NBER WORKING PAPER SERIES

PROJECTIVE PATERNALISM

Sandro Ambuehl
B. Douglas Bernheim
Axel Ockenfels

Working Paper 26119
http://www.nber.org/papers/w26119

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
July 2019

We thank Joshua Knobe, Yoram Halevy, Armin Falk, Glenn W. Harrison, Kiryl Khalmetski, Frank Schillbach, 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 number 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 agreement No 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. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

Projective Paternalism
Sandro Ambuehl, B. Douglas Bernheim, and Axel Ockenfels
NBER Working Paper No. 26119
July 2019
JEL No. P43,P48

ABSTRACT

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. We consider several hypotheses concerning CAs' motives. A conventional behavioral welfarist acts as a correctly informed social planner; a mistakes-projective paternalist removes options she wishes she could reject when choosing for herself; an ideals-projective paternalist seeks to align others' choices with her own aspirations. Ideals-projective paternalism provides the best explanation for interventions in the laboratory and rationalizes support for actual paternalistic policies.

Sandro Ambuehl
Department of Management UTSC
Rotman School of Management
University of Toronto
105 St. George Street
Toronto, ON, M5S 3E6
CANADA
sandro.ambuehl@utoronto.ca

Axel Ockenfels
Department of Economics
University of Cologne
D-50923 Köln
Germany
ockenfels@uni-koeln.de

B. Douglas Bernheim
Department of Economics
Stanford University
Stanford, CA 94305-6072
and NBER
bernheim@stanford.edu

An online appendix, data files, and code used for the analysis are available at:
http://www.nber.org/data-appendix/w26119

1 Introduction

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, 1971; Zamir, 1998). Paternalistically motivated social programs play a large role in the U.S. economy (Mulligan and Philipson, 2000; Moffitt, 2003; Currie and Gahvari, 2008), and in some important instances even the business models of private companies are paternalistically 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 who find themselves with the authority to influence and nudge others, 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 in situations where it is possible to compare choices with objective benchmarks (Kahneman, 2011). It is reasonable to assume that normative judgments are also susceptible to systematic confusion or bias.

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 verify the external validity of our findings, we also demonstrate 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 Architect construct choice sets that determine the opportunities available to others in the role of Choosers. Each choice option consists of two payments, one received "sooner" and the other received "later." Impatience is costly: larger earlier payments are associated with smaller total payments. We begin by showing that 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. By definition, interventions are paternalistic only if Choice Architects believe they increase the Chooser's well-being (Dworkin, 1972). According to both incentive-compatible and non-incentivized measures, our Choice Architects believe their interventions are helpful, and those who impose more severe restrictions believe the benefits are greater. We rule out alternative motives, such as the possibility that subjects intervene simply because they prefer to do something rather than remain idle: a condition

[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" (Bennyhoff and Kinniry Jr, 2011).

that removes the opportunity to take issue with the Choosers' objective function (by inducing Chooser's preferences experimentally) dramatically decreases the frequency of interventions.

Taking these initial findings as a starting point, we turn to the main focus of our investigation, which is to determine how Choice Architects decide which options are good for others and which are bad. As a benchmark case, we consider the hypothesis that Choice Architects are *conventional behavioral welfarists* – in other words, that they act as benevolent social planners who know the true distributions of choices and mistakes, and who try to maximize aggregate welfare. Under this hypothesis, there would be no relationship between the restrictions Choice Architects impose on others and either their own preferences over the pertinent options, or the mistakes that infect own choices among those options.

In evaluating the plausibility of this benchmark, one may well wonder how people arrive at their knowledge of others' preferences and mistakes. 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, distinguished according to whether paternalists reason about others based on their own mistakes, or based on their own preferences. *Mistakes-projective paternalism* is present when the paternalist assumes that others tend to share her susceptibility to error. A mistakes-projective paternalist behaves as if trying to help others avoid choices she herself would like to reject, but chooses nevertheless. This inclination generates a *negative correlation* between the choices she makes for herself and the restrictions she imposes on others. *Ideals-projective paternalism* is present when the paternalist behaves as if she assumes her own preferences are relevant for others, either because she thinks they tend to share her preferences,[2] or because she simply believes her preferences are valid and theirs are not. As a result, it generates a *positive correlation* between the choices she makes for herself and the restrictions she imposes on others. Accordingly, to distinguish between conventional behavioral welfarism, mistakes-projective paternalism, and ideals-projective paternalism, we focus on measuring the aforementioned correlation.

The hypothesis of ideals-projective paternalism provides the most compelling account of our subjects' behavior. We find that more patient Choice Architects impose greater patience on others, and that this pattern reflects their judgments about what the Choosers ought to do, rather than a greater inclination to intervene. Patient Choice Architects also believe more strongly that restrictions enhance the Chooser's well-being. Notably, a Choice Architect's proclivity to engage in ideals-projective paternalism is strongly correlated with her susceptibility to a recognized cognitive bias, the *false consensus effect* (Ross, Greene and House, 1977): ideals-projective paternalism is more prevalent among subjects who assume (incor-

[2]This form of projective paternalism is related to the concept of *projection bias* (Van Boven, Dunning and Loewenstein, 2000); i.e., the tendency to exaggerate the degree to which an individual's future tastes (rather than the tastes of others) resemble their current tastes.

rectly) that the choices of others tend to resemble their own. Despite this finding, we also show that ideals-projective paternalism is largely robust with respect to the availability and provision of information concerning others' choices and mistakes. Such knowledge affects the stringency of the restrictions Choice Architects impose on Choosers, but does not attenuate the strength of the relationship between these restrictions and Choice Architects' own preferences. We argue that these findings shed light on the mechanisms of ideals-projective paternalism.

The next step in our investigation is to assess whether paternalists agree with the types of normative judgments that often appear in analyses involving applied behavioral welfare economics. One common supposition among many behavioral economists is that time inconsistency arises at least in part from a tendency to place "too much" weight on immediate or near-term experiences ("present bias"). According to this normative hypothesis, removing the lure of immediacy by introducing "front-end delay" either eliminates the bias (as in the case of quasihyperbolic discounting), or reduces it (as in the case of hyperbolic discounting).[3] To the extent our Choice Architects agree with this normative principle and try to mimic welfarist social planners, the introduction of front-end delay should eliminate interventions, or should at least reduce their frequency. The same implication follows for mistakes-projective paternalists who recognize their own present bias. On the contrary, we find that people are slightly *more* likely to intervene in otherwise equivalent decision problems with front-end delay, and to impose greater patience when they intervene, despite predicting correctly that others exercise greater patience in these settings. Because people also choose more patiently for themselves with front-end delay, these patterns are consistent with ideals-projective paternalism.

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

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, some subjects choose between one of two options for themselves and decide whether to eliminate one of the options for others.[4] While these studies do examine the proclivity to intervene, as well as correlations between chosen interventions and

[3] A large number of studies have found that the introduction of front-end delay increases patience (Frederick et al., 2002). For a critical discussion of the common normative claim that people place "too much" weight on the present, see Bernheim (2016).

[4] 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); Gangadharan, Grossman and Jones (2015); Lusk, Marette and Norwood (2013); Schroeder, Waytz and Epley (2017).

the options subjects select for themselves, they do not address the associated questions that provide our focus in the current paper. With respect to the questions they do address, their experimental designs introduce two potentially important confounds: first, there are no opportunities to provide advice apart from imposing a restriction; second, the close juxtaposition of essentially identical decisions (for the subjects themselves and for others) may introduce spurious correlation through anchoring or a demand for consistency.

Several other lines of work bear on the empirical analysis of paternalism. One examines the tendency for professional advisors to steer their clients toward the same options they choose for themselves (Foerster, Linnainmaa, Melzer and Previtero, 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).[5] A third studies social disapproval of ostensibly *repugnant transactions* (Roth, 2007), such as paid organ donation (Basu, 2003, 2007; Leider and Roth, 2010; Elias, Lacetera and Macis, 2015a,b, 2016; Ambuehl, 2017; Ambuehl and Ockenfels, 2017; Clemens, 2017; Exley and Kessler, 2017).[6] 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, 2018, for a review). Finally, the literature on social preferences studies the general question of how people make decisions that affect others (see Fehr and Fischbacher, 2002; Cooper and Kagel, 2016, for reviews).

The empirical study of paternalism connects with various other branches of the literature. For example, 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; Benartzi, Beshears, Milkman, Sunstein, Thaler, Shankar, Tucker-Ray, Congdon and Galing, 2017, for reviews). Our findings raise questions about the objectivity of the principles underlying such judgments in domains that involve preferences, and thereby underscore the urgent need for objective principles of behavioral welfare analysis, as in Bernheim and Rangel (2009) and Bernheim (2016).[7] Our research also contributes more broadly to a literature in cognitive science on moral heuristics (see Sunstein, 2005; Gigerenzer, 2008, for reviews).

The remainder of this paper proceeds as follows. Section 2 formalizes the concepts of ideals-projective and mistakes-projective paternalism and derives their first-order implications. Section 3 outlines our experimental design. Section 4 demonstrates that subjects frequently act paternalistically, typically by

[5] 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); Laibson (2018).

[6] 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).

[7] For a summary of this method as well as empirical applications, see Bernheim and Taubinsky (2018).

forcing others to choose patiently. Section 5 documents our main finding concerning ideals-projective paternalism, explores underlying mechanisms, demonstrates its robust resistance to information provision, and explores the effect of introducing front-end delay. Section 6 documents projective paternalism regarding real-world paternalistic policies. Finally, Section 7 outlines directions for further research and concludes.

2 Conceptual framework

2.1 Preliminaries

Setting There are two agents, a Choice Architect (she) and Chooser (he). Consumption opportunities for the Chooser are indexed by $c \in C \equiv [\underline{c}, \overline{c}]$. The Choice Architect determines the subset $[\underline{r}, \overline{r}] \subseteq C$ of consumption opportunities available to the Chooser. The Chooser then selects $c \in [\underline{r}, \overline{r}]$. In an intertemporal choice application, for instance, higher indices may reflect greater patience, so the Choice Architect's selection of the restrictions $\underline{r}$ and $\overline{r}$ determines the least and most patient choices a Chooser may make.

Suppose the Choice Architect believes the Chooser would select c from the unrestricted set C, 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). The Choice Architect has imperfect information about the Chooser. Her beliefs about the Chooser's characteristics are given by a cumulative distribution function $F(u, m)$.[8]

Suppose further that, if faced with the same decision problem, the Choice Architect would select $c^A \in C$ for herself. She believes that another option, indexed u^A, would be ideal, and acknowledges that her actual choice c^A involves a mistake $m^A = u^A - c^A$.[9]

Varieties of paternalism We define varieties of paternalism according to the way in which the Choice Architect's beliefs about Choosers, $F(u, m)$, depend on her own type, (u^A, m^A). A *conventional behavioral welfarist* acts as a benevolent and correctly informed social planner of the type envisioned in many standard behavioral welfare analyses (see, e.g., Bernheim and Taubinsky, 2018). They do not consider their

[8] We permit the possibility that the Choice Architects' beliefs about what is good for the Chooser differ from the latter's own assessment, as only the former is relevant for our model. Consider the following examples. (i) A time-consistent Chooser has an annual discount factor of 0.8. The Choice Architect believes that the Chooser would be better off if he applied an annual discount factor of 0.95. Hence, even though the Chooser is time-consistent, the Choice Architect views his choices as mistaken. (ii) A Chooser is a naïve quasi-hyperbolic discounter with $\beta = 0.5$ and $\delta = 1$. The Choice Architect believes the Chooser would be best off if he behaved as if he were a time-consistent decision maker with $\delta = 1$. In this case she views the Chooser's unconstrained choice as mistaken, but only because he is time-inconsistent.

[9] We allow for the possibility that Choice Architects recognize biases and mistakes in others even if they are not aware of the same biases and mistakes in themselves (see, e.g., Fedyk, 2017). Furthermore, a Choice Architect who admits to mistakes does not necessarily exhibit a demand for commitment devices, and therefore need not be sophisticated in the sense of O'Donoghue and Rabin (1999). In particular, she may be unmotivated to correct her mistake, or she may simply view her failure to employ a commitment device as an additional mistake. Our definition simply requires that a Choice Architect subjectively characterizes certain actions (for herself) as mistakes, and not that she is necessarily willing to self-correct.

own inclinations when intervening in others' choices. In contrast, projective paternalists form their views about others by extrapolating from their understanding of themselves. A *mistakes-projective paternalist* believes that others are similar to herself with respect to the existence and intensity of mistakes—she believes m and m^A are positively correlated, but that u and u^A are unrelated. An *ideals-projective paternalist* projects her own ideals on others, thinking that others either do, or should, have ideals similar to her own—she believes u and u^A are positively correlated, but that m and m^A are unrelated. Naturally, hybrids of mistakes-projective and ideals-projective paternalism are also possible.[10]

In order to define these concepts formally, we let $F_u(\cdot)$ and $F_m(\cdot)$ denote the marginal distributions of beliefs about u and m, respectively. Moreover, for each u, we let $F_m(\cdot|u)$ denote the distribution of m conditional on u, and for each m we let $F_u(\cdot|m)$ denote the distribution of u conditional on m. We assume that the conditional distributions $F_u(\cdot|m)$ have the same support for all m and for all Choice Architects, and that they admit probability densities. Throughout, we say that a CDF H for a variable x is increasing in a parameter θ if, for all $\theta' > \theta''$, $H(\cdot, \theta')$ first-order stochastically dominates $H(\cdot, \theta'')$.

Definition

(i) A Choice Architect is a *conventional behavioral welfarist* if F does not depend on her type.

(ii) A Choice Architect is *mistakes-projective* if $F_m(\cdot|u)$ is increasing in m^A and $F_u(\cdot)$ does not depend on her type.

(iii) A Choice Architect is *ideals-projective* if $F_u(\cdot|m)$ is increasing in u^A and $F_m(\cdot)$ does not depend on her type.

These definitions reference the beliefs the Choice Architect would have held if her own values of u^A and m^A had been different. Ultimately, we will infer the predominant type in the population through correlations across subjects.

