

Notes: Whiskers indicate 95 percent confidence intervals with standard errors clustered by subject. Graph based on data from the Chooser Information condition, in which decisions involve menu 5.

Architects appear to select interventions that ostensibly help Choosers achieve their own goals. For Choosers who describe themselves as patient (panel B), this pattern reverses. Choice Architects impose more patience if a patient Chooser claims to be happy with his patience (perhaps to prevent the Chooser from accidentally selecting impatient options) than if he claims to be unhappy with it.

Column 4 of Table 4 makes these same points in a regression format. The dependent variable is our measure of the Choice Architect's mandate, defined as in the Section IIC. Using data from the Chooser Information condition, we regress mandates on indicators for each of the four Chooser self-descriptions (omitting one), as well as session and order fixed effects, clustering standard errors at the subject level. The results establish that Choice Architects' mandates differ significantly, both economically and statistically, across the four types of Choosers.[22]

[22] Online Appendix Section B.7 shows how Choosers' self-descriptions relate to Choice Architects' beliefs about the welfare effects of chosen restrictions. Interventions continue to be consistent with benevolent motives. The likelihood with which Choice Architects see their interventions as helpful is highest when the Chooser is impatient and generally unhappy with his choices, second highest when the Chooser is patient and generally happy with his choices (possibly because Choice Architects remove options they think Choosers might select by accident), and lowest when the Chooser is either impatient and happy or patient and unhappy (apparently because Choice Architects are disinclined to enforce impatience).

Next, we explore the mechanisms through which the information we provide affects interventions. The regression in column 5 of Table 4 shows how Choice Architects' beliefs about Choosers' unrestricted choices differ across the four self-descriptions. Recall that, in Stage 2 of the laboratory component, we elicit Choice Architects' beliefs about the selections Choosers would make from all unrestricted menus. Here we regress the mean of the early payment for this elicited distribution on indicators for the self-descriptions. The large and statistically significant differences between the four groups establish that the information changes Choice Architects' beliefs about the Choosers' inclinations. However, it is also apparent from the same regression that the information does not affect Choice Architects' interventions *only* by changing beliefs about Choosers' selections: conditional on knowing that the Chooser is unhappy, also knowing that he is patient rather than impatient drives beliefs and mandates in *opposite* directions (compare the pertinent coefficients in columns 4 and 5). The product of the differences in coefficients is negative and highly statistically significant ($p < 0.01$).[23]

E. *Potential Confounds*

Paternalism is not the only possible motivation for intervention. To test for specific spurious motives, we supplemented our experiment with some additional conditions, which we briefly summarize in this section. Finding little evidence of spurious motives, in particular for interventions that Choice Architects regard as helpful to Choosers, we conclude that our results primarily reflect paternalism. To conserve space, details of these analyses appear in online Appendix Section B.1.

The *Induced Chooser Preferences* condition, which we mentioned briefly in Section IB, addresses the concern that Choice Architects may intervene out of a general desire either to keep busy or to exercise active control over Choosers. By deploying Vernon Smith's induced preferences paradigm (Smith 1976), it removes reasons to take issue with the Chooser's objectives while preserving the spurious motives mentioned above, as well as paternalistic motives unrelated to disapproval of preferences, such as the preemption of accidental errors.

In this condition, each choice option consists of a pair of immediate payments, one in euros, the other in an experimental currency that converts to euros at the rate $\hat{\delta}_i$, which is known only to Chooser i. Choice Architects merely know the population distribution of exchange rates. Formally, these decision problems resemble the intertemporal problems in the Main condition: one can think of the two payments as "immediate" and "delayed" (even though we pay both out right after the session) and interpret $\hat{\delta}_i$ as the Chooser's discount factor. The main difference is that, in this setting, the Choice Architect has no basis for disputing the normative validity of the Chooser's $\hat{\delta}_i$, because we have exogenously induced it. In contrast, the use of induced preferences does not remove motivations for intervention stemming from the desire to keep busy or to exercise active control over Choosers. There are two rounds in this condition using, respectively, analogs of menu 3 and menu 4 (see Table 1).

[23] To perform this test, we estimate columns 4 and 5 jointly in a two-equation system OLS-regression with bootstrapped standard errors clustered on the subject level (1,000 samples).

Choice Architects are substantially less likely to withhold options in the Induced Chooser Preferences condition than in the Main condition. Moreover, we do not see a pattern in the Induced Chooser Preferences condition analogous to our finding in the Main condition that Choice Architects become more inclined to limit impatient choices when impatience is more costly. Choice Architects are also far less likely to believe that interventions benefit the Chooser in the Induced Chooser Preferences condition. Specifically, Choice Architects both restrict choice and describe the restriction as helpful in 26.0 percent in the Main condition (excluding the rounds with front-end delay), compared to only 11.7 percent of the rounds in the Induced Chooser Preferences condition. We conclude that the patterns of primary interest in the Main condition likely reflect paternalism arising from Choice Architects' disapproval of impatience rather than spurious motivations.

A limitation of the Induced Chooser Preferences condition is that it cannot rule out a more nuanced hypothesis: Choice Architects may indulge their desire to exercise active control only if they can rationalize their actions as beneficial. A second purpose of the Exogenous Restrictions condition, part of which we discussed in Section IIB, is to allow us to address this possibility.

To understand the logic of our test, consider a Choice Architect who would have constructed the two-option opportunity set had the default consisted of the three-option set. If we remove the least patient option exogenously, constructing the two-option opportunity set no longer involves an exercise of active control. To affect the outcome, the Choice Architect must now remove the middle option, which she can potentially rationalize as beneficial on the grounds that patience is virtuous. Thus, if the motives for interventions are spurious, the frequency with which Choice Architects withhold the middle option should rise. Accordingly, this condition includes a round in which the Choice Architect can remove options from a three-option choice set (menu 6), and a round in which we first remove the least patient option in that set exogenously. To account for unrelated boundary effects and noisy choosers, we include two otherwise identical versions of these rounds in which the default opportunity set consists of the most patient option by itself, so that the Choice Architect adds options instead of removing them.[24] These rounds disable the mechanism of concern, because a Choice Architect who enjoys exercising power, but who regards patience as virtuous, cannot rationalize the addition of impatient options. We find no evidence that Choice Architects are primarily motivated by a desire to exert control whenever they can rationalize active interventions as beneficial.

[24] Throughout this condition, when all three options are available, the Choice Architect can withhold either the least patient option, or the least patient and middle options; when we remove the least patient option exogenously, the Choice Architects can withhold the middle option. Thus, the Choice Architects' alternatives are more restricted than in the Main condition. In addition, as discussed in Section III, when all three options are available, elicited beliefs about welfare effects involve comparisons between the three-option set and either the two-option set consisting of the most patient and middle options, or the one-option set consisting of the most patient option. When we remove the least patient option exogenously, these elicited beliefs involve comparisons between the two-option set and the set consisting of the most patient option.

III. Projective Paternalism

We have shown that Choice Architects often withhold options for paternalistic reasons. But how do they decide whether particular options are good or bad for Choosers? By definition, paternalists are hesitant to rely on the judgments implicit in Choosers' decisions, and indeed may even question whether Choosers are aware of their own best interests. And yet we have also seen in Section IID that paternalists are sensitive to Choosers' preferences and subjective experiences. While Choice Architects evidently find this information relevant, we require them to make decisions about interventions without knowing much about the Choosers, as is often the case in real-world applications. In this section, we seek to clarify how Choice Architects arrive at their evaluations in such situations.

A. *Conceptual Framework*

To draw sharp inferences concerning the mechanisms governing Choice Architects' beliefs about Choosers, we interpret their behavior through the lens of a formal theory. Here, we develop a model of the Choice Architect's problem, precisely define various paternalistic mechanisms, and then identify their distinctive empirical signatures so we can distinguish among them.

Setting.—There are two agents, a Choice Architect (she) and a Chooser (he). Potential consumption opportunities for the Chooser are indexed by a real number $c \in \mathbb{R}$. The Choice Architect can set a lower bound (*mandate*), $\underline{r} \in [-\infty, \infty)$, on the set of available alternatives, where we interpret $\underline{r} = -\infty$ as the absence of a mandate. She can also set an upper bound (*cap*), $\bar{r} \in (-\infty, \infty]$, where we interpret $\bar{r} = \infty$ as the absence of a cap. The Chooser then selects an option in $[\underline{r}, \bar{r}]$. A surrogate choice is one in which the Chooser sets these bounds subject to the restriction that $\underline{r} = \bar{r}$; we use s to denote this common value. For our intertemporal choice application, the index captures the balance between earlier and later payments, and the Choice Architect's selection of restrictions determines the most and least patient choices a Chooser may make.

The Choice Architect's Beliefs about Choosers: Suppose the Choice Architect believes the Chooser would select c from the unrestricted set $\mathbb{R}$, but that option u would be best for him. We refer to u as the Chooser's *ideal choice*, and to $m = u - c$ as the Chooser's *mistake*: in each case, *according to the Choice Architect's subjective judgment*, which may or may not coincide with (i) the Chooser's judgment, (ii) the Choice Architect's belief about the Chooser's judgment, or (iii) any external judgment.[25]

The Choice Architect has imperfect information about the Chooser, and her beliefs about the Chooser's ideal choices and mistakes are given by a cumulative distribution function $F(u, m)$. Next, we describe some additional assumptions concerning these beliefs.