Optimal interventions The Choice Architect constructs the Chooser's opportunity set to maximize perceived welfare. Specifically, the Choice Architect believes that if a Chooser's choice c differs from his ideal u, he sustains a welfare loss of $l(c-u)$. Here, $l(z)$ is a twice-differentiable, strictly concave function with $l(0) = 0$. The Choice Architect chooses $\underline{r}$ and $\overline{r}$ to minimize the Chooser's expected loss. An additive term $\kappa(\cdot)$ with $\kappa(\overline{c} - \underline{c}) = 0$ captures the Choice Architect's feelings (either positive or negative) about restricting the Choosers' options, other than those arising from anticipated effects on the Chooser's

[10] As a purely logical matter, one could also entertain the possibility that u is correlated with m^A and m is correlated with u^A, but we see no natural explanations for such correlations.

welfare.[11] Hence, the Choice Architect's objective is given by

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

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

Our results concern the comparative statics for optimal restrictions $[\underline{r}^*,\overline{r}^*]$, defined as the values that maximize expression (1). For simplicity we focus on the case of $\kappa = 0$.[12]

It is worth noting that our formulation assumes the Choice Architect is concerned only with the Chooser's outcome, and not with his choice process. For example, the Choice Architect does not wish to restrict the Chooser's opportunity set in an attempt to lessen the cognitive effort associated with choosing from a large set.[13] Moreover, while we limit Choice Architects to specifying convex opportunity sets, Amador, Werning and Angeletos (2006) identify conditions under which this restriction is without loss of generality in a setting closely related to ours.

Additional assumptions Our formal results rely on the following two assumptions. First, in many domains, it is reasonable to assume that perceived mistakes are unidirectional. In the domain of intertemporal choice, for instance, there appears to be a widespread perception that people struggle to act sufficiently patiently rather than sufficiently impatiently. Similarly, general wisdom holds that people tend to exercise too little and excessively consume unhealthy foods. Accordingly, we assume that $P(m \geq 0) = 1$, where P denotes the probability measure induced by F.[14]

Assumption 1 $P(m \geq 0) = 1$.

Second, because we aim to infer the typical subject's paternalistic type from the correlation between mandates $(\underline{r}^*,\overline{r}^*)$ and their own choices c^A, we require an assumption concerning the distribution of types among Choice Architects. Intuitively, the assumption states that a Choice Architect who selects a higher value of c^A for herself has a higher perceived ideal u^A and a lower perceived mistake m^A. Formally, we define $G_{m^A}(\cdot|c^A)$ as the marginal distribution of Choice Architects' mistakes among those whose own

[11] A "pure" libertarian Choice Architect, for instance, satisfies $\kappa(\overline{c} - \underline{c}) = 0$ and $\kappa(z) = \infty$ for any $z < \overline{c} - \underline{c}$. Under this formulation of libertarianism, a libertarian who is forced to make a surrogate choice will be indistinguishable from an otherwise identical non-libertarian. This implication is consistent with the experimental results in Section 5.

[12] This assumption rules out the possibility that Choice Architects intervene due to a desire to exert control, which we would represent by assuming that $\kappa(d) < 0$ for $d > 0$. Section 4.3 provides an empirical test of that assumption.

[13] Similarly, by assumption, the Choice Architect does not restrict the Chooser's opportunities in order to spare the Chooser the disutility associated with overcoming temptation (as in Thaler and Shefrin (1981) or implicitly in Gul and Pesendorfer (2001)). Significantly, we structure our experiment in a way that eliminates considerations such as cognitive costs and temptation.

[14] That is, $P(A) = \int_A dF$.

choice is c^A, and $G_{u^A}(\cdot|c^A)$ as the distribution of Choice Architects' ideals among those whose own choice is c^A.

Assumption 2 $G_{u^A}(\cdot|c^A)$ is increasing in c^A and $G_{m^A}(\cdot|c^A)$ is decreasing in c^A.

2.2 Comparative statics

We begin with a lemma: under Assumption 1 (unidirectional mistakes), optimal interventions involve a minimum requirement $\underline{r}^*$ but no cap ($\bar{r}^* = \bar{c}$). The reason is that a paternalist sets restrictions to prevent mistakes. If no Chooser makes excessively high choices, the Choice Architect has no reason to impose an upper limit. As with all other results, proofs are in Appendix A.

Lemma 1 *The Choice Architect does not impose a binding upper bound:* $\bar{r}^* = \bar{c}$.

Correlations between mandates and choices Our first main result shows that it is possible to identify the dominant mode of paternalism by determining whether the correlation between Choice Architects' mandates and the choices they make for themselves is positive, negative, or zero.

Proposition 1

(i) The optimal mandate $\underline{r}^*$ *imposed by a mistakes-projective paternalist is increasing in her own mistake* m^A. *The distribution of optimal mandates imposed by mistakes-projective paternalists is decreasing in their own choices,* c^A.

(ii) The optimal mandate $\underline{r}^*$ *imposed by an ideals-projective paternalist is increasing in her own ideal* u^A. *The distribution of optimal mandates imposed by ideals-projective paternalists is increasing in their own choices,* c^A.

(iii) The optimal mandate $\underline{r}^*$ *imposed by a conventional behavioral welfarist is independent of her own type and choice* c^A.

In applying this result, we examine variation across Choice Architects, and ask whether those who choose larger values of c for themselves impose larger or smaller values of c on others. Specifically, in the context of intertemporal choice, under Assumption 2, more patient Choice Architects view themselves as pursing more patient ideals and as less prone to mistakes. These perceptions lead them to impose greater patience on others if they are ideals-projective paternalists, and to impose less patience if they are mistakes-projective paternalists.

To see the intuition underlying this result, observe that increasing a mandate $\underline{r}$ has two effects. It benefits Choosers with $u - m < \underline{r} < u$ by bringing their choice $\varphi(\underline{r}, \bar{r})$ closer to the ideal u. Simultaneously, it imposes a higher cost on Choosers with $u < \underline{r}$ as it increases the distance between their choice $\varphi(\underline{r}, \bar{r})$ and their ideal u.

Consider two mistakes-projective paternalists, i and j, who share the same ideals for themselves and have the same beliefs about the distribution of ideals for others, but differ with respect to the magnitudes of their own errors. Assuming i makes larger mistakes ($m_i^A > m_j^A$), she also believes that Choosers make larger mistakes. Accordingly, relative to j, i believes than an increase in the mandate involves larger incremental benefits and the same incremental cots. Her optimal mandate $\underline{r}^*$ will thus be higher than j's, even though her own choice is lower. Thus, mandates are negatively related to the choices mistakes-projective paternalists make for themselves.

Next, consider two ideals-projective paternalists, k and l, who are equally susceptible to mistakes when choosing for themselves, and who have the same beliefs about the distribution of mistakes for others, but who have different ideals for themselves. Assuming k has a higher ideal for herself than l ($u_k^A > u_l^A$), she also believes that Choosers have higher ideals. Accordingly, she must set a higher mandate to achieve the same degree of mistake-mitigation. It follows that her optimal mandate $\underline{r}^*$ will be higher, just like her own choice. Thus, optimal mandates are positively related to the choices ideals-projective paternalists make for themselves.

Finally, the result for conventional behavioral welfarists is a straightforward implication of the assumption that their beliefs are independent of their own types.

Beliefs about welfare effects Next we examine Choice Architects' beliefs about the welfare effects associated with setting restrictions. For this purpose, we define $\Delta W(\underline{r}^*) \equiv W(\underline{r}^*, \overline{c}) - W(\underline{c}, \overline{c})$. Clearly, a Choice Architect believes that her optimal restriction strictly increases the Chooser's expected welfare, $\Delta W(\underline{r}^*) > 0$, if and only if she believes that it will bind with positive probability. Focusing on that case, the next proposition derives the relation between $\Delta W(\underline{r}^*)$ and Choice Architects' type and mandate. Formally, we assume that beliefs about the distribution of ideals and mistakes $F(m, u)$ are continuously differentiable in the Choice Architect's own type (u^A, m^A).

Proposition 2 *Suppose $P(u - m < \underline{r}^*) > 0$. Then $\Delta W(\underline{r}^*) > 0$. Moreover:*

(i) For a mistakes-projective paternalist, $\Delta W(\underline{r}^*)$ is increasing in m^A and in $\underline{r}^*$. The distribution of $\Delta W(\underline{r}^*)$ among mistakes-projective paternalists is decreasing in their own choices c^A.

(ii) For an ideals-projective paternalist, $\Delta W(\underline{r}^*)$ may be increasing or decreasing in u^A and in $\underline{r}^*$. The distribution of $\Delta W(\underline{r}^*)$ among ideals-projective paternalists may be increasing or decreasing in their own choices c^A.

The proof of Proposition 2 relies on the envelope theorem, as does its underlying intuition. It is sufficient to consider the effect of a change in the belief distribution on welfare, holding the intervention fixed. Recall that the optimal mandate $\underline{r}^*$ balances the benefits and costs of restricting choice, as explained above. Fixing $\underline{r}^*$, an increase in m^A increases a mistakes-projective paternalist's beliefs about

m, which increases the benefit of imposing $\underline{r}^*$ to Choosers who would have made too low a choice. It does not affect the cost of imposing $\underline{r}^*$ to Choosers who are forced to make a choice that exceeds their ideal. Hence $\Delta W(\underline{r}^*)$ increases. In contrast, for an ideals-projective paternalist, an increase in u^A increases beliefs about u. Within the set of Choosers whom the intervention helps (those with $u - m < \underline{r}^* < u$), the shift in the distribution of u reduces the benefit of the intervention. Within the set of Choosers who would be better off without the intervention (those with $u < \underline{r}^*$), the shift in the distribution of u reduces the mandate's harm.[15] Whether the former effect outweighs the latter depends on parameter values, so $\Delta W(\underline{r}^*)$ may be increasing or decreasing in u^A. (The proof of this proposition in Appendix A provides numerical examples of each possibility.) The relation between $\Delta W(\underline{r}^*)$ and the optimal mandate, $\underline{r}^*$ follows from combining the preceding arguments with Proposition 1, as does the relation between $\Delta W(\underline{r}^*)$ and Choice Architects' own choices, c^A. A mistakes-projective paternalist's optimal mandate $\underline{r}^*$ is increasing in m^A, which implies a positive relation between $\Delta W(\underline{r}^*)$ and $\underline{r}^*$. For an ideals-projective paternalist, the ambiguous relation between $\Delta W(\underline{r}^*)$ and u^A translates into an ambiguous relation between $\Delta W(\underline{r}^*)$ and $\underline{r}^*$. Finally, while a conventional behavioral welfarist's beliefs about Choosers do not depend on her own type, depending on how her beliefs about mistakes relate to her beliefs about ideals, there may be a positive or negative relation between ΔW and $\underline{r}^*$.

Surrogate choices We are also interested in how Choice Architects make surrogate choices, denoted s^*, on behalf of the Chooser. This knowledge will help us determine empirically whether Choice Architects who impose stricter mandates have different beliefs concerning what is good for the Chooser, or whether they are merely more willing to act based on such beliefs. Our framework implies that ideals-projective paternalists' surrogate choices are increasing in their own choice. In contrast, these two decisions are independent both for conventional behavioral welfarists and for mistakes-projective paternalists. This result is a consequence of the fact that surrogate choices only depend on beliefs about ideals, which are uncorrelated with the choices mistakes-projective paternalists make for themselves.

Corollary 1

(i) The optimal surrogate choice s^* made by a conventional behavioral welfarist or by a mistakes-projective paternalist is independent of her own choice c^A.

(ii) The optimal surrogate choice s^* made by a projective paternalist is increasing in her own choice c^A.

Choice Architects' predictions and false consensus By asking Choice Architects to predict the choices unrestricted Choosers have made, we will elicit their beliefs concerning the distribution of $c =$

[15] Because the optimal intervention trades off the benefits to those who are forced to make a choice closer to their ideals with the harm to those forced to make a choice further from their ideals, the probability that the optimal intervention causes harm to some Choosers is generally positive.

$u - m$. In the context of our model, a Choice Architect believes the CDF of c is given by $Q(c) = \int \mathbb{1}(u - m \leq c) dF(u, m)$, where $\mathbb{1}$ denotes the indicator function.

The model implies that both ideals-projective and mistakes-projective paternalists will exhibit a false consensus effect: their predictions are positively related to the choice options they select for themselves. In contrast, the relation between predictions and mandates differs for the two types.

Proposition 3 *For both ideals-projective and mistakes-projective paternalists, Q is increasing in c^A. Moreover:*

(i) A mistakes-projective paternalist's optimal mandate $\underline{r}^*$ is decreasing in her prediction Q.

(ii) An ideals-projective paternalist's optimal mandate $\underline{r}^*$ is increasing in her prediction Q.

Intuitively, an increase in a mistakes-projective paternalist's own choice c^A implies that her mistake m^A is smaller. This property, in turn, implies a belief that the Chooser makes smaller mistakes, which leads to a higher prediction Q and a lower mandate $\underline{r}^*$. In contrast, an increase in an ideals-projective paternalist's own choice c^A implies her ideal u^A is larger. This property, in turn, implies a belief that the Chooser's ideal option is higher, which leads to a higher prediction Q and a higher mandate $\underline{r}^*$. Finally, a conventional behavioral welfarist's inferences do not vary with her own type; moreover, whether higher predictions coincide with higher or lower optimal mandates depends on the correlation between beliefs about ideals and beliefs about mistakes.

Front-end delay with timed payoffs Turning to decision problems that involve the timing of payoffs, we now examine the effect of introducing front-end delay (that is, a fixed delay in the receipt of all payments) on paternalistic interventions. As is well-known, front-end delay induces people to choose more patiently for themselves (Frederick et al., 2002). One common interpretation is that they become less prone to making mistakes arising from 'present bias;' in effect, front-end delay ostensibly removes the lure of immediacy that is responsible for excessively impatient choice. Conventional behavioral welfarists who adopt this view will impose less restrictive mandates when Choosers confront decision problems with front-end delay.[16] Likewise, a mistakes-projective paternalist who construes her own behavior this way would assume that others are also less likely to make mistakes once front-end delay is introduced. By the logic of Proposition 1, she would therefore be *less* inclined to impose patience on Choosers in settings with front-end delay. In either case, if the Choice Architect believes that front-end delay eliminates the pertinent biases (as the well-known model of quasihyperbolic discounting implies), she will not intervene at all once it is introduced.

[16]The formal proof of this statement is essentially the same as for the portion of Proposition 1 concerning mistakes-projective paternalism. Here, adding front-end delay induces a shift in beliefs about the distribution of mistakes without altering beliefs about the distribution of ideal points.

Front-end delay may also influence the Choice Architect's conception of her own ideal. A Choice Architect may have in mind a relatively impatient ideal when contemplating decisions (for herself) without front-end delay, and a relatively patient ideal when contemplating decisions (for herself) with front-end delay. Possible reasons include that the Choice Architect is naïve and fails to recognize her own errors when making decisions with immediate consequences, or that she genuinely takes the view that the greater patience she displays in the presence of front-end delay reflects a difference in the applicable ideals. Either way, if the Choice Architect is also ideals-projective, then, by the logic of Proposition 1, she will be *more* inclined to impose patience on Choosers in settings with front-end delay.[17]

In addition, according to Proposition 2, mistakes-projective Choice Architects should believe that the optimal intervention has a smaller welfare benefit once front-end delay is introduced, while ideals-projective Choice Architects could believe that the benefit is either larger or smaller. Our aim is to determine which of these patterns predominate.

3 Experiment Design

Section 3.1 describes the main types of decision problems we use to investigate projective paternalism. Section 3.2 provides an overview of the structure of the experiment. The remaining sections then present details concerning incentivization (Section 3.3), the Choosers' decisions (Section 3.4), and implementation (Section 3.5). For easier readability, this section condenses the presentation of our design. A comprehensive description of all design details appears in Appendix D.1.

3.1 Main decision problems

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 main building blocks for our analysis consist of answers to the following three questions: (i) What opportunity sets do Choice Architects construct for Choosers? (ii) Do they believe that withholding options helps or hurts the Chooser? (iii) What options do they choose for themselves? We discuss each of these elements in turn.

Constructing 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 choice option is a bundle of two

[17] More formally, the preceding analysis assumes that the addition of front-end delay increases c^A, and that Assumption 2 holds with respect to variation in c^A not only across Choice Architects but also within Choice Architects and across contexts (that is, when comparing choice in intertemporal allocation problems with and without front-end delay). Under these assumptions, to the extent Choice Architects believe they are more susceptible to mistakes when payoffs are immediate, front-end delay decreases m^A. To the extent some of the increase in c^A reflects a change of the Choice Architect's own ideal, front-end delay increases u^A. Proposition 1 then implies that front-end delay will moderate interventions through the mistakes-projection channel, but will intensify interventions through the ideals-projection channel.

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 each option will be included in the opportunity set—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 make choices that reflect their genuine views.

We will study the nature of the opportunity sets Choice Architects construct. Four design features narrow the set of possible justifications for paternalistic behavior. First, we ensure that Choice Architects can only influence Choosers' outcomes, not their decision processes. For example, in our experiment, 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.

Second, Choice Architects can send messages to the Chooser. Absent this opportunity, Choice Architects might remove options simply to convey their advice, rather than because they perceive a genuine need for intervention. In each round of the experiment, Choice Architects can write a note to the Chooser, with no restrictions on content or length. The Chooser observes the note before making his decision. Choice Architects can also convey disapproval of any option simply by clicking a button. In that case, the Chooser sees a red asterisk next to the corresponding choice option, accompanied by a statement that a previous participant advises against ranking that option highly.[18]

Third, Choice Architects' decisions concerning Choosers' opportunity sets have no material consequences for themselves. This feature mirrors a fundamental 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.

Fourth, we examine decision problems for which there are no information asymmetries between the Choice Architect and the Chooser besides those involving preferences. Thus our analysis excludes the types of information asymmetries present in expert-client relationships that might offer additional rationales for intervention.

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 the nudge's efficacy. In our setting, efficacy is unambiguous. Second, we study paternalistic decisions by individuals, rather than

[18] An analysis of the resulting communication appears in Appendix B.1.

groups. While many paternalistic policies result from group decision making (e.g., through voting), the judgments of individuals are always central to those processes. Third, by using opportunity sets involving delayed monetary payments, our experiment introduces a plausible *a priori* rationale for paternalism: people commonly view patience as virtuous and impatience as a reflection of weakness. However, we are not primarily concerned with the study of intertemporal preferences *per se*. We simply take this setting as useful for studying paternalistic decision making.[19]

Elicitation of beliefs about welfare effects A Choice Architect's decision to withhold options is paternalistic only if she believes it furthers the Chooser's own good (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 the opportunity set consisting of all three options on the left,[20] along with a subset of those options on the right. In some rounds, the subset on the right is the one the Choice Architect constructed; in other rounds it is given exogenously (specifics depend on condition, as explained below).

Choice Architects first answer a non-incentivized question that is easy to comprehend: *Which opportunity set is better for the future participant?* They select between *Opportunity Set Left*, *Both equal*, and *Opportunity 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 towards the Chooser, the elicitation is incentive-compatible. Each line of the list presents a binary choice of the form: *The payment of the future participant should be determined by ... (i) the opportunity set on the left, and the participant's completion payment will remain unchanged, OR (ii) the opportunity set on the right, and the participant's completion payment will be raised / lowered by €p*, with $p \in \{1, 0.5, 0.3, 0.1, 0.05, 0, -0.05, -0.1, -0.3, -0.5, -1\}$.[21] A Choice Architect who believes, for instance, that removing the least patient option improves the Chooser's well-being by €0.4 and is benevolent towards the Chooser will prefer the first option if $p = -0.5$, but will prefer the second option if $p = -0.3$. Generally, 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.[22]

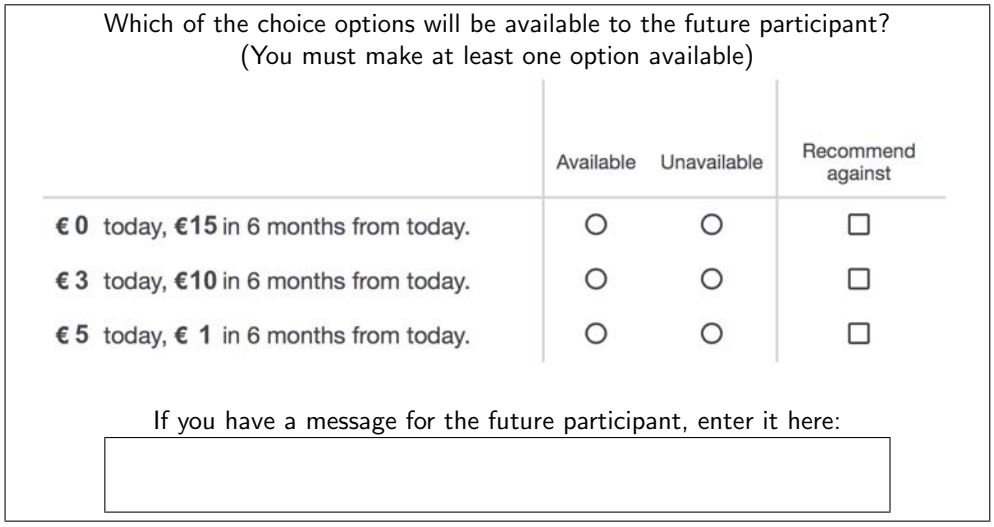
[19] Accordingly, questions about the extent to which choices involving money over time reveal intertemporal rates of substitution (e.g. Cohen, Ericson, Laibson and White, 2016; Andreoni, Gravert, Kuhn, Saccardo and Yang, 2018) are orthogonal to our paper. For completeness, Appendix B.2 presents data on Choice Architects' beliefs regarding the reasons why Choosers select impatient options. Explanations based on preferences and on liquidity constraints are rated as the most plausible. Explanations related to inattention, inability to choose the intended option, trust in the experimenter, and transaction costs are rated as implausible.

[20] The Exogenous Restriction condition (see Section 3.5) is an exception. In two of the four rounds of that condition, the opportunity set on the right consists of two options, while the one on the left consists only of the most patient option.

[21] It is worth highlighting two differences 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.

[22] A paternalist may believe that enlarging an opportunity set requires a positive compensating payment to the Chooser if the additional choice options create opportunities for errors.

A. Constructing the Chooser's opportunity set.



B. Beliefs about effect on Chooser's well-being

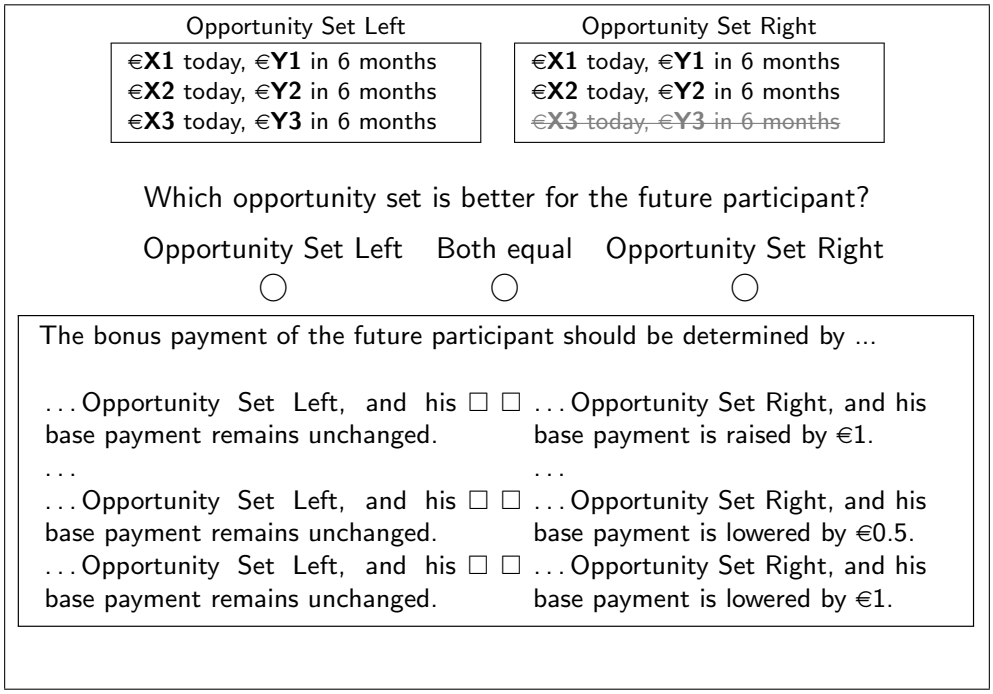


Figure 1: Decision screens for the Choice Architect in the Main condition. The Chooser is called a 'future participant.' Panel A: Constructing opportunity sets and communication. Panel B: Beliefs about effects on the Chooser's well-being.

We do not impose any restrictions on how subjects fill in these lists, such as monotonicity. A Choice Architect may express a preference that the Chooser obtain a higher rather than lower payment (if she is benevolent), that the Chooser obtain a lower rather than higher payment (if she is malevolent), or that the Chooser obtain a higher payment only up to some threshold.[23] The elicitation is incentive-compatible regardless of how the Choice Architect feels about the Chooser – e.g., whether she is benevolent or malevolent.

Choice Architects' own choices We test how Choice Architects decide which options are good for others by comparing the opportunity sets they construct for Choosers to the choices they make for themselves. Therefore, all Choice Architects complete six decision lists, such as the one shown in Figure 2. Each decision on each list is a choice between receiving €x_{early} the day of the experiment, or €x_{late} in t months after the experiment.[24] Subjects complete these lists in an online session three to six days before the laboratory session.

To ensure that our results are not artifacts of anchoring or a demand for consistency, we make it difficult for our subjects to relate the options they select for themselves to the options they make available to Choosers. Specifically, while the decisions they make for themselves hold the monetary amounts constant and vary the delay, the potential options for the Choosers keep the delay fixed and vary the monetary amounts. We also limit the potential influence of confounding mechanisms by requiring Choice Architects to make choices for themselves in an online session three to six days before they construct Choosers' opportunity sets in the laboratory, and by interspersing them with decision lists involving risk taking to obfuscate their purpose.[25] While these decisions involve no immediate payoffs, Choice Architects do make decisions with immediate payoffs during the laboratory session (see Section 3.5).

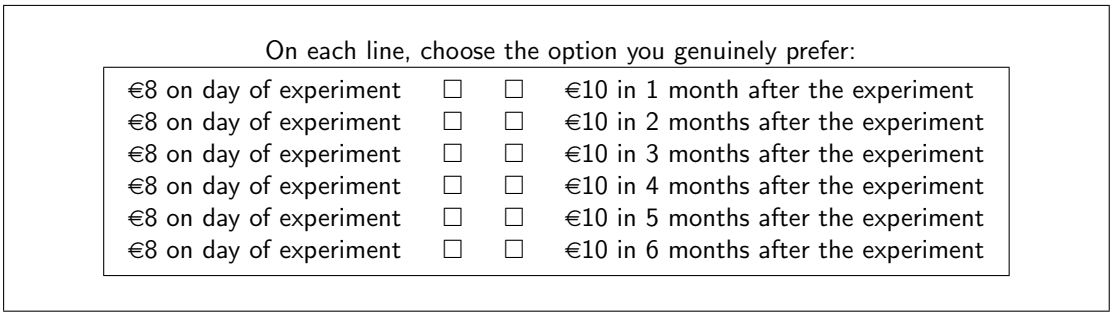


Figure 2: Choice Architects' own intertemporal choices.

[23] Benevolent Choice Architects will choose *(ii)* for high transfers and *(i)* for low transfers. Malevolent Choice Architects will display the opposite pattern, choosing *(i)* for high transfers and *(ii)* for low transfers. Choice Architects who want the Chooser to obtain a higher payment only up to a threshold choose *(i)* for low transfers and for transfers that exceed the threshold, but will choose *(ii)* for transfers of intermediate magnitude.

[24] We use $(x_{early}, x_{late}) \in \{(2,10),(5,10),(8,10),(2,15),(7,15),(12,15)\}$.

[25] For the decision lists involving risk taking, subjects decide between winning amount y with probability p and 0 otherwise, or a sure amount $z \in \{1,2,3,4,5,6\}$. We use $(y,p) \in \{(10,0.5),(6,0.7),(8,0.5),(5,0.8),(20,0.2),(13,0.3)\}$.