[25] For the intertemporal choice setting, the reader may be tempted to think of u and m in terms of quasi-hyperbolic discounting and the long-run criterion, but it is important to remember that our Choice Architects do not generally embrace that perspective (see Section IIC). These variables reflect whatever normative judgments they actually apply.

In many domains, it is reasonable to assume that perceived mistakes are unidirectional. Concerning intertemporal choice, there appears to be a widespread perception that people struggle to act sufficiently patiently rather than sufficiently impatiently. Indeed, as we saw in Section IIA, the dominant pattern by far in our experiment is for subjects to impose patience rather than impatience. Accordingly, letting P denote the probability measure induced by F, we impose the following assumption:

ASSUMPTION 1: $\Pr(m \geq 0) = 1$.[26]

This assumption involves a sign convention: the index of the ideal option is always at least as large as that of the chosen option. Thus, for our intertemporal choice setting, higher values of the index, c, denote greater patience, while $\underline{r}$ and $\bar{r}$ denote, respectively, the lowest and highest degrees of patience the Choice Architect permits.[27]

We also make some regularity assumptions concerning the Choice Architect's beliefs. To facilitate the statement of these assumptions, we decompose F into two components, as follows. We assume the Choice Architect believes the fraction γ of Choosers make normatively valid choices: m is identically equal to 0 and the CDF of u is given by $Q(u)$. Furthermore, we assume the Choice Architect believes the fraction $1 - \gamma$ of Choosers make normatively flawed choices: m is strictly positive and the CDF of (u,m) is given by $H(u,m)$. Thus, $F(u,m) = \gamma I_{m\geq 0} Q(u) + (1-\gamma)H(u,m)$, where $I_{m\geq 0}$ equals 1 if $m \geq 0$ and 0 otherwise.[28]

With this structure in place, we can state our technical assumptions.

ASSUMPTION 2: *Q and H are atomless CDFs with well-defined densities, and with* $\text{supp}(Q) = (-\infty,\infty)$ *and* $\text{supp}(H) = (-\infty,\infty) \times [0,\infty)$. *Furthermore, the mean of u is finite for both Q and H, and the mean of m is finite for H.*

Objective Function: The Choice Architect constructs the Chooser's opportunity set to maximize her perception of the Chooser's expected welfare. Specifically, the Choice Architect believes that if a Chooser's selection c differs from his ideal u, he sustains a welfare loss of $-l(c-u)$. We assume that $l(z)$ is differentiable with $l(0) = 0$, and that z and $l'(z)$ have opposite signs, so that $z = 0$ maximizes $l(z)$. We also assume for technical simplicity that $|l'|$ is bounded above, and that it is bounded away from zero outside any neighborhood of 0.[29]

We can write the Choice Architect's objective function as follows:

$$W(\underline{r},\bar{r}) = \int l\big(\varphi_{u,m}(\underline{r},\bar{r}) - u\big)\, dF(u,m), \tag{1}$$

[26] Our results continue to hold when Assumption 1 is violated as long as $\Pr(m < 0)$ is not too large.

[27] In a setting involving exercise, where the concern is that mistakes generally involve inactivity, higher values of c would correspond to more frequent and vigorous workouts. In a setting involving nutrition, where the concern is that mistakes generally involve excessive quantities of unhealthy foods, higher value of c would correspond to healthier consumption.

[28] Studies in Behavioral Public Economics often employ the parallel assumption that the population includes a nonnegligible group of "rational" consumers along with "behavioral" consumers (see Bernheim and Taubinsky 2018). Here our assumption concerns the Choice Architect's beliefs about Choosers, rather than their actual characteristics.

[29] This assumption guarantees the existence of various integrals used in the proofs of our formal results. We can relax the assumption as long as the existence of these integrals is maintained.

where $\varphi_{u,m}(\underline{r}, \bar{r})$ denotes the Choice Architect's belief about the selection a Chooser of type (u, m) will make when choosing from the restricted set $[\underline{r}, \bar{r}]$. We assume the Choice Architect believes the Chooser will select $c = u - m$ if that option is available, and the next closest available option otherwise. Accordingly, $\varphi_{u,m}(\underline{r}, \bar{r}) = \max\{\min\{c, \bar{r}\}, \underline{r}\}$.

It would be straightforward to add a term, $\kappa(\underline{r}, \bar{r})$, to the objective function for the purpose of capturing the Choice Architect's feelings (either positive or negative) about restricting the Chooser's options, other than those arising from anticipated effects on the Chooser's welfare. For simplicity we focus on the case of $\kappa \equiv 0$.[30]

Our results concern the comparative statics governing the optimal restrictions $[\underline{r}^*, \bar{r}^*]$, defined as the values that maximize expression (1), as well as the optimal surrogate choices s^*.

The Choice Architect's Own Choices: Our empirical analysis focuses in large part on the relation between the choices Choice Architects make for themselves and the restrictions they impose on others. Accordingly, let c^A denote the selection the Choice Architect makes for herself when faced with the same decision problem as above. She believes that option u^A would be ideal. If $u^A \neq c^A$, she acknowledges that her actual choice c^A involves a mistake $m^A = u^A - c^A$.

Our next assumption, which concerns the distribution of types among Choice Architects, will eventually allow us to connect Choice Architects' own choices, c^A, to their mandates, $\underline{r}^*$. Formally, define $G_{m^A}(\cdot\,|c^A)$ as the marginal distribution of mistakes, and $G_{u^A}(\cdot\,|c^A)$ as the marginal distribution of ideals, among Choice Architects who choose c^A for themselves. We say that a CDF Γ for a variable x is increasing in a parameter θ if, for all $\theta' > \theta''$, $\Gamma(\cdot\,|\theta')$ first-order stochastically dominates $\Gamma(\cdot\,|\theta'')$.

ASSUMPTION 3: *$G_{u^A}(\cdot\,|c^A)$ is increasing in c^A and $G_{m^A}(\cdot\,|c^A)$ is decreasing in c^A.*

Intuitively, given the identity $c^A = u^A - m^A$, a higher value of c^A tends to indicate that u^A is higher and m^A is lower. Formally, the assumption rules out a strong positive correlation between ideals and mistakes among Choice Architects.[31]

[30] Our formulation of the Choice Architect's problem, together with the assumption $\kappa = 0$, rules out three types of motives that could affect interventions. First, we exclude the possibility that Choice Architects intervene due to a desire to exert control, which we would represent by assuming that $\kappa(\underline{r}, \bar{r}) > 0$ for $\underline{r} \neq -\infty$ and/or $\bar{r} \neq \infty$. This assumption is consistent with the experimental results in Section IIE. Second, we abstract from pure libertarianism, which we might otherwise model by assuming that, for all restrictions, κ is negative and extremely large in magnitude for some fraction of the population. Assuming this concern for autonomy is uncorrelated with other aspects of preferences, libertarians and non-libertarians will behave identically when forced to make surrogate choices for others, consistent with our experimental results in Section IIIB. Third, in our model, the Choice Architect is concerned only with the Chooser's outcome, not with his choice process. For example, Choice Architects do not restrict Choosers' opportunity sets to save them the cognitive costs of making more complex choices, or to spare them the disutility associated with overcoming temptation (as in Thaler and Shefrin 1981 or implicitly in Gul and Pesendorfer 2001). Importantly, we structure our experiment in a way that eliminates motives for intervention such as the cognitive costs or temptation experienced by Choosers.

[31] The assumption is satisfied when mistakes and ideals are independently distributed, when they are negatively correlated, and when they are not too strongly positively correlated. If u^A and m^A follow a bivariate Gaussian distribution with variances $\sigma^2_{u^A}$ and $\sigma^2_{m^A}$, respectively, and correlation ρ^A, then Assumption 3 is satisfied if and only if $\rho^A < \min\{\sigma_{u^A}/\sigma_{m^A}, \sigma_{m^A}/\sigma_{u^A}\}$.

The Paternalist's Decision.—We begin with the fundamental trade-off between commitment and flexibility inherent in the paternalist's optimization problem. If the Choice Architect knew everything about the Chooser including his preferences, she would form a view as to which alternative is best for him and then eliminate any potential for mistakes by removing all other options. However, when the Choice Architect is not perfectly informed about the Chooser, any restriction may end up ruling out the best choice. As a result, optimal interventions often leave the Chooser with a degree of autonomy. Moreover, under Assumption 1 (unidirectional mistakes), an upper bound on the choice of c imposes costs on Choosers with high values of u without preventing any mistakes. Accordingly, as shown in the following proposition, while the optimal intervention may involve a mandate ($\underline{r}^* > -\infty$), it never includes a cap (i.e., we always have $\bar{r}^* = \infty$).[32] The proposition also guarantees the existence of optimal mandates and surrogate choices. All proofs appear in online Appendix Section A.

PROPOSITION 1:

(*i*) *A Choice Architect does not impose a cap on the Chooser's options*: $\bar{r}^* = +\infty$.

(*ii*) *A Choice Architect may impose a mandate* $\underline{r}^* > -\infty$ *on the Chooser's options*, *but always leaves the Chooser with some flexibility*: *the set of optimizers is non-empty*, *closed*, *and bounded*.[33]

(*iii*) *A Choice Architect's optimal surrogate choice*, s^*, *is well defined*: *the set of optimal surrogate choices is non-empty*, *closed*, *and bounded*.

We cannot guarantee the uniqueness of the optimal mandate or surrogate choice without additional technical restrictions. In what follows, we assume that the Choice Architect breaks ties in favor of the Chooser's autonomy and thus selects the smallest maximizer (which, according to the proposition, is well-defined). The consistent application of other rules, such as selecting the largest maximizer, yields identical conclusions.