3.2 Conditions, timing and additional elicitations

The experiment consists of an online component and a laboratory component. The online component serves to elicit Choice Architects' own intertemporal choices using the six decision lists described above.

The laboratory component consists of three stages. Stage 1 includes 14 rounds of paternalistic decisions. The *Main* condition comprises four rounds, each of which proceeds as described above. These rounds employ the option menus shown in Table 1. To examine the effect of front-end delay, we randomly select either menu 1 or 2 and delay both the early and the late payment for each option in that menu by one week. Each Choice Architect also participates in three additional conditions for which we alter the decision problems described above to test specific hypotheses about the mechanisms underlying projective paternalism. We will describe these conditions when we detail the hypothesized mechanisms, as they come up in Section 4.

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, as well as in 6 additional rounds, the comparator subset is the one the Choice Architect has constructed herself. In two other rounds, that subset either includes only the most patient option, or only the most patient and middle options; we randomize between these two possibilities at the subject level. Finally, in two rounds, the Choice Architect compares a two-option opportunity set to an opportunity set consisting of the most patient option alone.[26]

We collect additional data in Stage 2 of the laboratory component. First, Choice Architects make surrogate choices for Choosers. For these decisions, which are otherwise identical to those in Stage 1, we require Choice Architects to construct opportunity sets consisting of a single choice option. Second, we administer a test to determine whether subjects recall 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 from the study. The purpose of the test is to incentivize subjects to pay attention.[27] Third, we ask Choice Architects to adjust the completion payment of a Chooser other than the one affected by their paternalistic decisions. They can either 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. The remaining elicitations are, like our supplemental conditions, designed to test specific hypotheses. We will describe them in conjunction with those hypotheses in Section 4.

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 infor-

[26] These decisions concern the option menu consisting of the following bundles of immediate and delayed payments: $(0, 15), (4, 6), (5, 1)$.

[27] Appendix 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.

mation to evaluate the generalizability of our findings. The experiment ends with a brief memory check on choices subjects made in the Online component.

Table 2 provides a schematic overview of the experiment from the Choice Architect's perspective. Appendix D.1 presents comprehensive detail about all design elements.

Option menu 1		
Option	today	in 6 months
Most patient	€0	€15
Middle	€3	€10
Least patient	€5	€1

Option menu 2		
Option	today	in 6 months
Most patient	€0	€15
Middle	€3	€9
Least patient	€6	€1

Option menu 3		
Option	today	in 6 months
Most patient	€0	€15
Middle	€2	€12
Least patient	€4	€2

Option menu 4		
Option	today	in 6 months
Most patient	€0	€15
Middle	€2	€9
Least patient	€4	€2

Table 1: Menus of options from which Choice Architects construct opportunity sets in the Main condition.

Online component	
Decision lists on intertemporal and risky choice	
Laboratory component	
Stage 1	Decisions concerning Choosers
	Each of 14 rounds: 1. Constructing Choosers' opportunity sets 2. Revelation of welfare beliefs
Stage 2	Additional elicitations (detailed later)
Stage 3	Policy judgments

Table 2: Schematic overview of the experiment.

3.3 Incentives

Choice Architects' decisions concerning Choosers. Because our experiment involves four times as many Choice Architects as Choosers, there is a 25% chance that we will match any given Choice Architect with a Chooser. For those who are matched, we randomly draw one of the rounds in which the Choice Architect makes a decision concerning the Chooser. With 50% probability, we implement the opportunity set specified by the Choice Architect in the first half of that round. With the remaining 50% probability, we determine the Chooser's opportunity set by randomly drawing a line from the decision list in the second

half of the round and implementing the Choice Architect's selection. Separately, with 25% probability, we implement the Choice Architect's decision concerning the completion payment for another randomly assigned Chooser. We do not match any Choice Architect with the same Chooser for both purposes. Choice Architects know that Choosers will participate in a subsequent laboratory session. We inform Choice Architects of the matching and implementation probabilities described above, 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 the choice problems that determine their own payment have not been manipulated by any subject.

Choice Architects' own payment. A Choice Architect's own payment is determined either by the online component, or by one of three supplemental task blocks in the laboratory component, each with 25% probability.[28] If the attention test is selected, the Choice Architect receives €1 for each correct answer on the eight questions. If another task is selected, her payment is determined according to one randomly selected decision within the task block. In addition, each Choice Architect receives a showup-payment of €4.5 and a completion payment of €8.[29]

3.4 Choosers

Choosers participate in separate sessions after all Choice Architect sessions are completed. Each Chooser ranks one set of three options according to his preferences. He receives the option he has ranked most highly among those the Choice Architect matched to that Chooser has made available. In addition, each Chooser receives a showup-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 Chooser's completion payment based on the Choice Architect's decision.

3.5 Implementation

We conducted the experiment with 303 subjects in the role of Choice Architect at the Cologne Laboratory for Economic Research. We recruited an additional 100 subjects for the Choice Distribution Information condition described in Section 5.2. Separately, 124 subjects participated as Choosers in subsequent sessions.[30]

[28] Subjects learn at the beginning of the online component that there is a 25% chance their payment will be determined entirely by a single decision from the online component, and a 75% chance it will be determined by the laboratory component, but they do not learn at the outset what the latter component will entail.

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

[30] We ran 16 sessions between June 14 and July 4, 2018, that lasted for approximately 90 minutes each. In addition, after we completed the Choice Architects' sessions, we conducted five sessions with subjects in the role of Chooser, who made a single choice each. Appendix D.2 provides details, and Appendix B.4 details Chooser choices. Prior to fielding our laboratory experiment, we conducted pilot experiments on Amazon Mechanical Turk. Data from each pilot are consistent with our conclusions concerning projective paternalism.

The experiment is computer-based. We display all instructions on-screen, and intersperse comprehension checks which subjects must complete correctly in order to continue. We refer those failing the comprehension checks back to the instructions, and then give them additional chances to pass (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. The invitation email informs subjects of this feature and asks them to open a PayPal account if they do not already have one.[31]

4 Paternalistic interventions

4.1 Restrictions on opportunity sets

We begin by studying the characteristics of the opportunity sets Choice Architects construct, focusing on the Main condition, and on rounds without front-end delay. Figure 3 shows the frequencies with which Choice Architects leave specific types of options unavailable, averaged across rounds. The prevailing tendency is for Choice Architects to prevent impatient choice. They remove the least patient option 33% of the time, the middle option 11.5% of the time, and the most patient option 5.1% of the time. When we limit attention to the 86.3% of Choice Architects who are altruistic in the sense that they choose to costlessly increase a second Chooser's completion payment by €1 in Stage 2, the frequency with which they remove the most patient option in the Main condition drops by half (to 2.5%). The removal frequency for the middle option falls slightly (to 10.6%), and there is no change for the least patient option.[32] The pattern of withholding impatient options and making patient options available occurs within each of the four option menus in the Main treatment (see Appendix B.5 for details).

Imposing a minimum degree of patience is also the modal behavioral pattern on the individual level: 44.9% of our subjects remove at least one option from at least one of the opportunity sets, and never remove an option without also excluding less patient options.[33] Choice Architects who never remove any option in the Main condition comprise the second largest category (38.3%). We refer to this group as

[31] PayPal accounts are free. Subjects received exactly the amount of money promised in the experiment; all transaction fees were paid by the researchers.

[32] Of all Choice Architects, 86.3% increase the second Chooser's completion payment, 11.2% leave it unchanged, and 2.5% reduce it.

[33] With only three options, a Choice Architect can violate this monotonicity condition only if she blocks the middle option in at least one round of the Main condition. Of the Choice Architects who block a least patient option, 26% block a middle option at least once. There are several potential explanations for the observation that Choice Architects tend to impose patience without limiting it, but this experiment is not designed to disentangle them. Choice Architects may simply view patience as virtuous, or they may balance the cost of removing the best options for some Choosers against the benefit of blocking options that others may select in error. In the latter case, the optimality of a lower bound on patience depends on the joint distribution of ideals and mistakes. As Lemma 1 shows, a Choice Architect will impose only a lower bound on patience if she believes that mistakes only lead to overly impatient choice.

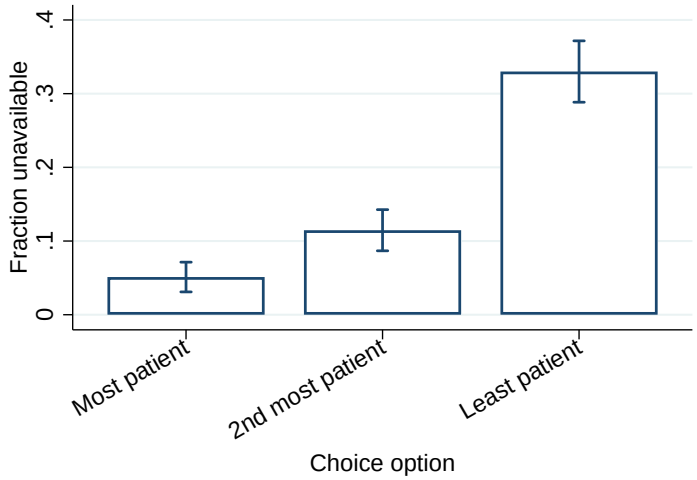


Figure 3: Frequency of withheld choice options, by type, averaged across the four option menus in the Main condition, excluding opportunity sets with front-end delay. Whiskers display 95%-confidence intervals with estimates of standard errors clustered at the subject level.

Libertarians. A small fraction (4.3%) of Choice Architects impose an upper bound on patience.[34] The remaining 12.5% of Choice Architects impose non-monotonic restrictions.[35]

4.2 Are interventions benevolently motivated?

The removal of a choice option is paternalistic only if the Choice Architect believes that it benefits the Chooser (Dworkin, 1972). On average, our Choice Architects hold this belief. Panel A of Figure 4 displays Choice Architects' mean response to the question of whether the full opportunity set, the opportunity set they have constructed, or neither is better for the Chooser, separately for those who make one, two, and three options available. Choice Architects who withhold options indicate a strong belief that their intervention benefits the Chooser, and the belief is stronger for those who withhold more options. Panel B shows that the same relationship holds for beliefs about compensating variations, the elicitation of which is incentive-compatible.

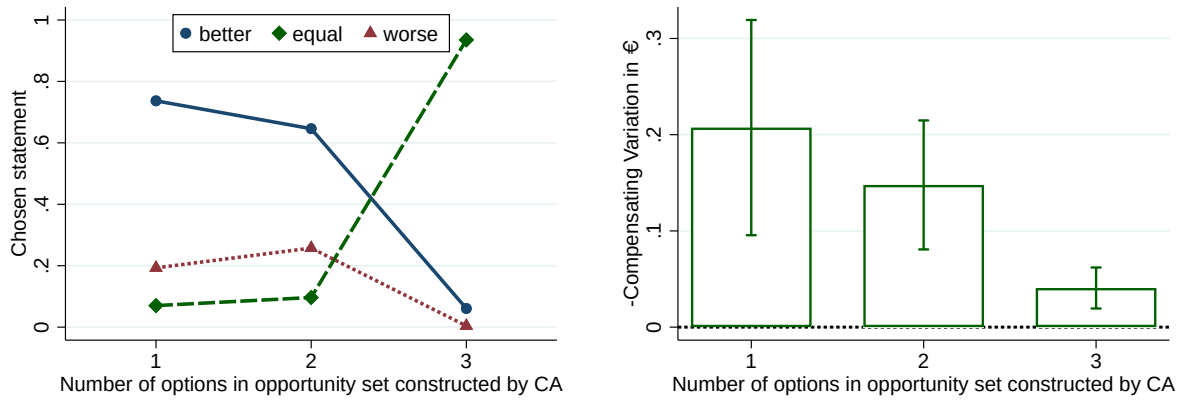
We formalize these comparisons by performing the regressions reported in Table 3. Column 1 is an ordered probit. The dependent variable measures whether the Choice Architect considers the opportunity set she has constructed better, equally good, or worse for the Chooser than the unrestricted set. The independent variables include indicators for the number of options removed (one or two), and for whether

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

[35] Among subjects who prevent impatient options, 89.7% elect to costlessly increase a second Chooser's payment; only 0.7% decrease the payment. Among Libertarians, the corresponding numbers are 92.2% and 0.0%. In contrast, among those who prevent patient options, only 46.2% of subjects behave altruistically, and 23.1% behave spitefully. For those who violate monotonicity, these percentages are 68.4% and 13.2%, respectively. Accordingly, genuine paternalism (intervention motivated by concern for the Chooser's well-being) is probably more common among Choice Architects who enforce patience than among those who intervene in some other fashion.

Dependent variable	(1) Belief smaller opportunity set better for Chooser	(2)	(3) Negative of Compensating Variation	(4)
Smaller set	Endogenous	Exogenous	Endogenous	Exogenous
Method	Ordered probit	Ordered probit	Interval regression	Interval regression
Effect of number options removed				
1	0.634***		0.094***	
	(0.144)		(0.036)	
2	1.035***		0.156**	
	(0.247)		(0.062)	
Mean # options removed		0.926***		0.107**
in Main condition		(0.187)		(0.045)
No options removed		-0.287		-0.013
in Main condition		(0.202)		(0.051)
Reduced opportunity set consists		0.317**		0.006
of most patient option only		(0.139)		(0.041)
Social preferences				
Altruist	0.193	0.133	0.038	0.078
	(0.194)	(0.223)	(0.036)	(0.063)
Spiteful	-0.957*	-1.238***	0.297	-0.175
	(0.537)	(0.361)	(0.239)	(0.239)
Cut 1	-0.535*	0.888**		
	(0.307)	(0.452)		
Cut 2	1.385***	1.239***		
	(0.310)	(0.455)		
Mean of dependent variable	-	-	0.079	-0.022
			(0.015)	(0.020)
Observations	909	606	817	475
Number of subjects	303	303	289	254

Table 3: Subjects' beliefs about the welfare effects of withholding options. Each column represents a separate regression. 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. For columns 3 and 4, it is the negative of the Choice Architect's beliefs about the compensating variation of reducing the opportunity set. The dependent variables in columns 1 and 3 reflect comparisons between the maximal opportunity set and the one the Choice Architect has constructed. In columns 2 and 4 it reflects comparisons to opportunity sets from which we exogenously remove the (middle and) least patient option(s). We include the two rounds from the Exogenous Removal condition in which the larger choice set contains three options. All regressions include session, order, and option menu fixed effects. Numbers in parentheses indicate standard errors, clustered by subject. $^{*}p < 0.1,^{**} p < 0.05,^{***} p < 0.01$.