Varieties of Paternalism: We distinguish between various modes of paternalism, differentiated by the way in which the Choice Architect's beliefs about Choosers depend on her own type, (u^A, m^A). A *projective paternalist* forms her views about others by extrapolating from her understanding of herself. A Choice Architect exhibits *mistakes-projective paternalism* to the extent she thinks others are similar to her with respect to the mistakes they make. A Choice Architect exhibits *ideals-projective*

[32] For our intertemporal choice problem, given our sign convention, $\underline{r}^*$ is the minimum degree of patience the Choice Architect permits, and hence the maximum amount the Chooser can elect to receive early. Likewise, $\bar{r}^*$ is the maximum degree of patience the Choice Architect permits, and hence the minimum amount the Chooser can elect to receive early.

[33] The existence of residual flexibility hinges on our assumption that the underlying set of feasible options is unbounded. When the set of feasible options is bounded (as in our experiment), it may be optimal for the Choice Architect to eliminate all discretion.

paternalism to the extent she thinks others either have, or should have, ideals similar to her own. Naturally, a projective paternalist can be both mistakes-projective and ideals-projective. In contrast, a *conventional behavioral welfarist* acts as a benevolent social planner whose information about others may be incomplete, but who has accurate information about the underlying population distributions of key decision parameters, as envisioned in many standard behavioral welfare analyses (see, e.g., Bernheim and Taubinsky 2018). She does not consider her own inclinations when intervening in others' choices.

To define these concepts precisely, we specialize to a semi-parametric representation of the Choice Architect's beliefs. Specifically, we assume that the distribution of beliefs over (u, m) is bivariate Gaussian, with parameters that depend on the Choice Architect's characteristics, as specified below. To maintain Assumption 1, we truncate m at zero.[34]

ASSUMPTION 4: *Let* $(u, z_m) \sim \mathcal{N}(\mu_u, \mu_m, \sigma_u, \sigma_m, \rho)$. *Then* $(u, m) = (u, \max\{0, z_m\})$.

This assumption yields a mass point at $m = 0$ that corresponds to the share γ of Choosers who the Choice Architect believes behave according to their ideals. Thus it implies specific functional forms for Q and H. We now turn to our formal definitions.

DEFINITION 1:

(*i*) *Choice Architects practice* mistakes-projective paternalism *if* μ_m *is strictly increasing and* μ_u *is weakly increasing in* m^A.

(*ii*) *Choice Architects practice* ideals-projective paternalism *if* μ_u *is strictly increasing and* μ_m *is weakly increasing in* u^A.

(*iii*) *A Choice Architect is a* conventional behavioral welfarist *if* μ_u *and* μ_m *are the actual population means* (*and hence independent of the Choice Architect's type*).

Each of the first two definitions involves a *direct effect* and a *compensatory effect*. The direct effects are intuitive: to the extent Choice Architects assume others either share or should share their ideals, we would expect μ_u to be strictly increasing in u^A (and similarly for mistakes). To understand the compensatory effects, imagine Choice Architects have sufficient information about others' behavior to correctly identify the population average of unrestricted choice (i.e., the mean of c, μ_c). Then, if an increase in u^A produces an increase in μ_u, it must also produce a fully compensating increase in μ_m. Likewise, with imperfect information about μ_c, one would

[34] It is straightforward to generalize all of our results to a setting with two jointly Gaussian latent characteristics, z_u and z_m, such that u is a function of z_u, $u(z_u)$, and m is function of z_m, $m(z_m)$, where u and m are increasing in their arguments, in each case strictly so on some open interval. This approach allows us to accommodate our assumption that $\Pr(m \geq 0) = 1$ in a variety of ways without ruling out natural distributional forms. For example, as an alternative to truncation, one could then assume that $m(z_m) = \alpha \exp(z_m)$.

expect a partially compensating increase in μ_m. Similar considerations apply to the inferences about others that Choice Architects draw from their own mistakes.[35]

Testable Implications: Our main theoretical result uses Assumption 3 to show that each mode of paternalism has a distinctive empirical signature involving the sign of the correlation between, on the one hand, the selections that Choice Architects make for themselves, and on the other hand, their mandates and surrogate choices.

PROPOSITION 2:

(*i*) *Through the mistakes-projective mechanism, the distributions of the optimal mandate* $\underline{r}^*$ *and optimal surrogate choice* s^* *conditional on the Choice Architect's own selection,* c^A*, are decreasing in* c^A *(weakly in the case of surrogate choices).*[36]

(*ii*) *Through the ideals-projective mechanism, the distributions of the optimal mandate* $\underline{r}^*$ *and optimal surrogate choice* s^* *conditional on the Choice Architect's own selection,* c^A*, are increasing in* c^A*.*

(*iii*) *A conventional behavioral welfarist's optimal mandate* $\underline{r}^*$ *and optimal surrogate choice* s^* *are independent of her own choice* c^A*.*

The intuition for Proposition 2 is straightforward. Under Assumption 3, an increase in c^A generates an upward shift in the distributions of u^A and a downward shift in the distribution of m^A. Concretely, a consumer who acts more patiently probably aspires to greater patience and is less prone to mistakes involving excessive impatience. Through ideals-projective paternalism, the increase in u^A leads to an upward shift in the distribution of u (the main effect), and possibly an upward shift in the distribution of m (the compensatory effect). Both of these effects result in higher mandates: concretely, greater enforcement of patience. Through mistakes-projective paternalism, the decrease in m^A leads to a downward shift in the distribution of m (the main effect) and possibly a downward shift in the distribution of u (the compensatory effect). Both of these effects result in lower mandates: concretely, less enforcement of patience. As a result, greater patience when acting for oneself goes hand-in-hand with the imposition of greater patience to the extent Choice Architects are ideals-projective paternalists, and with the imposition of less patience to the extent they are mistakes-projective paternalists. For a conventional behavioral welfarist, own choices play no role. The intuition for surrogate choices is essentially the same, except that surrogate choice disables effects operating through m.

Accordingly, in the next subsection, we examine the empirical relationships between, on the one hand, the patience that Choice Architects display when choosing

[35] Our notions of projective paternalism assume that the inferences a Choice Architect draws from her own characteristics pertain to the means, μ_u and μ_m, rather than to the variances or correlations, which we treat as invariant across Choice Architects. For the sake of tractability, we abstract from all forms of heterogeneity among Choice Architects other than in u^A and m^A.

[36] Recall that, when we say a distribution is increasing or decreasing in a parameter, we mean in the sense of first-order stochastic dominance.

for themselves and, on the other hand, their mandates and surrogate choices. We ask whether these relationships are consistent with the signature patterns implied by ideals-projective or mistakes-projective paternalism.

B. *Distinguishing between Variants of Projective Paternalism*

The Relationship between Patience and Mandates.—We have seen that mistakes-projection induces a *negative* relation between the options that Choice Architects select for themselves and those they force on others, whereas ideals-projection induces a *positive* relation, and conventional behavioral welfarism implies the absence of a relation. To differentiate between these hypotheses, we study this relationship empirically. First we construct a measure of the degree of patience that the Choice Architect displays when choosing for herself in the experiment's online component. Specifically, we calculate the percentile rank of the number of months she is willing to delay the receipt of the larger payment averaged over the six decision lists. Measuring patience in this way enables us to avoid assumptions about the structure of Choice Architects' intertemporal preferences. To avoid ambiguity, we focus on the 291 (of 303) Choice Architects who respected monotonicity in all of these lists.[37] Next, we relate this measure of patience to Choice Architects' interventions in the Main condition, using the *Mandate* variable defined in Section IIC (i.e., the maximum amount that the Choice Architect permits the Chooser to receive immediately).

Figure 6, which excludes libertarian subjects (38 percent of our sample), depicts our main result: those who have chosen more patiently for themselves in the online component of our experiment also impose significantly more patience on Choosers. The most patient non-libertarian Choice Architects' mandates are stricter than those of the least patient Choice Architects by about €1. This difference is just under two-thirds of the standard deviation of mandates (€1.64) across this population, and one-fifth of the greatest possible effect (€5) in our experiment. According to Proposition 2 in Section IIIA, this finding is consistent with the signature pattern associated with ideals-projective paternalism.

We formalize these observations by regressing Choice Architects' mandates in the Main condition on their patience percentiles. The unit of observation is a single decision in a single round. We control for session, order, and menu fixed effects, and cluster standard errors by subject. For the regression in column 1 of Table 6, we exclude libertarians. As Figure 6 suggests, we find that the mandate tightens by €1.04 as patience rises from the bottom of the population distribution to the top, and the effect is highly statistically significant. Column 2 shows that the effect is smaller (€0.61) but still highly statistically significant once all subjects, including the libertarians, are included. The estimated effect is attenuated because libertarians never intervene, and because a Choice Architect's patience percentile does not predict whether she is libertarian (see column 3, which reports a subject-level regression of a libertarian indicator variable on the Choice Architect's patience percentile).

[37] The 3.97 percent of subjects with multiple switches is low compared to other studies using multiple-decision lists, such as Holt and Laury (2002).

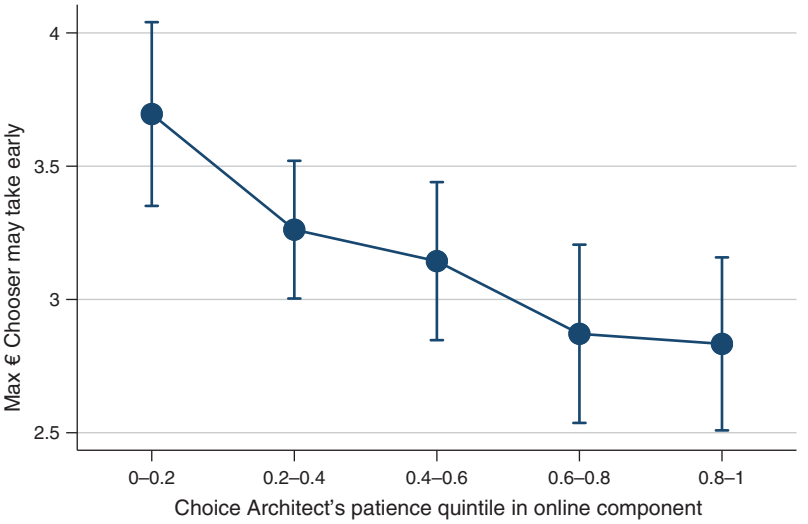

FIGURE 6. IDEALS-PROJECTIVE PATERNALISM

Notes: Dependent variable: maximum amount of money the Chooser is allowed to receive immediately as function of the Choice Architects' own patience. Sample: we exclude subjects classified as libertarian. Patience percentiles are unrelated to libertarianism, see Table 6. Whiskers indicate 95 percent confidence intervals.

TABLE 6—RELATIONSHIPS BETWEEN CHOICE ARCHITECTS' PATIENCE AND MEASURES OF THEIR MANDATES, SURROGATE CHOICES, AND WELFARE BELIEFS

Variables:	Mandate	Mandate	Libertarian	Surrogate choice	Surrogate choice	Welfare belief
Non-libertarian subjects only:	Yes			Yes		Yes
	(1)	(2)	(3)	(4)	(5)	(6)
Patience percentile	−1.044	−0.614	0.011	−1.589	−1.655	0.509
	(0.270)	(0.197)	(0.091)	(0.242)	(0.183)	(0.130)
Mean of dependent variable	3.080	3.633	0.385	0.876	0.865	0.285
	(0.091)	(0.067)	(0.029)	(0.088)	(0.069)	(0.044)
Observations	537	873	291	518	837	537
Number of subjects	179	291	291	179	291	179

Notes: Method: OLS. Unit of observation: subject-round pairs for column 1, 2, 4, 5, and 6; subjects for column 3. Dependent variables: *Mandate* is the maximum amount the Choice Architect allows the Chooser receive immediately (as in Section IIC). *Libertarian* is a dichotomous variable indicating that the Choice Architect never withholds options. *Surrogate choice* is the amount the Chooser receives immediately if the Choice Architect's surrogate choice is implemented. *Welfare belief* is the Choice Architect's belief about the welfare effect of the chosen restriction (as in Section IIB). Controls: *Patience percentile* is the Choice Architect's percentile rank according to the average number of months she is willing to delay the receipt of the larger payment in the online tasks. Other controls include session, order, and (with the exception of columns 3 and 6) menu fixed effects. Samples: All regressions limited to subjects who responded monotonically to all multiple-decision lists in the online component. Columns 1, 2, and 3 based on rounds from the Main condition without front-end delay. Columns 4 and 5 based on the corresponding surrogate choice rounds from Stage 2 of the laboratory component. The number of observations is smaller for surrogate choices because some of these choices were not recorded in the first two sessions. Column 6 based on the same two rounds of the Exogenous Restrictions condition used in Section IIB. Columns 1, 4, and 6 restricted to non-libertarian subjects. Standard errors: clustered at the subject level.

The Relationship between Patience and Surrogate Choices.—Next, we ask whether Choice Architects differ in their judgments about what is good for Choosers, or merely in their propensities to intervene based on those judgments. To address

this question, we examine Choice Architects' *surrogate choices*. Specifically, as mentioned in Section IB, in Stage 2 of the laboratory session we require Choice Architects to select a single item for the Chooser from each menu encountered previously in the experiment. Here we study each Choice Architect's surrogate choices for all of the Main-condition menus, excluding the one with front-end delay. Because these surrogate choices force Choice Architects to intervene, they directly reveal judgments about what is good for the Chooser. Column 4 shows that among non-libertarian subjects, the relation between Choice Architects' surrogate choices and their patience percentiles is even stronger than the relation between their mandates and their patience percentiles. Thus, the latter relationship reflects differences in judgments about what is good for the Chooser, and not merely differences in the propensity to intervene. Column 5 replicates column 4 using the entire sample. As the coefficient of interest changes only slightly, we infer that libertarians are similar to non-libertarians with respect to their judgments about others, except for their willingness to intervene. The positive relation between surrogate choices and own choices is part of the signature pattern associated with ideals-projective paternalism.

The Relationship between Patience and Welfare Beliefs.—Next, we ask whether patience on the part of Choice Architects is related to their beliefs about the benefits of exogenously specified restrictions. For this purpose, we use data from the Exogenous Restrictions condition, much as in Section IIB. Regressing our measure of beliefs about the welfare effects of restrictions on our measure of the Choice Architect's patience along with session and order fixed effects, we find that more patient Choice Architects are more likely to believe patience-promoting interventions are beneficial (see column 6 of Table 6).[38] Though not a formal implication of Proposition 2, this finding goes hand-in-hand with the tendency for more patient Choice Architects to impose greater patience, and it suggests that the key comparative static is indeed a consequence of paternalism.

The Relationship between Front-End Delay and Mandates.—In Section IIC, we rejected the hypothesis that Choice Architects intervene to control perceived present bias by showing that the introduction of front-end delay increases the value Choice Architects believe Choosers derive from restrictions. According to our estimates, it also increases the restrictiveness of interventions, and while that effect is not statistically significant, we certainly find no evidence that interventions decline. A natural explanation for this initially puzzling phenomenon now emerges: it is simply a consequence of ideals-projective paternalism. First recall that the introduction of front-end delay causes Choice Architects to behave more patiently when choosing for themselves. Perhaps because they are not fully aware of their inconsistencies (as evidence from Fedyk 2017 implies), they behave as if their ideal level of patience, u^A, is higher with front-end delay than without (consistent with our Assumption 3). Projecting this greater level of ideal patience onto others, they impose more restrictive mandates. Any compensatory change in their beliefs about others' susceptibility to mistakes reinforces this tendency.

[38] Using our incentivized measure of negative compensating variation, the corresponding coefficient estimate is 0.191 (SE 0.094).

C. *Are Interventions Misguided?*

The fact that people engage in projective paternalism does not necessarily mean their interventions are objectively misguided. For example, an ideals-projective paternalist could have correct beliefs about the choices others would make if given the freedom to choose, but nevertheless decompose each of those choices into an ideal component and a mistaken component based on her own subjective values. To conclude that projective paternalists err when tightening or loosening constraints on others' choices, we must first demonstrate that their beliefs about others are systematically mistaken, and then confirm that those beliefs impact their interventions.

We begin by demonstrating that Choice Architects suffer from *false consensus bias* with respect to their' beliefs about others' choices. For the regression shown in column 1 of Table 7, the dependent variable is the Choice Architect's beliefs about the mean amount a Chooser will receive immediately if allowed to choose without restrictions from a given menu, calculated from the distribution elicited in the relevant round of Stage 2 during the laboratory session (as in Section IIC). Focusing on the menus presented in the Main condition without front-end delay, we relate this variable to the Choice Architect's own patience percentile, defined as in Section IIIB. As with all other regressions in the table, we exclude libertarian subjects, control for session, order, and menu fixed effects, and present standard errors clustered at the subject level. Striking evidence of the false consensus effect emerges, in the sense that people tend to believe others will behave as they themselves behave. Compared to the most patient Choice Architects, the least patient ones think unrestricted Choosers would elect to receive an additional €1.30 immediately (26 percent of the maximum, €5, that Choosers can receive immediately).

Next, we ask whether Choice Architects' skewed perceptions of others' inclinations infect the restrictions they impose. We have already seen in Section IV that Choice Architects tailor their interventions based on information they receive about Choosers' propensities. As a general matter, one would therefore expect to find a strong relationship between their beliefs about others' unrestricted choices and their chosen restrictions. The hypothesized relationship is precisely what we observe in column 2 of Table 7. For this regression, the dependent variable is the Choice Architect's mandate, defined once again as the maximum amount she allows the Chooser to receive immediately (see Section IIC). Focusing on the menus presented in the Main condition without front-end delay and excluding libertarian subjects, we relate this variable to a measure of the Choice Architects' beliefs about the Chooser's impatience. To maintain comparability of the coefficient estimate to the effect of the Choice Architects' own patience, we use the percentile rank of the inverse of the mean amount she believes the Chooser will receive immediately if allowed to choose without restrictions. We also control for the Choice Architect's patience percentile. Because the Choice Architect's own preferences could drive their interventions directly while also incidentally impacting their beliefs about others through the false consensus effect, the omission of this control variable could potentially induce spurious correlation between the Choice Architect's mandate and her beliefs. The significant negative coefficient for the beliefs percentile implies, as expected, that Choice Architects impose more restrictive mandates when they believe Choosers are more patient. Indeed, Choice Architects' mandates are more strongly related to their