(A) (B)

Figure 4: Beliefs about the effect of withholding choice options on the Chooser's well-being. Panel A: Statements regarding whether the Choice Architect considers the opportunity set she has constructed better, equally good, or worse for the Chooser, respectively. Panel B: Incentive-compatibly elicited beliefs about the (negative) compensating variation of withholding options. Whiskers indicate 95% confidence intervals, clustered by subject.

the Choice Architect has chosen to costlessly increase a second Choosers' completion payment by €1 (altruist) or decrease it by €1 (spiteful). We include session, order, and option menu fixed effects, and cluster standard errors by subject. On average, Choice Architects who remove a single option have stronger beliefs that their action benefits the Chooser than those who make all options available, and the difference is statistically significant. For Choice Architects who remove two options, the estimated coefficient is even larger, although the increment is only marginally significant ($p = 0.11$). We also see that spiteful Choice Architects believe much less strongly that their action benefits the Chooser.[36]

Although we are primarily interested in Choice Architects' beliefs about the welfare effects of the opportunity sets they themselves construct, the endogeneity of those sets potentially introduces bias. Accordingly, we also examine opportunity sets from which we have removed impatient options exogenously. Column 2 of Table 3 exhibits an ordered probit regression in which the dependent variable measures whether the Choice Architect considers the exogenously restricted opportunity set better, equally good, or worse for the Chooser than the unrestricted set. We drop the indicators for the number of options removed, and add three new variables: the mean number of options removed in the Main condition, a variable that indicates whether the subject removed no options in the Main condition, and a variable indicating whether the exogenously reduced opportunity set consists only of the most patient option. According to the regression, Choice Architects who remove a larger number of options in the Main condition have significantly stronger beliefs that the exogenous removal of impatient options is good for the Chooser.

[36] We caution that this estimate is based on only nine spiteful subjects.

Columns 3 and 4 replicate columns 1 and 2 using dependent variables based on our second measure of welfare effects: beliefs about the compensating variation of removing impatient options, elicited through an incentive-compatible procedure. Here we use interval regression, including only those subjects whose choices in the multiple-decision lists are consistent with preferring that the Chooser receive a higher payment over a lower payment. We replicate the pattern of results obtained with non-incentivized beliefs. We conclude that Choice Architects' behavior is, for the most part, benevolently motivated, and therefore consistent with paternalism.

4.3 Other potential motives

Paternalism is not the only possible motivation for intervention. Various alternatives merit consideration. One possibility is that subjects perceive the imposition of a restriction as an active choice and laissez faire as a passive choice, and prefer to take action ("keep busy") rather than do nothing. Formally, the Choice Architect may experience a fixed utility bonus when picking something other than a default opportunity set. Another possibility is that subjects restrict opportunity sets out of a desire to exert control by taking actions that limit Choosers' options. Formally, the Choice Architect may experience a utility bonus when picking something more restrictive than a default opportunity set.[37] Some Choice Architects may also have weak preferences and randomize among the available alternatives. While all of these alternative hypotheses can account for the observation that Choice Architects frequently restrict Choosers' options, there is no reason to think they would give rise to the particular pattern of patience-promoting interventions and the associated beliefs about welfare effects documented above. To explore these alternatives in greater depth, we include two additional conditions. Each of these conditions comprises several rounds, which are randomly positioned within the first stage of the laboratory component (see Table 2 for a schematic overview, and Appendix D.1 for comprehensive detail).

The Induced Chooser Preferences condition We designed this condition to suppress paternalistic motivations by removing reasons to take issue with the Chooser's objectives. (Some other potential reasons for paternalistic interventions, such as dampening the consequences of trembling-hand errors, remain.) Because it does not affect the other hypothesized motives, any change in behavior is presumably attributable to paternalism.

To understand the design of this condition, recall that choice options in the Main condition are monetary bundles (X, Y) with Y paid later than X. Assuming income effects are negligible for the stakes used in this experiment, we can write the Chooser's utility from such bundles as $X + \delta Y$, where δ is the Chooser's discount factor. For the Induced Chooser Preferences condition, we also present Choosers with menus of bundles of the form (X, Y), but pay them $X + rY$ the day of the experiment. Each

[37] See Pikulina and Tergiman (2018) for a targeted experimental investigation into preferences for exerting control.

Chooser learns the value of the parameter $r \in \{0, 0.5, 1\}$ before making a choice.[38] Choice Architects only learn the distribution of r—each value is equally likely and is drawn independently for each Chooser. We describe X and Y as gold and silver tokens, respectively, to be exchanged for Euros the day of the experiment, and we employ the values of X and Y used for option menus 3 and 4 in Table 1.

Averaging across rounds in the Main condition, the overall frequency with which Choice Architects exclude at least one option from an opportunity set is 39.8% (s.e. 2.3 percentage points). For the Induced Chooser Preferences condition, that frequency drops precipitously, to 22.6% (s.e. 2.1 percentage points). Not only are average removal rates lower, but the monotonic relationship between larger values of X and removal frequencies that characterizes the Main condition vanishes entirely. In particular, removal rates for the options in the Induced Chooser Preferences condition with the largest, middle, and smallest values of X are 12.9% (s.e. 1.6%), 5.1% (s.e. 1.1%), and 7.8% (s.e. 1.3%), respectively.[39] Moreover, Choice Architects are far less likely to believe that the opportunity to intervene benefits the Chooser. Across all rounds and Choice Architects in the Main condition, subjects indicate in 29.6% (s.e. 2.2%) of cases that the unrestricted opportunity set is worse for the Chooser. This number drops to 15.3% (s.e. 1.8%) in the Induced Chooser Preferences condition. Similarly, the elicited compensating variations are more than twice as large in the Main condition as in the Induced Chooser Preferences condition: 0.079 (s.e. 0.015) versus 0.038 (s.e. 0.011). Moreover, with induced preferences, the relation between the number of options removed and the Choice Architect's beliefs about the Chooser's welfare is not statistically significant.[40] These results strengthen our previous interpretation that most of the interventions observed in the Main condition, as well as the patterns of interventions, are attributable to paternalism.

The Exogenous Restriction condition The Induced Chooser Preferences condition refutes the hypothesis that Choice Architects intervene out of a general desire to restrict Choosers' opportunity sets. It leaves room, however, for a more nuanced version of that hypothesis: Choice Architects may wish to exert control as long as they can rationalize their actions as beneficial. Arguably, such rationalization is difficult in the Induced Chooser Preferences condition, and consequently we use the Exogenous Restriction condition to test this hypothesis. This condition consists of four rounds. In one, Choice Architects decide between making all options available, removing the least patient option, and removing the two least patient options. In another, we exogenously remove the least patient option. The Choice Architect

[38] We chose these parameter values to minimize complexity. Another parameterization strategy would have been to match the distribution of r to the representative Choice Architect's subjective distribution of δ. Unfortunately, that subjective distribution is difficult to measure, and may differ from one Choice Architect to the next.

[39] The higher removal rate for the first option may reflect a variety of motives, including fairness concerns. A Chooser facing $r = 0$ can obtain no more than €4 in the Induced Chooser Preferences condition. A Chooser facing $r = 1$, in contrast, obtains €15 from choosing the first option. By removing that option, the Choice Architect can reduce inequality across Choosers.

[40] We estimate regressions that parallel those in columns 1 and 3 of Table 3. Using the non-incentivized measure, we find a coefficient of 0.32 (s.e. 0.20) for removing one option, and a coefficient of -0.18 (s.e. 0.51) for removing two options. With the incentivized measure, the corresponding parameter estimates are 0.06 (s.e. 0.04) and 0.14 (s.e. 0.12). None of these parameter estimates are statistically significantly different from zero at the 10% level.

must decide between offering this reduced opportunity set and removing the middle option.[41] (We describe the other two rounds below.) Under the hypothesis that Choice Architects are solely concerned with the consequences of their actions for Choosers, the exogenous removal of the least patient option should not affect the availability of the middle option. In contrast, under the alternative hypothesis, the availability of the middle option should decline. To see why, 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 control. To exert control, the Choice Architect must now remove the middle option. Because people typically perceive patience as virtuous, the Choice Architect can rationalize her intervention as beneficial.[42]

Indeed, we find that the availability frequency for the middle option falls slightly from 89.7% when the least restrictive opportunity set contains three options, to 85.8% when we remove the least patient option exogenously. However, because the number of possible responses differs across these two rounds, the reduction in the availability of the middle option could be attributable to subjects who randomize among their alternatives. To control for the effects of random choice, we include two additional rounds that are identical to the first pair, except that the default is the most restrictive opportunity set (i.e., just the most patient option), rather than the least restrictive set. By taking action, the Choice Architect adds rather than removes options. Random choosers will add options, but subjects who are motivated by the desire to restrict others' opportunity sets will not.

In these two rounds, we find that the exogenous removal of the least patient option causes a similar decline in the availability of the middle option (76.2% and 73.9%). The difference-in-differences is far from statistically significant in a regression with session and order fixed effects ($p > 0.6$). While the statistical power of this test is limited, we conclude that there is no evidence indicating that Choice Architects are primarily motivated by a desire to exert control, even when they can potentially rationalize such action as beneficial.[43]

5 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

[41] In each one of these four rounds, we use the same three bundles of immediate and delayed payments, $(0, 15), (4, 6), (5, 1)$.

[42] The literature on pure versus impure altruism in charitable giving, starting with Andreoni (1993), tests similar implications (see Vesterlund, 2016, for a review).

[43] This analysis is based on within-subject variation. The results do not change qualitatively if we restrict attention to cross-subject variation by only including each Choice Architect's first decision from among these four rounds. The resulting estimates of the decline in the frequency with which Choice Architects offer the middle option are virtually the same regardless of whether we specify the most restrictive set or the least restrictive set as the default option: it falls from 89.0% to 81.7% in the first instance, and from 78.3% to 71.1% in the second ($p > 0.9$ for the difference-in-differences).

to rely on the judgments implicit in Choosers' decisions, and indeed may even question whether Choosers are aware of their own best interests.

In Section 2, we differentiated between three varieties of paternalism: conventional behavioral welfarism, mistakes-projective paternalism, and ideals-projective paternalism. In all three cases, a paternalist holds beliefs about the distribution of ideals and biases, and maximizes the Chooser's expected well-being. A conventional behavioral welfarist either understands the true distribution of behavioral proclivities or forms a belief about it based on information concerning the Choosers. In contrast, the two forms of projective paternalism portray Choice Architects as arriving at their interventions by considering how they themselves would act in the Chooser's position.[44] A mistakes-projective paternalist views herself as prone to particular decision errors, makes the assumption that the Chooser shares that susceptibility, and intervenes to remedy it. An ideals-projective paternalist assumes that others either share or ought to share her ideals, and intervenes to promote them. We have seen that mistakes-projection induces a *negative* relation between the options 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 the relation between Choice Architects' interventions and the choices they make for themselves. For each round of the Main condition, we construct a variable measuring the Choice Architect's *mandate*, defined as the smallest delayed payment a Choice Architect forces the Chooser to receive.[45] For example, if a Choice Architect offers the most patient and middle alternatives from option menu 4 of Table 1, the mandate is €9. We also construct a measure of the degree of patience 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.[46]

Figure 5, which excludes the 38% libertarian subjects, 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 exceed those of the least patient Choice Architects by about €3. This difference is almost a quarter of the average gap between the late payments for the most patient and least patient options in our option menus (€13.50). According to Proposition 1 in Section 2, this pattern points to ideals-projective paternalism.

[44] Because we focus on behavior rather than cognitive processes, these are technically "as-if" hypotheses.

[45] Our results are robust with respect to alternative definitions of this variable, such as the maximal amount of money the Choice Architect allows the Chooser to receive early.

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

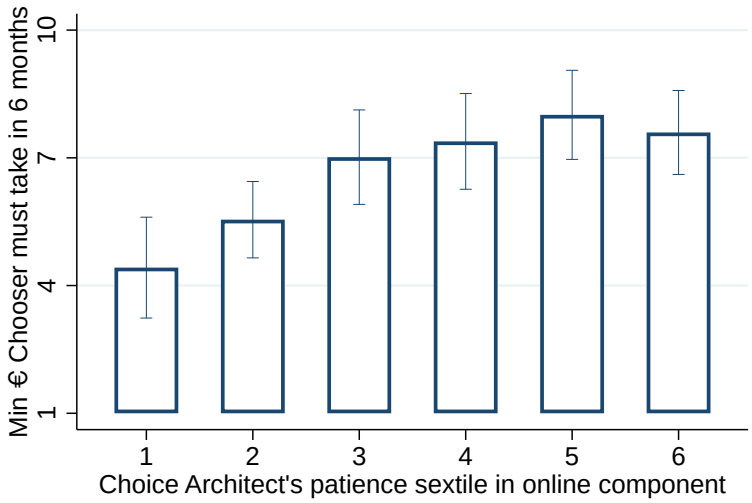


Figure 5: Ideals-projective paternalism. The figure excludes subjects classified as libertarian.

We formalize these observations by regressing the Choice Architect's mandate on the their patience percentile. The unit of observation is a single intervention in a single round. We use data from the Main condition, control for session, order, and option menu fixed effects, and cluster standard errors by subject. For the regression in column 1 of Table 4, we exclude libertarians. As above, we find that the mandate increases by €3.33 as patience moves from the lowest percentile to the highest, and the effect is highly statistically significant. Column 2 shows that the effect is smaller (€1.90), but still highly statistically significant, once all subjects, including the libertarians, are included. This attenuation reflects the fact that a Choice Architect's patience percentile does not predict whether she is libertarian (column 3).

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 *surrogate choices* from rounds of the Main condition in which we require Choice Architects to select a single item for the Chooser from the same menus as before (Stage 2 in Table 2). These surrogate choices force Choice Architects to intervene, and therefore directly reveal their 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 as strong as 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, except for their their willingness to intervene. According to Corollary 1 in Section 2, the positive relation between surrogate choices and own choices is consistent only with ideals-projective paternalism; for conventional behavioral welfarists and for mistakes-projective paternalists, there would be no relation.