Table 7—Accuracy and Relevance of Beliefs

Dependent variables:	Belief mean (1)	Mandate (2)	Believed relevance (3)	Actual relevance (4)	More restrictive than believed (5)
Patience percentile	−1.272	−0.540	0.011	0.041	0.320
	(0.167)	(0.318)	(0.029)	(0.013)	(0.064)
Belief percentile		−1.136			
		(0.365)			
Mean of dependent variable	1.467	3.100	0.095	0.056	0.209
	(0.062)	(0.089)	(0.010)	(0.004)	(0.023)
Observations	537	537	537	537	537
Number of subjects	179	179	179	179	179

Notes: Method: OLS. Unit of observation: subject-round pairs. Dependent variables: *Belief mean* is the Choice Architect's belief about the mean amount a Chooser will receive immediately if allowed to choose without restrictions (as in Section IIC). *Mandate* is the maximum amount the Choice Architect allows the Chooser to receive immediately (as in Section IIC). *Believed relevance* is the frequency with which the Choice Architect's mandate would bind according to her beliefs about Choosers' unrestricted selections. *Actual relevance* is the frequency with which the Choice Architect's restriction actually binds according to Choosers' unrestricted selections. *More restrictive than believed* is a dichotomous variable indicating whether the actual relevance is greater than the believed relevance. Controls: *Patience percentile* is the Choice Architect's percentile rank according to the average number of months she is willing to delay the receipt of the larger payment in the online tasks. *Beliefs percentile* is the Choice Architects' percentile rank in terms of her beliefs about the Chooser's patience, as measured by the inverse of the mean amount the Chooser will receive immediately if allowed to choose without restrictions. Additional controls include session, order, and menu fixed effects. Samples: Regressions limited to non-libertarian subjects who responded monotonically to all multiple-decision lists in the online component. Column 1 based on rounds from Stage 2 of the laboratory component that elicited Choice Architects' beliefs concerning Choosers' selections for the same menus as in the Main condition (excluding the one with front-end delay). Columns 2–5 based on rounds from the Main condition without front-end delay. Columns 3 and 5 also use data on Choice Architects' beliefs concerning Choosers' selections, gathered in Stage 2 of the laboratory component, in combination with the Choice Architect's chosen restriction to construct the dependent variable. Columns 4 and 5 also use data on unrestricted choices, gathered in Stage 2 of the laboratory component, in combination with the Choice Architects' chosen restrictions to construct the dependent variable. Standard errors: clustered at the subject level.

percentiles ranks according to their beliefs about Choosers' unrestricted selections than to their percentiles ranks according to their own preferences.[39]

So far, we have shown that false consensus bias infects Choice Architects' beliefs about others' choices, and that those beliefs are closely related to their mandates. It follows that systematic bias also infects Choice Architects' decisions about interventions, which means these decisions are objectively misguided, even accepting the Choice Architects' aims. For a stark illustration of this point, see columns 3 through 5 of Table 7. For column 3, the dependent variable is the *Believed relevance* of the restriction the Choice Architect imposes, defined as the fraction of Choosers for whom the Choice Architect thinks the mandate would bind. Specifically, for each menu encountered in the Main condition (excluding the round with front-end delay), we identify the least patient choice the Choice Architect permits, and then calculate the frequency with which the Choice Architect believes unrestricted Choosers would select less patient options according to the subjective distribution

[39] A possible objection to the regression in column 2 is that, because the two percentiles are strongly correlated, the regression may spuriously load their effects on the one that contains less measurement error. Online Appendix Section B.8 displays the results of a two-stage least-squares version of the same regression that addresses this concern. We find that the coefficient of the beliefs percentile more than doubles, and it remains highly statistically significant, while the coefficient of the patience percentile changes sign, declines slightly in magnitude, and is statistically insignificant.

elicited in Stage 2 of the laboratory session. We relate this variable to the Choice Architect's patience percentile and the usual fixed effects. As indicated in the table, these subjects believe, on average, that they force 9.5 percent of Choosers to choose more patiently.

The regression reveals no systematic relation between these probabilities and Choice Architects' own patience: more patient Choice Architects do not intend to impose restrictions that enforce patience with higher frequency. And yet, that is precisely what they do. For column 4, the dependent variable is the *Actual relevance* of the Choice Architect's restriction, defined as the fraction of Choosers for whom the restriction would actually bind. Specifically, for each menu encountered in the Main condition (excluding the round with front-end delay), we identify the least patient option the Choice Architect permits, and calculate the frequency with which unrestricted individuals choose less patiently than that.[40]

Because more patient Choice Architects impose more stringent mandates, an increase in the Choice Architect's patience percentile significantly increases the fraction of Choosers affected by the selected restriction. As a result, more patient Choice Architects are substantially more likely to underestimate the restrictiveness of their interventions. A simple tabulation reveals that 29.5 percent of Choice Architects in the most patient (non-libertarian) quartile underestimate the restrictiveness of their mandates, compared with only 9.0 percent of those in the least patient quartile. This conclusion also emerges from the regression in column 5, which employs a dichotomous dependent variable indicating whether the actual effect of the Choice Architect's restriction is greater than she believes. The estimated coefficient for the Choice Architect's patience percentile is highly statistically significant, which means that greater patience implies a significantly greater propensity to underestimate the restrictiveness of chosen mandates.[41]

Because Choice Architects' biased beliefs influence their interventions, it is natural to wonder whether informational policies could correct these errors and thereby improve the interventions, at least from the Choice Architects' perspectives. We address this question in online Appendix Section B.4. Consistent with an existing literature that documents the resistance of the false consensus effect to informational interventions (Krueger and Clement 1994; Engelmann and Strobel 2012), we find that neither the Chooser-specific information provided in the Chooser Information condition, nor information about the distribution of unrestricted choices, attenuates the false consensus bias or ideals-projective paternalism.

[40] For the actual unrestricted choice frequencies, we use the distribution of selections Choice Architects make for themselves from each of the pertinent menus in Stage 2 of the laboratory session (see online Appendix Section B.9). Because there is only one Chooser for every four Choice Architects and each Chooser makes a single unrestricted choice, estimated frequencies based on Chooser data would be noisy. Significantly, we select Choice Architects and Choosers from the same population, and Choice Architects are aware of this fact.

[41] While we are mainly interested in the effect of patience on misperceptions of restrictiveness, it is worth noting that most of our Choice Architects actually overstate the restrictiveness of their chosen mandates. We interpret this finding with caution, however, due to the well-known tendency for elicited beliefs to be biased toward uniformity. Here, average beliefs about the distribution of choices across types of options (56 percent, 25 percent, and 19 percent for the most, second-most, and least patient options, respectively) are more uniform than the actual frequencies (78 percent, 14 percent, and 8 percent), and this pattern is responsible for the apparent prevalence of overstated restrictiveness.

IV. Support for Paternalistic Policies

We conclude this investigation by showing that subjects' judgments about real-world paternalistic policy proposals relate meaningfully to the decisions they make as Choice Architects in our experiment. Additionally, we demonstrate that ideals-projective paternalism extends to the policy domain.

To this end, in Stage 3 of the experiment, we ask subjects to rate four policy proposals involving taxes on sugary drinks, alcohol, and tobacco, as well as restrictions on short-term, high-interest loans. Because our subjects live in Germany, we focus primarily on tax policies for Switzerland, which should purge the ratings of personal interests, at least in principle. We ask subjects to assume the tax policies would be budget-neutral, so responses do not reflect general attitudes about the size of government. For each policy, we elicit the extent to which the subject supports or opposes its implementation.[42] We also elicit beliefs about the impact of each policy on the welfare of the average citizen.[43]

After subjects express these judgments and answer additional non-incentivized questions, they provide information about themselves that relates to the impacted activities. These variables allow us to distinguish between ideals-projective paternalism and mistakes-projective paternalism in the domain of real-world paternalistic policy proposals. Specifically, we elicit subjects' body mass index,[44] their average alcohol consumption, their frequency of binge drinking (defined as the consumption of five or more units of alcohol for men, or four or more units for women, within a two-hour period),[45] their cigarette consumption, and their experience with short-term, high-interest loans. In addition, subjects provide information about their credit card debt in the online portion of the experiment (see online Appendix Section D.1 for details).[46]

Relation of Policy Support to Laboratory Behavior.—For each policy proposal, panel A of Table 8 lists the fractions of subjects who express a given level of support or opposition. Support outweighs opposition in each case. It is particularly pronounced for increased tobacco taxes and sugary drinks taxes.

To examine the relation between these judgments and the decisions subjects make as Choice Architects in our experiment, we regress measures of their support for the policies on the average mandates they impose on Choosers. In the interests

[42] Concerning tax policies, response options are that Switzerland should "definitely not introduce such a tax," "probably not introduce such a tax," "probably introduce such a tax," or "definitely introduce such a tax." Because small, short-term, high-interest loans are not available in Switzerland (possibly due to a lack of demand), questions regarding lending restrictions pertain to Germany. The response options are that the market for short-term loans in Germany should "be greatly restricted," "be somewhat restricted," "remain unchanged," "be somewhat liberalized," or "be greatly liberalized." We encode these options as 2, 1, 0, −1, and −2, respectively.

[43] The question about alcohol taxes concerns adolescents and young adults rather than average citizens, but is otherwise identical.

[44] Subjects can click a button to open a window that asks them to enter their height h in cm and weight w in kg. The window then displays their body mass index using the formula $\text{BMI} = w/(h/100)^2$.

[45] The experiment defines a unit of alcohol as 0.2 liters of beer, 0.1 liter of wine, or 1 shot of schnapps or liquor.