Finally, columns 6 and 7 show that more patient Choice Architects not only impose stricter restrictions, but also hold more positive beliefs about the welfare effects they cause. According to Proposition 2, this pattern is inconsistent with mistakes-projective paternalism (intuitively, greater patience implies smaller mistakes, which implies lower benefits from restrictions), but it is consistent with ideals-projective paternalism.

VARIABLES	(1) Min. € Chooser takes late	(2) Min. € Chooser takes late	(3) Libertarian	(4) Surrogate choice	(5) Surrogate choice	(6) Welfare belief non-inc.	(7) Welfare belief - CV
Non-libertarian subjects only	Yes			Yes		Yes	Yes
Method	OLS	OLS	Probit	OLS	OLS	Ordered probit	Interval regression
Patience %-ile	3.328***	1.901***	0.022	3.974***	4.107***	0.858***	0.169**
	(0.843)	(0.654)	(0.235)	(0.678)	(0.519)	(0.219)	(0.067)
Cut 1						-0.067	
						(0.353)	
Cut 2						1.141***	
						(0.358)	
Mean of dep. var.	6.845	4.862	0.383	12.960	13.020	2.282	0.100
	(0.280)	(0.225)	(0.028)	(0.229)	(0.179)	-	(0.022)
Observations	537	873	873	518	837	537	471
Number of subjects	179	291	291	179	291	179	171

Table 4: Relationship between Choice Architect's mandates (in the Main condition) and their patience percentiles (in the online component). We exclude subjects who responded non-monotonically to any of the multiple-decision lists in the online component. We also exclude decisions with front-end delay. Columns 6 and 7 use the two rounds of the Exogenous Restriction condition in which the subjects compare a three-option opportunity set to a smaller opportunity set. The number of observations in column 7 is smaller than in column 6 because the former excludes observations with multiple switching points in the choice list eliciting CV. The number of observations for surrogate choices is smaller because some of these choices were not recorded in the first two sessions. All regressions include session, order, and option menu fixed effects. Numbers in parentheses indicate standard errors, clustered at the subject level. $^{*}p < 0.1$,$^{**}p < 0.05$,$^{***}p < 0.01$.

5.1 Predictions of Choosers' selections and the false consensus effect

We now study the relation between Choice Architects' mandates and their predictions of Choosers' selections. This relation is interesting for two reasons. First, according to Proposition 3, it provides another basis for distinguishing between mistakes-projective and ideals-projective paternalism. Second,

we can compare predictions to the true distribution of choices. As we will see, this comparison links projective paternalism to a well-known bias, the false consensus effect.[47]

We elicit Choice Architects' beliefs about the distribution of choices from each option menu made by ten previous subjects who faced no constraints.[48] Subjects drag and drop ten tags labelled "Participant" into three bins representing the choice options, as shown in Figure 6. (Subjects perform these tasks in Stage 2 of the laboratory component; see Table 2 for a schematic overview, and Appendix D.1 for comprehensive detail). There is a 25% chance that the Choice Architects' payment is determined by one round randomly selected from this belief elicitation stage.[49] In that case, she receives €10 minus the number of tags we must reassign to make the elicited distribution coincide with the observed distribution of choices for that round.[50] For our econometric analysis, we use the elicited information to construct the Choice Architects' beliefs about the mean delayed payment selected by unrestricted subjects.

For Column 1 of Table 5, we regress Choice Architects' mandates on their beliefs, focusing on non-libertarian subjects. A €1 increase in beliefs about the mean selection is associated with a €0.57 increase in mandates. According to Proposition 3, the direction of this relationship is consistent with ideals-projective paternalism, but not with mistakes-projective paternalism. Column 2 performs the same regression for the entire sample, including libertarian subjects. The magnitude of the coefficient declines to 0.34, but is still highly statistically significant.

Next we investigate the connection between ideals-projective paternalism and the false consensus effect. First we regress our measure of Choice Architects' beliefs on their patience percentiles, controlling for session, order, and option menu fixed effects, clustering standard errors by subject. The estimates in Column 3 show that, compared to the least patient Choice Architect, the most patient Choice Architect predicts that Choosers will opt for an additional €3.57 with a half-year delay. By Proposition 3, this manifestation of the false consensus effect is consistent with either form of projective paternalism.

Next we document an *individual-level* relation between false consensus bias and projective paternalism. Let r_i denote Choice Architect i's patience percentile from the online component (averaged across decision lists), and let b_i denote i's beliefs about the mean delayed payment unrestricted Choosers would elect to receive (averaged across rounds). Using $\bar{r}$ and $\bar{b}$ to denote the population means of r_i and b_i, we measure i's susceptibility to false consensus bias as $m_i = \frac{b_i - \bar{b}}{r_i - \bar{r}}$. We partition the pool of Choice Architects into four

[47] In Section 5.2, we address the possibility that a false consensus effect arises due to the lack of information concerning Choosers.

[48] For Choice Architects in sessions 2 and onwards, we use the unrestricted choices that Choice Architects in session 1 had made when choosing from the unrestricted option menus in Stage 2 (see Table 2). For Choice Architects in the first session, we calibrated the benchmark choices according to choices made in online pilots.

[49] Half of our subjects, chosen at random, complete these elicitations before Stage 1. All our regressions control for the ordering of these tasks.

[50] Our elicitation procedure is the balls-in-bins method described in Delavande, Giné and McKenzie (2011). Formally, if the subject places x_i workers in bin i, for $i = 1, \ldots, n$, and the true distribution of choices is $y_1, \ldots, y_n$, then the subject's payoff is given by $10 - \sum_{i=1}^{n} |y_i - x_i|$. Truthful revelation is optimal for a risk-neutral subject. Subjects understand this scheme more easily than alternatives. While risk aversion theoretically generates a tendency towards overly dispersed beliefs, Choice Architects' risk preferences, elicited in the online component, predict neither the location nor the dispersion of elicited belief distributions.

Figure 6: Belief elicitation. Participants drag and drop each of the tags labelled *Participant* into the bins corresponding to the choice options.

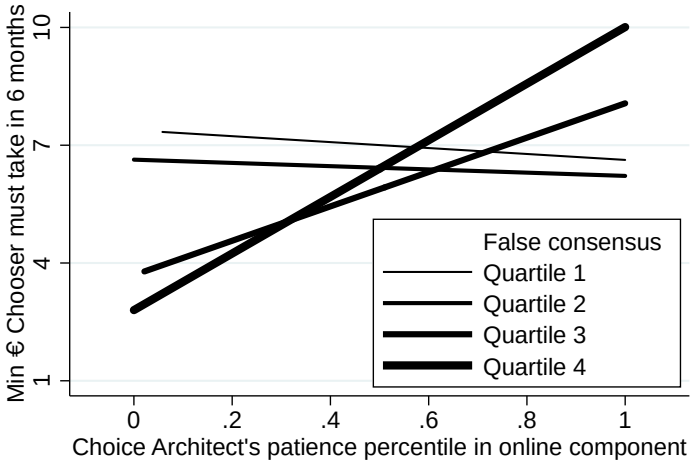


Figure 7: Projective paternalism by false consensus bias quartile. The figure excludes subjects classified as libertarian.

quartiles based on m_i, and study the relationship between mandates and patience percentiles separately within each quartile.

VARIABLES	(1) Min. € Chooser takes late	(2) Min. € Chooser takes late	(3) Mean belief	(4) Min. € Chooser takes late	(5) Surrogate choice	(6) Min. € Chooser takes late	(7) Surrogate choice
Non-libertarian subjects only	Yes						
Method	OLS	OLS	OLS	OLS	OLS	OR-IV	OR-IV
Mean belief	0.572***	0.340***					
	(0.088)	(0.076)					
Patience %-ile			3.566***				
			(0.371)				
Patience %-ile							
× *false consensus Q1*				-1.437	1.476	-0.733	2.079
				(1.561)	(1.072)	(4.011)	(2.116)
× *false consensus Q2*				0.114	1.066	1.130	-0.245
				(1.020)	(0.831)	(2.938)	(2.065)
× *false consensus Q3*				3.677***	4.897***	3.355	5.580*
				(1.161)	(0.829)	(3.458)	(3.042)
× *false consensus Q4*				4.389***	10.630***	5.587**	11.55***
				(1.599)	(1.194)	(2.488)	(1.369)
False consensus quartile fixed effects				Yes	Yes	Yes	Yes
Mean of dep. var.	6.845	4.862	11.32	4.893	13.05	4.893	13.05
	(0.280)	(0.225)	(0.137)	(0.231)	(0.181)	(0.231)	(0.181)
Observations	561	909	873	873	837	873	837
Number of subjects	187	303	291	291	291	291	291

Table 5: Projective paternalism and beliefs about Choosers' unrestricted selections. We exclude subjects who responded non-monotonically to any of the multiple-decision lists in the online component. All regressions control for option menu fixed effects. Columns 1 to 5 also control for session and order fixed effects. Numbers in parentheses indicate standard errors of the estimates, clustered by subject. The number of observations for surrogate choices is smaller because some of these choices were not recorded in the first two sessions. The number of subjects in regressions involving Choice Architects' patience percentiles is smaller because we exclude subjects with multiple switching points in the choice lists eliciting patience. $^{*}p < 0.1,^{**}p < 0.05,^{***}p < 0.01$.

Figure 7 depicts the results graphically. Choice Architects whose susceptibility to false consensus bias falls within the first and second quartiles exhibit no projective paternalism. They are paternalistic, imposing average mandates of around €7, but these mandates are unrelated to their own patience. For Choice Architects in the third quartile, however, there is a strong relationship between mandates and patience percentiles. The relationship is even stronger for Choice Architects in the top quartile. Hence, the

greater a Choice Architect's false consensus bias, the more she projects her ideals onto others.[51] Column 4 of Table 5 presents estimates of the same model as column 1, except that we interact the Choice Architects' patience percentile with indicators for each of the four false consensus quartiles (while also adding quartile fixed effects). The difference between the mandates imposed by the most and least patient Choice Architects in the top false consensus quartile is significant both statistically and economically (€4.39); it is roughly a third of the €13.50 difference between the most and least patient options. While the difference between the coefficient estimates for the third and top quartiles is not statistically significant, both coefficients are significantly larger than those for the first and second quartiles ($p < 0.05$ for each pairwise comparison). Column 5 shows that the same pattern is present in Choice Architects' surrogate choices. Hence, false consensus bias affects a Choice Architect's judgments about what is good for others, and not merely her propensity to act on those judgments. The pairwise differences between the coefficients for the second, third, and fourth quartiles in column 5 are all statistically significant ($p < 0.01$ in each case).

A possible issue with the previous analysis is that subjects may differ according to the noisiness of their choices. OLS might thus yield inconsistent estimates of the relation between projective paternalism and false consensus bias. To address this issue, we make use of the fact that our experiment elicits all variables required to estimate the regressions shown in columns 4 and 5 multiple times, which allows us to apply OR-IV estimation (Gillen, Snowberg and Yariv, forthcoming).[52] As columns 6 and 7 show, the resulting estimates exhibit the same qualitative pattern as the OLS estimates.

The explanatory power of beliefs in accounting for paternalistic interventions raises the question of whether Choice Architects merely remove the options they expect no Chooser will select. Two pieces of evidence show this is not the case. First, if a Choice Architect believes that an option will never be chosen, she should also believe that the Chooser's well-being does not depend on that option's availability. In contrast, as demonstrated in Section 4.2, Choice Architects believe that withholding impatient options significantly increases Choosers' welfare. Second, elicited beliefs about Choosers' unrestricted choices show that Choice Architects frequently remove options they believe many Choosers would have selected. Among Choice Architects who removed the least patient option, 59.5% (s.e. 3.7%) believe Choosers would have selected it with positive probability (averaged across the four decisions in the Main condition). For Choice Architects who removed the middle and most patient option, the respective numbers are 63.8% (s.e. 6.1%) and 87.0% (s.e. 7.5%), respectively.

[51] There is no relationship between a Choice Architect's false consensus quartile and the likelihood we classify her as libertarian ($p > 0.4$ in a joint test with session and order fixed effects).

[52] Specifically, we calculate each Choice Architects' mean patience rank separately for the three decision lists involving a delayed payment of €10, and for those involving a delayed payment of €15. We also calculate subjects' beliefs concerning Choosers' unrestricted selections separately for option menus 1 and 2, and for option menus 3 and 4. This procedure yields two indices of false consensus bias for each Choice Architect that use no overlapping data. We interact each index with the Choice Architect's patience percentile for the choices that were used to calculate the index. Monte Carlo simulations show that, in some circumstances, OLS would spuriously suggest a relation between projective paternalism and false consensus bias even if none exists, while OR-IV would correctly detect the absence of a relation.

5.2 The role of information

Projective paternalism could potentially be a consequence of scarce information about the Chooser's preferences. We now study whether information provision attenuates projective paternalism, using the *Chooser Information* condition. It resembles the Main condition, except that Choice Architects can condition their intervention on information about their Chooser. Specifically, Choosers endorse one of the four statements listed in Table 6. These statements allow them to describe themselves as patient or impatient, and as generally happy or unhappy with their intertemporal choices. The condition consists of four rounds, one for each of the four statements, which we randomly intermingle with the other ten rounds in Stage 1 of the laboratory sessions (see Table 2).[53] If projective paternalism is simply a consequence of scarce information, this condition should attenuate the relationship between mandates and patience percentiles.

- 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).

Table 6: Statements for the Chooser Information condition.

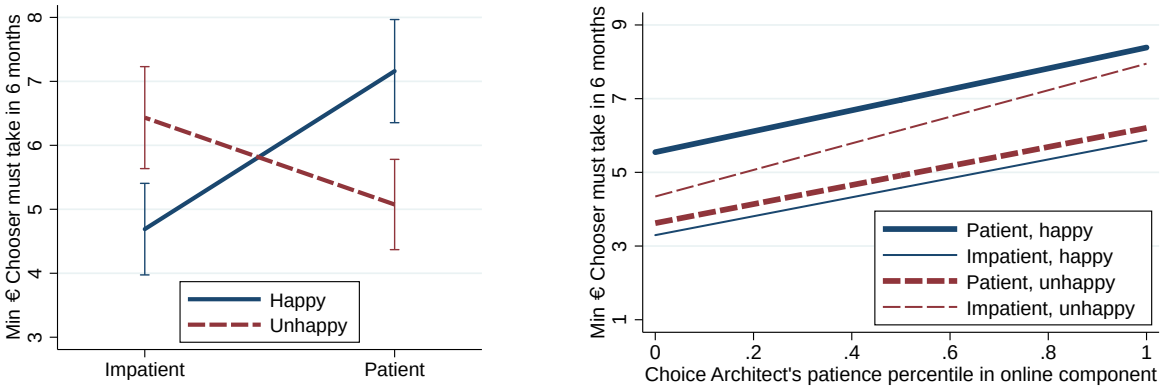
Panel A of Figure 8 shows that Choice Architects respond to this information. It displays average mandates by Chooser statement. For Choosers who claim to be generally happy with their choices, Choice Architects impose more patience when Choosers also claim that they generally behave patiently rather than impatiently. For Choosers who claim to be generally unhappy with their intertemporal choices, this relationship reverses. Despite the fact that Choice Architects treat these statements as informative and relevant to their intervention decisions, this information does not attenuate ideals-projective paternalism, as Panel B shows. More patient Choice Architects impose greater patience on Choosers, and do so to roughly the same degree, regardless of how the Chooser describes himself. We obtain these results even though, on average, Choice Architects believe that a substantial fraction of Choosers falls into each of the four categories.[54]

[53] Each of these rounds concerns the following three bundles of immediate and delayed payments: $(0, 15), (3, 7), (4, 1)$. 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.

[54] On average, Choice Architects believe that 31.8% of Choosers will classify themselves as patient and happy, 25.4% as impatient and happy, 17.2% as patient and unhappy, and 25.6% as impatient and unhappy. We elicited these beliefs either

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
VARIABLES	Min. € Chooser takes late			Belief € Chooser takes late			Min. € Chooser takes late	Surrogate choice
Chooser								
Patient, happy	5.092***	4.932***	3.791***	12.08***	10.35***	8.847***		
	(0.308)	(1.094)	(0.996)	(0.178)	(0.844)	(1.112)		
Impatient, happy	3.383***	3.210***	2.069**	8.482***	6.796***	5.864***		
	(0.251)	(1.080)	(0.934)	(0.209)	(0.847)	(1.076)		
Patient, unhappy	3.719***	3.478***	2.368**	11.05***	9.296***	7.959***		
	(0.253)	(1.056)	(0.960)	(0.182)	(0.843)	(1.083)		
Impatient, unhappy	4.657***	4.413***	2.966***	9.049***	7.321***	6.275***		
	(0.293)	(1.044)	(0.960)	(0.200)	(0.840)	(1.071)		
Patience %-ile								
× 1		1.483**			3.454***		1.678**	4.427***
		(0.670)			(0.373)		(0.649)	(0.542)
× *Chooser*								
Patient, happy			1.367			3.952***		
			(0.934)			(0.498)		
Impatient, happy			1.368*			2.990***		
			(0.809)			(0.637)		
Patient, unhappy			1.316*			3.680***		
			(0.787)			(0.524)		
Impatient, unhappy			1.885**			3.186***		
			(0.955)			(0.585)		
× *Choice Dist. Info. tr.*							-1.177	-0.709
							(1.260)	(1.040)
Choice Dist. Info. tr.							0.815	0.541
							(0.851)	(0.857)
Session and order FE		Yes	Yes		Yes	Yes	Yes	Yes
p-values								
Patient = impat.								
if happy	0.000	0.000	1.000	0.000	0.000	0.240		
if unhappy	0.000	0.000	0.560	0.000	0.000	0.460		
Diff-in-diff	0.000	0.000	0.670	0.000	0.000	0.610		
Joint test	0.000	0.000	0.920	0.000	0.000	0.660		
Observations	1,212	1,164	1,164	1,212	1,164	1,164	1,548	1,446
Number of subjects	303	291	291	303	291	291	387	387

Table 7: Effect of information provision on projective paternalism. We exclude subjects who responded non-montonically to any of the multiple-decision lists in the online component for columns 2, 3, 5, and 6. *p*-values labeled *Joint test* correspond to the hypothesis that all of the coefficients associated with each Chooser statement are equal to each other; those in columns 1, 2, 4, and 5 refer to the level effects of Chooser statements, while those in columns 3 and 6 refer to the interaction effects between patience percentile and Chooser statements. All regressions include session and order fixed effects. Columns 7 and 8 include subjects in the Choice Distribution Information treatment and opportunity sets with front-end delay. They also include opportunity set and front-end delay fixed effects. Numbers in parentheses indicate standard errors of the estimates, clustered by subject. $^{*}p < 0.1,^{**}p < 0.05,^{***}p < 0.01$.



(A) (B)

Figure 8: Projective paternalism and the provision of Chooser information. Panel A: Mandates by Chooser statement, averaged across Choice Architects. Whiskers indicate 95% confidence intervals with standard errors clustered by subject. Panel B: Projective paternalism by Chooser statement. Figures exclude subjects classified as libertarian.

To document these results formally, we regress mandates on indicators for each of the four Chooser statements along with various combinations of additional explanatory variables discussed below, clustering standard errors on the subject level. We only employ data from the Chooser Information condition. In contrast to Figure 7, we also include subjects classified as libertarians.

Table 7 displays the results. Column 1 shows that the level effects associated with the various Chooser statements differ significantly from each other both economically and statistically. Column 2 adds the Choice Architect's patience percentile as a right-hand-side variable; its coefficient is positive and highly statistically significant.[55] Thus, ideals-projective paternalism prevails despite the provision of Chooser-specific information. As a further check on the robustness of ideals-projective paternalism, we estimate a specification that allows the coefficient of Choice Architects' patience percentile to vary freely across the four Chooser statements (see column 3). The estimated coefficients are all similar, and we cannot reject the hypothesis that they are identical, either in pairwise comparisons or in a joint test ($p > 0.5$ for each test).

Next, we explore the mechanisms by which the information we provide affects the Choice Architects interventions. Column 4 reports a regression of Choice Architects' beliefs about the mean delayed payment selected by unrestricted Choosers on indicators for the four Chooser statements. The large and statistically significant differences between the four coefficients establish that the information changes Choice Architects' beliefs about the Choosers' inclinations. However, it is also apparent from the same

directly before or directly after eliciting beliefs about the distribution of choices by ten unrestricted subjects (randomized on the individual level).

[55] The magnitude of the estimated coefficient is comparable to its counterpart in Table 4. However, this comparison is imperfect because the two estimates are based on different option menus.

regression that the information does not affect Choice Architects interventions *only* by changing beliefs about Choosers' selections. In particular, we know that, as a general matter, an increase in the mean of the Choice Architect beliefs about the Chooser's unrestricted selection increases the chosen mandate (see Section 5.1). And yet, 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 1 and 4). The difference in differences is highly statistically significant.[56]

The next two columns show, however, that despite its effects on beliefs and mandates, Chooser information does not alleviate the false consensus effect—neither when we include the Choice Architects' patience percentile as a single regressor (column 5), nor when we allow for different effects across the four Chooser statements (column 6). This finding is consistent with previous research showing that the false consensus effect is surprisingly robust with respect to information provision (Engelmann and Strobel, 2012).[57]

In addition to the Chooser-specific information analyzed above, 100 additional Choice Architects participated in the *Choice Distribution Information* condition. The experiment for these Choice Architects proceeded in the same fashion as for all other Choice Architects, with the exception that, in each round, these Choice Architects could click a button to view the distribution of previous selections Choosers made when all options were available.[58] Remarkably, 39.7% of non-libertarian subjects in this treatment never viewed any distributions of Choosers' selections, even though they could do so costlessly, at the click of a button.[59]

The regression reported in column 7 allows us to ask whether the option to view information on the distribution of unrestricted Choosers' selections meaningfully affect ideals-projective paternalism. Because of the small sample size and the limited number of Choice Architects who viewed the information, the confidence interval for the key coefficient is large, encompassing both the absence of projective paternalism and the possibility that it is unchanged. However, we continue to find a strong relation between surrogate choices and Choice Architects' patience percentiles, even in the Choice Distribution Information treatment ($p < 0.01$, column 8). Accordingly, there is evidence of projection even when Choice Architects have access to full information about the distribution of unrestricted Choosers' selections.

Taken as a whole, our results show that ideals-projective paternalism is robust with respect to information provision.

[56] $p < 0.001$ in a two-equation system OLS-regression with bootstrapped standard errors clustered on the subject level (1,000 samples).

[57] The finding is also consistent with the possibility that the information we provide to subjects is imperfect.

[58] Upon clicking the button they observed, for each of the three options in the menu, a line of text of the following form: *Out of 100 previous Choosers, N choose X today and Y in six months.* We calibrated the numbers N based on online pilot studies. Each Choice Architect in each session session faced a 25% chance of being assigned to the Choice Distribution Information condition.

[59] Surprisingly, this fraction is similar for libertarians and non-libertarians (37.5% for libertarians). Libertarians may choose to view the information out of curiosity. Alternatively, some of the Choice Architects we classify as libertarian might have been willing to intervene had the distribution of past choices been different.

5.3 Front-end delay

Next, we examine the effect of adding front-end delay (one incremental week) to all payments. As we explained in Section 2.1, under the view that the desire for immediate gratification is a bias (see, e.g., Frederick, Loewenstein and o'Donoghue, 2002), the introduction of front-end delay should eliminate or at least diminish the incentive to intervene. Mistakes-projective paternalism reinforces this tendency, while ideals-projective paternalism creates a countervailing tendency.

	(1)	(2)	(3)	(4)	(5)	(6)
VARIABLES	€ Choice Architect takes late	Mean belief	Min. € Chooser takes late	Surrogate Choice	Welfare belief non-inc.	Welfare belief -CV
Method	OLS	OLS	OLS	OLS	Ordered probit	Interval regression
Front-end delay	0.747***	0.350***	0.462*	0.064	0.149**	0.046**
	(0.195)	(0.117)	(0.265)	(0.202)	(0.069)	(0.020)
Cut 1					-0.917***	
					(0.210)	
Cut 2					0.922***	
					(0.209)	
Mean of dep. var.	13.26***	11.22***	4.712***	12.88***	-	0.086***
	(0.196)	(0.134)	(0.220)	(0.184)	-	(0.014)
Observations	1,212	1,212	1,212	1,136	1,212	1,093
Number of subjects	303	303	303	303	303	290

Table 8: Paternalism with front-end delay. All regressions include session, order, and option menu fixed effects. Numbers in parentheses indicate standard errors of the estimates, clustered by subject. The number of observations regarding surrogate choice is smaller because some of these choices were not recorded in the first two sessions. $^{*}p < 0.1$,$^{**}p < 0.05$,$^{***}p < 0.01$.

Choice Architects choose for themselves among options with front-end delay in the penultimate part of Stage 2 of the laboratory sessions (see Table 2). We present them with all of the menus they encountered earlier in the experiment when constructing opportunity sets. One of these choices determines the Choice Architects' payment with 25% probability.[60]

[60]Potentially, mechanisms such as anchoring or a demand for consistency may create an artificial relation between the Choice Architect's own choices during the laboratory component and the opportunity sets she has previously constructed for the Choosers. We test for such mechanisms as follows. First, we rank subjects according to the mean amount of money they choose to receive with a half-year delay in Stage 2 of the laboratory experiment for the three-option menus used in the Main condition (excluding the round with front-end delay). Second, we define a variable d_i as the difference between the rank we just defined and the rank of a Choice Architect's patience elicited in the online component. Third, we define m_i as the percentile rank of the mean mandate a Choice Architect imposes on Choosers in the Main condition (excluding the round with front-end delay). We then examine the relation between d_i and m_i. We continue to maintain the assumption that the confounding mechanisms do not apply to the relation between mandates and choices in the online component (see Section 3.1). If the confounding mechanisms are quantitatively important, then we should observe a positive relation between m_i and d_i: if the selected mandates influence subsequent choices that Choice Architects make for themselves through anchoring or a demand for consistency, then those who impose higher mandates should exhibit larger discrepancies between their patience ranks in the online component and Stage 2 of the laboratory component. If the confounding mechanisms play

Column 1 of Table 8 presents a regression of the delayed payment for the Choice Architect's selected option on an indicator for front-end delay. We use all choices in the Main condition, include session, order, and option menu fixed effects, and cluster standard errors by subject. The estimates show that the addition of front-end delay yields an increase in the selected delayed payment that is both statistically and economically significant (€0.75). Choice Architects, moreover, predict that unrestricted Choosers will exhibit this pattern, although to a smaller degree (€0.35, as shown in column 2).[61]

Turning to Choice Architects' interventions, our results contradict conventional behavioral welfarism and mistakes-projective paternalism, but are consistent with ideals-projective paternalism. The introduction of front-end delay leads Choice Architects to impose a mandate that is €0.46 higher ($p < 0.1$, column 3). However, front-end delay does not, on average, alter surrogate choices (column 4). Accordingly, it strengthens Choice Architects' inclinations to intervene conditional on a particular judgment of what is good for the Chooser, but leaves those judgments unchanged. We caution, however, that the estimated impact on surrogate choices may be attenuated due to ceiling effects.[62]

Interventions remain paternalistic in the presence of front-end delay. Indeed, Choice Architects believe that the restrictions they impose benefit the Chooser to a greater extent with front-end delay than without it, according to both non-incentivized statements (column 5) and elicited compensating variation (column 6).[63]

6 Real-world policies

We conclude this investigation by showing that projective paternalism extends to judgments about real-world paternalistic policies. Additionally, we demonstrate that subjects' paternalistic judgments across the experimental and policy domains are related.

To this end, in Stage 3 of the experiment, we ask subjects to rate four policy proposals concerning 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 on tax policies for Switzerland, so that personal interests would not influence their answers, at least in principle. We ask subjects to assume that the tax policies would be budget-neutral, so responses do not reflect general attitudes about the size of government. For each

no role, then we should observe no relation between d_i and m_i. Formally, when we regress d_i on m_i, we obtain a coefficient estimate of 0.03 (s.e. 0.05). This relation is far from statistically significant ($p = 0.56$).