[46] Misreporting of these characteristics is a potential concern. For example, people tend to over-report height and under-report weight (Gorber et al. 2007). A strong correlation remains, however, between reported and measured BMI (Nawaz et al. 2001). Because our interest centers on the signs of correlations, underreporting of BMI does not qualitatively affect our conclusions. Similar statements apply to self-reported alcohol consumption (Sobell and Sobell 1995) and self-reported cigarette smoking (West et al. 2007).

Table 8—Experimental Decisions and Support for Real-World Paternalistic Policies

Policy:	All	Increase alcohol tax	Increase tobacco tax	Introduce sugary drinks tax	Tighten restrictions on short-term lending
	(1)	(2)	(3)	(4)	(5)
Panel A. Summary statistics for dependent variables					
Distribution of support					
Strongly opposed (−2)	—	0.154	0.055	0.087	0.060
Weakly opposed (−1)	—	0.305	0.102	0.196	0.191
Neutral (0)	—	—	—	—	0.342
Weakly in favor (1)	—	0.328	0.280	0.362	0.278
Strongly in favor (2)	—	0.213	0.563	0.355	0.129
Panel B. Relation between laboratory choice and policy attitudes					
B.1 Dependent variable: support for policy proposal					
Average mandate imposed on	−0.113	−0.129	−0.126	−0.113	−0.086
Chooser (max. € paid immediately)	(0.041)	(0.066)	(0.051)	(0.066)	(0.052)
Observations	403	403	403	403	403
B.2 Dependent variable: beliefs about welfare effect of policy proposal					
Average belief exog. smaller choice set better for Chooser	0.185	0.220	0.235	0.210	0.075
	(0.054)	(0.086)	(0.084)	(0.076)	(0.075)
Observations	403	403	403	403	403
Panel C. Projective paternalism with actual policies					
Alcohol consumption					
Alcohol units/week		−0.050			
		(0.034)			
log(days binge drinking/year)		−0.165			
		(0.055)			
Tobacco consumption					
Smoker yes/no			−0.881		
			(0.284)		
Cigarettes/day			−0.046		
			(0.039)		
Body mass index				−0.062	
				(0.028)	
Debt					
Credit card debt (in €1,000)					−0.573
					(0.114)
Other short-term debt yes/no					0.101
					(0.388)
Controls	—	Yes	Yes	Yes	Yes
Observations	—	403	403	398	351

Notes: Method: OLS. Unit of observation: subject. Dependent variables: *Support for policy* proposal is the support expressed for various policies, coded −2 (strong opposition) to 2 (strong support), averaged over policies for "All." A "neutral" response was possible only for the question about short-term lending. Because we asked subjects about *loosening* restrictions on short-term lending, we reverse-coded these responses for easier comparability (so that higher values correspond to greater support for tightening restrictions). *Beliefs about welfare effects* is the belief about the welfare effect of various policies, coded −2 (significantly worse off) to 2 (significantly better off), averaged over policies for "All." Controls: *Average mandate imposed on Chooser* is the maximum amount the Choice Architect allows the Chooser receive immediately (as in Section IIC), averaged over rounds involving menus 3 and 4 in the Main condition. *Average belief exog. smaller choice set better for Chooser* (as in Section IIB) encodes whether the Choice Architect considers the smaller opportunity set better, equally good, or worse for the Chooser than the unrestricted set, coded as 1, 0, and −1, respectively, averaged across the four rounds of the Exogenous Restriction Condition. All regressions control for gender, age, self-reported political attitudes, log monthly expenses, high school GPA, university faculty at which the subject's major field of study is offered, as well as for session and order fixed effects. Regressions in panel B in addition control for weekly alcohol consumption, log days of binge drinking per year (defined as the consumption of at least 4 units of alcohol for females, 5 for males, within a period of two hours; National Institutes on Alcohol Abuse and Alcoholism 2018), smoking status, number of cigarettes smoked per day, body mass index, credit card debt, and for having taken a short term loan. For variables measured with interval precision, we use midpoints for analysis. For control variables that subjects chose not to disclose, the regressions in panel B impute population means, and include indicators for whether a variable's value was missing. Samples: includes 303 subjects who participated in the main experiment, plus 100 subjects who participated in the Choice Distribution Information Condition (see online Appendix Section B.4). Regressions in panel C exclude subjects who chose not to disclose the personal characteristic of interest. Standard errors: clustered by subject.

of obtaining easily interpretable coefficients, we use OLS, and defer ordered probit estimates to online Appendix Section B.10. We encode strong opposition as -2, weak opposition as -1, weak support as 1, and strong support as 2.[47] We relate these measures of policy support to the Choice Architects' mandates (i.e., the maximum amount the Choice Architect allows the Chooser receive immediately, as in Section IIC), averaged over rounds involving menus 3 and 4 in the Main condition.[48] Other controls include gender, age, field of study, high school GPA, monthly expenses, political orientation, and personal behaviors related to each of the policy domains listed above, as well as session and order fixed effects.[49] Panel B.1 in Table 8 displays the results. Focusing first on average support for the four paternalistic policies (column 1), we find that the coefficient of interest is negative and statistically significant ($p < 0.01$). Thus, subjects who impose tighter restrictions on Choosers in the laboratory also express greater support for paternalistic policies in the real world. We also perform separate regressions for each of the policy proposals and report the key coefficient estimate from each in columns 2 to 5. In each case, we find that Choice Architects who enforce more patience in the laboratory express greater support for actual paternalistic interventions. These relations are significant at the 5 percent level for alcohol and tobacco taxes and at the 10 percent level for the remaining two policies.

Next, we examine the relation between beliefs about the welfare effects of actual paternalistic policies, and those concerning exogenous restrictions on choice in the laboratory. To measure the latter beliefs, we encode the Choice Architects' statements as to whether an exogenously specified restriction would help Choosers, leave them equally well off, or hurt them, as 1, 0, and -1, respectively (as in Section IIB). We average this variable across the four rounds of the Exogenous Restrictions condition. For the regressions shown in panel B.2 of Table 8, we include the same control variables as in panel B.1. On average across the four policies (column 1), subjects who believe choice restrictions are beneficial in the laboratory also tend to believe that paternalistic interventions are beneficial in practice. The same pattern emerges for each of the sin taxes (columns 2–4). For restrictions on short-term lending, we do not find statistically significant effects. Because German universities do not charge tuition, personal debt may be less salient than alcohol, tobacco, and sugary drinks for our student subjects.

While we cannot exclude the possibility that non-paternalistic considerations such as the prevention of externalities factor into subjects' assessments of the four policies, Choice Architects' selections and statements in the laboratory session are not subject to such confounds. Consequently, it is reasonable to conclude from the robust relationships documented in Table 8 that attitudes toward these policies depend in significant part on the types of paternalistic inclinations we document in the laboratory.

[47] For lending restrictions, we code neutrality as 0.

[48] We confine attention to menus 3 and 4 so that we can take an average over the same set of menus for each Choice Architect without including rounds with front-end delay.

[49] Subjects could decline to provide pieces of information such as their body mass index. In that case, we set the value of the corresponding variable equal to its population mean, and we include dummy variables to indicate such replacements. Online Appendix Section B.11 displays a version of Table 8 in which subject-specific control variables are excluded from the regressions. If anything, the inclusion of control variables leads to an increase in the magnitude of the coefficient estimates.

Projective Paternalism with Actual Policies.—Next, we investigate the relationships between respondents' policy judgments and their own characteristics, with the objective of distinguishing between mistakes-projective and ideals-projective paternalism. Focusing on the example of alcohol taxes, mistakes-projective paternalism should yield a positive relationship between alcohol consumption and support for alcohol taxation in politically unrelated jurisdictions, where the impact will fall on others. Intuitively, heavier drinkers are more likely to be problem drinkers. To the extent they project their problem drinking on others, they will conclude that others would benefit from measures that limit alcohol consumption. Ideals-projective paternalism has the opposite implication. Intuitively, light drinkers are less likely to enjoy consuming alcohol. To the extent they project their lack of enjoyment on others, they will conclude that the costs of limitating alcohol consumption are relatively low. Therefore, we should expect to observe a negative relationship between alcohol consumption and support for alcohol taxes.

To test these competing implications, we estimate OLS regressions relating subjects' support for each policy to pertinent personal behaviors and characteristics.[50] We include all subjects from all treatments, and include the same set of additional control variables as in the regressions of Table 8.

Panel C displays the results. Column 2 shows that subjects who binge drink more frequently are significantly *less* likely to express support for alcohol taxes on others, exactly as ideals-projective paternalism predicts. Because binge drinking is a good proxy for problem drinking, this finding strongly contradicts the hypothesis of mistakes-projective paternalism. Conditional on the level of binge drinking, greater weekly alcohol consumption also appears to reduce support for alcohol taxes on others, but the effect is not statistically significant.[51] Similarly, column 3 shows that German subjects are significantly more likely to express support for tobacco taxes in Switzerland if they do not smoke themselves. In column 4, we find that people with lower BMIs express stronger support for sugary drinks taxes on others, again consistent with ideals-projective paternalism. Similarly, subjects are less likely to support restricting other people's access to the market for short-term, high-interest lending when they have larger credit card balances (column 5).[52] While results such as those of Allcott, Lockwood, and Taubinsky (2019) suggest that the relationships between policy support and subject characteristics may partially be driven by differences in domain-specific knowledge such as health literacy, this factor cannot plausibly explain the strong relationship between laboratory choice and policy attitudes documented in panel B of the table. Overall, the data on judgments about real-world paternalistic policies therefore manifests the signature pattern associated with ideals-projective paternalism.

In principle, the observed relationship between one's own behavior and support for paternalistic policies may also be due to systematic variation in beliefs about the effectiveness of a given policy. Online Appendix Section C reports a vignette experiment with US subjects in which we control for variations in beliefs about

[50] Online Appendix Section B.10 presents corresponding estimates obtained from ordered probit models.

[51] The coefficient on weekly consumption becomes statistically significant (and remains negative) when we remove the control for binge drinking from the regression.

[52] Only 12 out of 403 subjects report ever having taken a short-term loan other than through their credit cards.

efficacy. In cases where the pattern of support for paternalistic policies points to ideals-projective paternalism, controlling for beliefs about efficacy leaves that pattern qualitatively unchanged.[53]

V. Conclusion

This paper examines when, why, and how people intervene in others' choices. In a setting involving intertemporal trade-offs, Choice Architects frequently remove options that are attractive to impatient decision-makers. They believe their interventions benefit the Chooser, and are thus acting paternalistically. We examine and reject two simple hypotheses about their motives. First, we rule out the possibility that Choice Architects intervene with the objective of controlling excessive impatience arising from perceived present bias, a motive practitioners of behavioral welfare economics often attribute to benevolent planners. Second, we rule out the possibility that Choice Architects intervene simply to impose their own judgments, without regard to the inclinations or subjective experiences of the affected individuals.

How do Choice Architects judge what is good for others? Ideals-projective paternalism emerges from our analysis as the key mechanism. An ideals-projective paternalist acts *as if* she believes others share, or ought to share, the ideals to which she aspires for herself. We show that ideals-projective paternalism is related to the false consensus effect (an objective fallacy). As a result, even though more patient Choice Architects do not intend to force a larger fraction of Choosers to choose more patiently, this is precisely what they do. We also find that ideals-projective paternalism extends to assessments of real-world paternalistic policies, which strongly correlate with behavior in the laboratory.

Throughout, we have remained agnostic about the effects of Choice Architects' interventions on Choosers' welfare. Finding an objective basis for making such assessments is challenging. For example, from a libertarian perspective, any intervention is welfare-reducing. Alternatively, if one believes that, given the high cost of impatience in our experiment, the most patient option always dominates the other alternatives, then removing the least patient option, or both the least patient and middle options, is weakly welfare-enhancing, because it helps Choosers avoid accidental errors. Existing evidence suggests, however, that people have a positive willingness to pay for autonomy (Fehr, Herz, and Wilkening 2013; Bartling, Fehr, and Herz 2014; Owens, Grossman, and Fackler 2014; Lübbecke and Schnedler 2018; Ackfeld and Ockenfels 2021).

There are many questions we hope future research will clarify. For example, how much of their own resources are subjects willing to give up in order to impose paternalistic restrictions on others, and do these amounts differ for mistakes-projective and ideals-projective paternalists? We also hope to extend the empirical study of paternalistic decision-making to subject pools consisting of "professional paternalists" such as medical doctors and policymakers. In contexts where objective benchmarks are available, existing evidence suggests that nearly everyone exhibits behavioral biases (Stango and Zinman 2019), including elected politicians (Sheffer et al. 2018).

[53] That survey asks about alcohol taxes, retirement savings mandates, restrictions on short-term, high-interest lending, and sugary drinks taxes. We observe statistically significant indications of ideals-projective paternalism for the first two policies. The evidence does not support mistakes-projective paternalism for any policy.

REFERENCES

Ackfeld, Viola, and Axel Ockenfels. 2021. "Do People Intervene to Make Others Behave Prosocially?" Unpublished.

Allcott, Hunt, Benjamin B. Lockwood, and Dmitry Taubinsky. 2019. "Regressive Sin Taxes, with an Application to the Optimal Soda Tax." *Quarterly Journal of Economics* 134 (3): 1557–626.

Altmann, Steffen, Armin Falk, and Andreas Grunewald. 2017. "Incentives and Information as Driving Forces of Default Effects." Unpublished.

Ambuehl, Sandro. 2017. "An Offer You Can't Refuse? Incentives Change How We Think." CESifo Working Paper 6296.

Ambuehl, Sandro, B. Douglas Bernheim, and Axel Ockenfels. 2021. "Replication Data for: What Motivates Paternalism? An Experimental Study." American Economic Association [publisher], Inter-university Consortium for Political and Social Research [distributor]. https://doi.org/10.3886/E120765V1.

Ambuehl, Sandro, Muriel Niederle, and Alvin E. Roth. 2015. "More Money, More Problems? Can High Pay Be Coercive and Repugnant?" *American Economic Review* 105 (5): 357–60.

Ambuehl, Sandro, and Axel Ockenfels. 2017. "The Ethics of Incentivizing the Uninformed: A Vignette Study." *American Economic* 107 (5): 91–95.

Bartling, Björn, Alexander Cappelen, Henning Hermes, Marit Skivenes, and Bertil Tungodden. 2020. "Free to Fail? Paternalistic Preferences in the US." Unpublished.

Bartling, Björn, Ernst Fehr, and Holger Herz. 2014. "The Intrinsic Value of Decision Rights." *Econometrica* 82 (6): 2005–39.

Basu, Kaushik. 2003. "The Economics and Law of Sexual Harassment in the Workplace." *Journal of Economic Perspectives* 17 (3): 141–57.

Basu, Kaushik. 2007. "Coercion, Contract and the Limits of the Market." *Social Choice and Welfare* 29: 559–79.

Benartzi, Shlomo, John Beshears, Katherine L. Milkman, Cass R. Sunstein, Richard H. Thaler, Maya Shankar, Will Tucker-Ray, William J. Congdon, and Steven Galing. 2017. "Should Governments Invest More in Nudging?" *Psychological Science* 28 (8): 1041–55.

Bernheim, B. Douglas. 2016. "The Good, the Bad, and the Ugly: A Unified Approach to Behavioral Welfare Economics." *Journal of Benefit-Cost Analysis* 7 (1): 12–68.

Bernheim, B. Douglas, and Antonio Rangel. 2009. "Beyond Revealed Preference: Choice-Theoretic Foundations for Behavioral Welfare Economics." *Quarterly Journal of Economics* 124 (1): 51–104.

Bernheim, B. Douglas, and Dmitry Taubinsky. 2018. "Behavioral Public Economics." In Handbook of Behavioral Economics: *Applications and Foundations 1*, Vol. 1, edited by B. Douglas Bernheim, Stefano DellaVigna, and David Laibson, 381–516. New York: Elsevier.

Besley, Timothy. 1988. "A Simple Model for Merit Good Arguments." *Journal of Public Economics* 35 (3): 371–83.

Bisin, Alberto, Alessandro Lizzeri, and Leeat Yariv. 2015. "Government Policy with Time Inconsistent Voters." *American Economic Review* 105 (6): 1711–37.

Carlin, Bruce Ian, Simon Gervais, and Gustavo Manso. 2013. "Libertarian Paternalism, Information Production, and Financial Decision Making." *Review of Financial Studies* 26 (9): 2204–28.

Clemens, Michael A. 2018. "Testing for Repugnance in Economic Transactions: Evidence from Guest Work in the Gulf." *Journal of Legal Studies* 47 (S1): S5–S44.

Currie, Janet, and Firouz Gahvari. 2008. "Transfers in Cash and In-Kind: Theory Meets the Data." *Journal of Economic Literature* 46 (2): 333–83.

Daniels, David P., and Julian J. Zlatev. 2019. "Choice Architects Reveal a Bias Toward Positivity and Certainty." *Organizational Behavior and Human Decision Processes* 151: 132–49.

Delavande, Adeline, Xavier Giné, and David McKenzie. 2011. "Measuring Subjective Expectations in Developing Countries: A Critical Review and New Evidence." *Journal of Development Economics* 94 (2): 151–63.

DellaVigna, Stefano, and Ulrike Malmendier. 2004. "Contract Design and Self-Control: Theory and Evidence." *Quarterly Journal of Economics* 119 (2): 353–402.

Dworkin, Gerald. 1972. "Paternalism." *Monist* 56 (1): 64–84.

Elías, Julio J., Nicola Lacetera, and Mario Macis. 2015a. "Markets and Morals: An Experimental Survey Study." *PLOS One* 10 (6).

Elías, Julio J., Nicola Lacetera, and Mario Macis. 2015b. "Sacred Values? The Effect of Information on Attitudes toward Payments for Human Organs." *American Economic Review* 105 (5): 361–65.

Elías, Julio J., Nicola Lacetera, and Mario Macis. 2019. "Paying for Kidneys? A Randomized Survey and Choice Experiment." *American Economic Review* 109 (8): 2855–88.

Engelmann, Dirk, and Martin Strobel. 2012. "Deconstruction and Reconstruction of an Anomaly." *Games and Economic Behavior* 76 (2): 678–89.

Exley, Christine L., and Judd B. Kessler. 2017. "Equity Concerns Are Narrowly Framed: Why Money Cannot Buy Time." Unpublished.

Faravelli, Marco, Priscilla Man, and Randall Walsh. 2015. "Mandate and Paternalism: A Theory of Large Elections." *Games and Economic Behavior* 93: 1–23.

Fedyk, Anastassia. 2017. "Asymmetric Naïveté: Beliefs about Self-Control." Unpublished.

Fehr, Ernst, Holger Herz, and Tom Wilkening. 2013. "The Lure of Authority: Motivation and Incentive Effects of Power." *American Economic Review* 103 (4): 1325–59.

Foerster, Stephen, Juhani T. Linnainmaa, Brian T. Melzer, and Alessandro Previtero. 2017. "Retail Financial Advice: Does One Size Fit All?" *Journal of Finance* 72 (4): 1441–82.

Frederick, Shane, George Loewenstein, and Ted O'Donoghue. 2002. "Time Discounting and Time Preference: A Critical Review." *Journal of Economic Literature* 40 (2): 351–401.

Gangadharan, Lata, Philip J. Grossman, and Kristy Jones. 2015. "Donor Types and Paternalism." Unpublished.

Gigerenzer, Gerd. 2008. "Moral Intuition = Fast and Frugal Heuristics?" In *Moral Psychology, Volume 2: The Cognitive Science of Morality: Intuition and Diversity,* edited by Walter Sinnot-Armstrong, 1–26. Cambridge, MA: MIT Press.

Glaeser, Edward L. 2005. "Paternalism and Psychology." NBER Working Paper 11789.

Gorber, S. Connor, M. Tremblay, D. Moher, and B. Gorber. 2007. "A Comparison of Direct vs. Self-Report Measures for Assessing Height, Weight and Body Mass Index: A Systematic Review." *Obesity Reviews* 8 (4): 307–26.

Gruber, Jonathan, and Botond Kőszegi. 2001. "Is Addiction 'Rational?' Theory and Evidence." *Quarterly Journal of Economics* 116 (4): 1261–1303.

Gruber, Jonathan, and Botond Kőszegi. 2004. "Tax Incidence When Individuals Are Time-Inconsistent: The Case of Cigarette Excise Taxes." *Journal of Public Economics* 88 (9–10): 1959–87.

Gul, Faruk, and Wolfgang Pesendorfer. 2001. "Temptation and Self-Control." *Econometrica* 69 (6): 1403–35.

Holt, Charles A., and Susan K. Laury. 2002. "Risk Aversion and Incentive Effects." *American Economic Review* 92 (5): 1644–55.

Ifcher, John, and Homa Zarghamee. 2020. "Behavioral Economic Phenomena in Decision-Making for Others." *Journal of Economic Psychology* 77: 102180.

Jacobsson, Fredric, Magnus Johannesson, and Lars Borgquist. 2007. "Is Altruism Paternalistic?" *Economic Journal* 117 (520): 761–81.

Kahneman, Daniel. 2011. *Thinking Fast and Slow*. New York: Macmillan.

Kataria, Mitesh, M. Vittoria Levati, and Matthias Uhl. 2014. "Paternalism with Hindsight: Do Protégés React Consequentialistically to Paternalism?" *Social Choice and Welfare* 43 (3): 731–46.

Krawczyk, Michal, and Lukasz Patryk Wozny. 2017. "An Experiment on Temptation and Attitude towards Paternalism." Unpublished.

Krueger, Joachim, and Russell W. Clement. 1994. "The Truly False Consensus Effect: An Ineradicable and Egocentric Bias in Social Perception." *Journal of Personality and Social Psychology* 67 (4): 596–610.

Laibson, David. 2018. "Private Paternalism, the Commitment Puzzle, and Model-Free Equilibrium." *AEA Papers and Proceedings* 108: 1–21.

Leider, Stephen, and Alvin E. Roth. 2010. "Kidneys for Sale: Who Disapproves, and Why?" *American Journal of Transplantation* 10 (5): 1221–27.

Linnainmaa, Juhani T., Brian T. Melzer, and Alessandro Previtero. Forthcoming. "The Misguided Beliefs of Financial Advisors." *Journal of Finance*.

Locke, John. 1764. *Two Treatises of Government*. London: A. Millar et al.

Loewenstein, George, and Emily Haisley. 2007. "The Economist as Therapist: Methodological Ramifications of 'Light' Paternalism." In *The Foundations of Positive and Normative Economics*, edited by Andrew Caplin and Andrew Schotter, 210–48. Oxford: Oxford University Press.

Loewenstein, George, Ted O'Donoghue, and Matthew Rabin. 2003. "Projection Bias in Predicting Future Utility." *Quarterly Journal of Economics* 118 (4): 1209–48.

Lübbecke, Silvia, and Wendelin Schnedler. 2018. "Don't Patronize Me! An Experiment on Rejecting Paternalistic Help." Unpublished.

Lusk, Jayson L., Stephan Marette, and F. Bailey Norwood. 2013. "The Paternalist Meets His Match." *Applied Economic Perspectives and Policy* 36 (1): 61–108.

Madarász, Kristóf. 2012. "Information Projection: Model and Applications." *Review of Economic Studies* 79 (3): 961–85.

Mill, John Stuart. 1869. *On Liberty*. London: Longmans, Green, Reader, and Dyer.
Moffitt, Robert A. 2003. "The Negative Income Tax and the Evolution of US Welfare Policy." *Journal of Economic Perspectives* 17 (3): 119–40.
Moffitt, Robert. 2006. "Welfare Work Requirements with Paternalistic Government Preferences." *Economic Journal* 116 (515): F441–58.
Mulligan, Casey B., and Tomas J. Philipson. 2000. "Merit Motives and Government Intervention: Public Finance in Reverse." NBER Working Paper 7698.
National Institute on Alcohol Abuse and Alcoholism. 2018. "Recommended Alcohol Questions." https://www.niaaa.nih.gov/research/guidelines-and-resources/recommended-alcohol-questions.
Nawaz, Haq, Wendy Chan, Mustapha Abdulrahman, David Larson, and David L. Katz. 2001. "Self-Reported Weight and Height: Implications for Obesity Research." *American Journal of Preventive Medicine* 20 (4): 294–98.
O'Donoghue, Ted, and Matthew Rabin. 1999. "Doing It Now or Later." *American Economic Review* 89 (1): 103–24.
O'Donoghue, Ted, and Matthew Rabin. 2006. "Optimal Sin Taxes." *Journal of Public Economics* 90 (10–11): 1825–49.
Owens, David, Zachary Grossman, and Ryan Fackler. 2014. "The Control Premium: A Preference for Payoff Autonomy." *American Economic Journal: Microeconomics* 6 (4): 138–61.
Rizzo, Mario J., and Glen Whitman. 2019. *Escaping Paternalism: Rationality, Behavioral Economics, and Public Policy*. Cambridge: Cambridge University Press.
Ross, Lee, David Greene, and Pamela House. 1977. "The 'False Consensus Effect': An Egocentric Bias in Social Perception and Attribution Processes." *Journal of Experimental Social Psychology* 13 (3): 279–301.
Roth, Alvin E. 2007. "Repugnance as a Constraint on Markets." *Journal of Economic Perspectives* 21 (3): 37–58.
Saint-Paul, Gilles. 2004. "Policy Consequences of Limited Cognitive Ability." *Spanish Economic Review* 6: 97–105.
Schroeder, Juliana, Adam Waytz, and Nicholas Epley. 2017. "Endorsing Help for Others That You Oppose for Yourself: Mind Perception Alters the Perceived Effectiveness of Paternalism." *Journal of Experimental Psychology: General*: 146 (8): 1106–25.
Sheffer, Lior, Peter John Loewen, Stuart Soroka, Stefaan Walgrave, and Tamir Sheafer. 2018. "Nonrepresentative Representatives: An Experimental Study of the Decision Making of Elected Politicians." *American Political Science Review* 112 (2): 302–21.
Smith, Vernon L. 1976. "Experimental Economics: Induced Value Theory." *American Economic Review* 66 (2): 274–79.
Sobell, Linda C., and Mark B. Sobell. 1995. "Alcohol Consumption Measures." In *Assessing Alcohol Problems: A Guide for Clinicians and Researchers*, Vol. 2, edited by, 75–100. Bethesda, MD: National Institute on Alcohol Abuse and Alcoholism.
Stango, Victor, and Jonathan Zinman. 2019. "We Are All Behavioral, More or Less: Measuring and Using Consumer-Level Behavioral Sufficient Statistics." NBER Working Paper 25540.
Sunstein, Cass R. 2005. "Moral Heuristics." *Behavioral and Brain Sciences* 28 (4): 531–41.
Thaler, Richard H., and Hersh M. Shefrin. 1981. "An Economic Theory of Self-Control." *Journal of Political Economy* 89 (2): 392–406.
Thaler, Richard H., and Cass R. Sunstein. 2003. "Libertarian Paternalism." *American Economic Review* 93 (2): 175–79.
Uhl, Matthias. 2011. "Do Self-Committers Mind Other-Imposed Commitment? An Experiment on Weak Paternalism." *Rationality, Markets and Morals* 2 (40).
Van Boven, Leaf, David Dunning, and George Loewenstein. 2000. "Egocentric Empathy Gaps between Owners and Buyers: Misperceptions of the Endowment Effect." *Journal of Personality and Social Psychology* 79 (1): 66–76.
West, Robert, Witold Zatonski, Krzysztof Przewozniak, and Martin J. Jarvis. 2007. "Can We Trust National Smoking Prevalence Figures? Discrepancies between Biochemically Assessed and Self-Reported Smoking Rates in Three Countries." *Cancer Epidemiology Biomarkers & Prevention* 16 (4): 820–22.
Zamir, Eyal. 1998. "The Efficiency of Paternalism." *Virginia Law Review* 84 (2): 229–86.
Zlatev, Julian J., David P. Daniels, Hajin Kim, and Margaret A. Neale. 2017. "Default Neglect in Attempts at Social Influence." *Proceedings of the National Academy of Sciences* 114 (52): 13643–48.