[61] Relatedly, the subjects in Fedyk (2017) predict that others' decisions will be more patient when made in advance than when the earliest consequence of the decision arises immediately.

[62] A possible explanation for the observed pattern is that, with immediate rewards, the Chooser has private information about his own momentary preference shocks. In contrast, with front-end delay, this informational asymmetry may be smaller, providing a justification for more restrictive interventions.

[63] Comparing Choice Architects' own choices in Stage 2 to the choice sets they construct for Choosers provides additional evidence against mistakes-projective paternalism. According to mistakes-projective paternalism, we should observe that Choice Architects who choose an impatient option for themselves will tend to remove that option for the Chooser. Yet, conditional on choosing the least patient or middle option for herself, the chance that a Choice Architect removes this option for the Chooser never exceeds 12.5% for any opportunity set (see Appendix A5).

policy, we elicit the extent to which the subject supports or opposes its implementation.[64] We also elicit beliefs about the impact of each policy on the welfare of the average citizen.[65] After subjects provide these judgments and answer additional non-incentivized questions,[66] they provide information about themselves that relates to the impacted activities. Specifically, we elicit subjects' body mass index,[67] their average alcohol consumption, their frequency of binge drinking (defined as the consumption of four or more units[68] of alcohol for men, or five or more units for women, within a two-hour period), 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 Table 2 for a schematic overview, and Appendix D.1 for comprehensive detail).[69]

Projective paternalism with real-world paternalistic policies How do respondents' own characteristics relate to their policy judgments? Focusing on the example of alcohol taxes, mistakes-projective paternalism entails a positive relation between alcohol consumption and support for alcohol taxation, which could arise if heavier drinkers appreciate aids to limit their drinking, and believe that others would also benefit from those aids. Ideals-projective paternalism predicts the opposite relationship. If people project their own preferred level of alcohol consumption on others, lighter drinkers, for instance, will tend to believe that heavier drinkers would be better off with lower alcohol consumption. Therefore, we should expect a negative relationship between own alcohol consumption and support for alcohol taxes.

To test these hypotheses, we estimate ordered probit models relating subjects' support for a policy to the subset of personal characteristics pertaining to that policy. We include all subjects from all treatments and control for session fixed effects. Table 9 displays the results. Column 1 shows that our subjects are more likely to express support for alcohol taxes the less they themselves drink on average, and the less often they binge drink, exactly as ideals-projective paternalism predicts. Because we elicit German subjects' opinions concerning Swiss policy, the relation is probably not attributable to differences in subjects' willingness to bear taxes themselves. Column 2 shows that these results are robust with respect to the inclusion of controls for gender, age, monthly expenses, and political orientation. Although the coefficient of weekly consumption is no longer statistically significant at the 5% level, its magnitude and

[64] 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.

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

[66] See Appendix D.4.

[67] 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 as BMI $= \frac{w}{(h/100)^2}$.

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

[69] With respect to all of these measures, misreporting 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. Parallel statements hold regarding self-reported alcohol consumption (Sobell and Sobell, 1995) and self-reported cigarette smoking (West et al., 2007).

Stated support for policy	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Increase alcohol tax		Increase tobacco tax		Introduce sugary drinks tax		Tighten restrictions on short-term lending	
Alcohol consumption								
Alcohol units / week	-0.055**	-0.052*						
	(0.027)	(0.027)						
log(days binge drinking / year)	-0.123***	-0.117***						
	(0.044)	(0.045)						
Tobacco consumption								
Smoker yes / no			-0.952***	-0.981***				
			(0.211)	(0.221)				
Cigarettes / day			-0.017	-0.018				
			(0.029)	(0.028)				
Body Mass Index					-0.049**	-0.047**		
					(0.021)	(0.021)		
Debt								
Credit card debt (in €1,000)							-0.614***	-0.717***
							(0.167)	(0.177)
Other short-term debt yes / no							0.088	0.105
							(0.384)	(0.392)
Male		-0.165		-0.268**		-0.038		0.154
		(0.120)		(0.118)		(0.117)		(0.121)
Age		0.009		-0.022**		-0.002		0.018
		(0.012)		(0.010)		(0.012)		(0.016)
Monthly expenses (in €1,000)		0.172		0.248*		0.081		0.068
		(0.166)		(0.147)		(0.164)		(0.168)
Political orientation		-0.036		0.058		0.077		0.021
		(0.054)		(0.061)		(0.058)		(0.060)
Cut 1	-1.774***	-1.501***	-2.088***	-2.560***	-2.533***	-2.424***	-1.862***	-1.292**
	(0.215)	(0.368)	(0.225)	(0.351)	(0.540)	(0.587)	(0.463)	(0.608)
Cut 2	-0.767***	-0.483	-1.402***	-1.858***	-1.727***	-1.613***	-0.844*	-0.273
	(0.201)	(0.360)	(0.200)	(0.337)	(0.526)	(0.574)	(0.459)	(0.610)
Cut 3	0.205	0.492	-0.429**	-0.865***	-0.775	-0.658	0.104	0.680
	(0.199)	(0.360)	(0.199)	(0.329)	(0.521)	(0.568)	(0.460)	(0.610)
Cut 4							1.048**	1.635***
							(0.466)	(0.613)
Observations	403	403	403	403	398	398	351	351

Table 9: Support of real-world paternalistic policies and respondent characteristics. Each column shows a separate ordered probit regression. For population means, support is measured as *should definitely not do, should probably not do, should probably do, should definitely do*. Sample size in some columns falls short of 403 because some subjects refused to disclose personal characteristics. Binge drinking is defined as the consumption of at least 4 (females) or 5 (males) units of alcohol within a period of two hours (National Institutes on Alcohol Abuse and Alcoholism, 2018). Subjects choose from intervals; we use the midpoint of each interval for analysis. 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). All regressions include session fixed effects. Numbers in parentheses indicate standard errors of the estimates, clustered by subject. $^{*}p < 0.1$,$^{**}p < 0.05$,$^{***}p < 0.01$.

standard error both change only slightly.[70] Columns 3 and 4 show that German subjects are substantially and significantly more likely to express support for tobacco taxes in Switzerland if they do not smoke themselves. In columns 5 and 6, we focus on the relationship between support for taxes on sugary drinks and body mass index. Because body mass index depends on many factors other than sugary drink consumption, these two measures are perhaps more distantly related than the corresponding variables used for the alcohol and cigarette tax policies. Nonetheless, we find that people with lower body mass indexes express stronger support for sugary drinks taxes, again consistent with ideals-projective paternalism. Similarly, subjects are less likely to support restricting the market for short-term, high-interest lending when they have larger credit card balances (columns 7 and 8).[71]

Expressions of support for these policies are consistent with paternalistic motivations. For each of the three tax policies, an increase in the support for that policy is associated with an economically and statistically significant increase in the magnitude of the perceived social benefits (see Appendix B.7 for details). For the case of limits on short-term, high-interest lending, a similar relation obtains except at the highest two levels of support.[72] Overall, the evidence on judgments about real-world paternalistic policies is consistent with ideals-projective paternalism, and inconsistent with both conventional behavioral welfarism and mistakes-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. This would be the case, for instance, if heavier individuals are more likely to believe that losing weight is difficult, and hence think that the demand elasticity for sugary drinks is lower. Appendix C reports a vignette experiment with US subjects in which we control for variations in beliefs about efficacy. Whenever the pattern of support for paternalistic policies points to ideals-projection, controlling for beliefs about efficacy leaves that pattern qualitatively unchanged.[73]

Experimental interventions and judgments of real-world policies Subjects' judgments of real-world paternalistic policy proposals are not only consistent with ideals-projective paternalism, but also relate meaningfully to the decisions they make as Choice Architects in our experiment. To demonstrate this point, we regress mandates in the Main condition on the first principal component of the four policy judgments, controlling for session, order, and option menu fixed effects; see Column 1 of Table 10. We find a positive relationship that is statistically significant at the 5% level. Column 2 adds

[70] Moreover, the coefficient on weekly consumption becomes statistically significantly (and remains negative) when weekly consumption is the only alcohol-related predictor in the regression, regardless of whether we control for demographic characteristics. In describing the alcohol tax proposal, the survey explicitly referenced binge drinking.

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

[72] We cannot exclude the possibility, however, that non-paternalistic considerations such as the prevention of externalities also enter subjects' assessment of the policies. Choice Architects' behavior in stages 1 and 2 of the laboratory part is not subject to such confounding mechanisms.

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

controls for the same demographic characteristics of Choice Architects used in Table 9, as well as for each of each of the behaviors that relate to the four real-world policy proposals. We also control for educational background, which may be related both to support for real world paternalistic policies (for instance, through better health knowledge) and to mandates in the experimental treatment (through the relationship between cognitive ability and patience, as documented in Dohmen, Falk, Huffman and Sunde, 2010). We separately include high school GPA, as well as final grades in mathematics and German literature. These additions strengthen the estimated relationship between policy judgments and experimental behavior. As a placebo test, columns 3 and 4 perform parallel regressions for the Induced Chooser Preferences condition, using the mandated number of silver tokens, which serve as the currency to which the induced discount factor applies. There is no evidence of a relationship between policy judgments and interventions in that condition.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Min € Chooser must take late		Min silver tokens Chooser must take		Belief that smaller opportunity set is better		−Compensating Variation	
Conditions	Main		Ind. Chooser Pref.		Exogenous Restriction		Exogenous Restriction	
Method	OLS		OLS		Ordered probit		Interval regression	
Policy support (1st princ. comp.)	0.347**	0.477***	-0.128	-0.052				
	(0.135)	(0.153)	(0.088)	(0.095)				
Welfare judgment about policies (1st princ. comp.)					0.123***	0.181***	0.021*	0.036***
					(0.044)	(0.051)	(0.012)	(0.013)
Cut 1					0.634*	0.560		
					(0.334)	(0.906)		
Cut 2					0.881***	0.821		
					(0.334)	(0.907)		
Controls	No	Yes	No	Yes	No	Yes	No	Yes
Observations	1,612	1,364	806	682	806	682	638	551
Number of subjects	403	341	403	341	403	341	343	293

Table 10: Experimental decisions and support for real-world paternalistic policies. Controls consist of all the predictive variables in Table 9, as well as high school GPA and grades on final examinations in high school mathematics and German literature. All regressions include session, order, and option menu fixed effects. Columns 5 - 8 control for whether the Choice Architects' welfare beliefs pertain to removing the least patient or the two least patient options. Numbers in parentheses indicate standard errors of the estimates, clustered by subject. $^{*}p < 0.1,^{**}p < 0.05,^{***}p < 0.01$.

We also find a positive relationship between Choice Architects' beliefs about the welfare effects of laboratory interventions and of real-world paternalistic policies. We regress Choice Architects' beliefs about the welfare effects of exogenously removing impatient options on the first principal component of Choice Architects' judgments regarding the welfare effects of the four real-world policies. Column

5 uses non-incentivized beliefs as the dependent variable. It shows that there is a significant positive relationship, which remains unchanged when we control for the same collection of variables as in Table 9 (column 6). Similar results obtain when we use beliefs regarding compensating variations as the dependent variable (columns 7 and 8). Overall, both interventions and welfare judgments pertaining to the timing of payments are strongly associated with subjects' assessments of real-world paternalistic policies.

7 Conclusion

This paper examines when, why, and how people intervene in others' choices. In a setting involving intertemporal tradeoffs, we find that Choice Architects frequently remove options that are attractive to impatient decision makers. Choice Architects believe their interventions benefit the Chooser, and are thus acting paternalistically. How do Choice Architects judge what is good for others? This is a difficult task because, 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. Ideals-projective paternalism emerges from our empirical analysis as the key organizing principle. An ideals-projective paternalist acts *as if* she believes other 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), and that it is robust with respect to the provision of information about the Choosers. It is also consistent with the otherwise surprising finding that the introduction of front-end delay does not reduce, and may increase, the inclination to intervene, and it extends to subjects' assessment of real-world paternalistic policies. Ideals-projective paternalism contrasts with mistakes-projective paternalism, according to which subjects view themselves as prone to particular types of decision errors, make the assumption that others share that susceptibility, and intervene to remedy it. It also contrasts with conventional behavioral welfarism, which holds that subjects either understand the true distribution of behavioral proclivities or form a belief about it based on information concerning the Choosers. Both of these alternative hypotheses are more closely aligned with widespread perspectives in behavioral welfare economics, but neither finds support in our data.

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.[74] Existing evidence does suggest, however, that people have a positive willingness to pay for autonomy (Fehr, Herz and Wilken-

[74] We have frequently encountered both of these viewpoints among audiences of economists.

ing, 2013; Bartling, Fehr and Herz, 2014; Owens, Grossman and Fackler, 2014; Lübbecke and Schnedler, 2018).

There are many questions we hope future research will clarify. For example, do Choice Architects intervene because they take issue with the Chooser's objectives or because they question the Chooser's competence to achieve his own objectives? Do Libertarian subjects abstain from intervening due to matters of principle, or simply because they do not feel confident in their assessment of what is good for others? We also hope to extend the empirical study of paternalistic decision making to subject pools consisting of "professional paternalists" such as medical doctors and policy makers. In contexts where objective benchmarks are available, existing evidence suggests that nearly everyone exhibits behavioral biases (Stango and Zinman, 2019), including elected politicians (Sheffer, Loewen, Soroka, Walgrave and Sheafer, 2018).

References

Altmann, Steffen, Armin Falk, and Andreas Grunewald, "Incentives and information as driving forces of default effects," *unpublished*, 2017.

Amador, Manuel, Ivan Werning, and George Marios Angeletos, "Commitment vs. Flexibility," *Econometrica*, 2006, *74* (2), 365–396.

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

_ **and Axel Ockenfels**, "The Ethics of Incentivizing the Uninformed. A Vignette Study," *The American Economic Review Papers and Proceedings*, 2017, *107* (5), 91–95.

_ **, Muriel Niederle, and Alvin E. Roth**, "More Money, More Problems? Can High Pay be Coercive and Repugnant?," *American Economic Review, Papers & Proceedings*, 2015, *105* (5), 357–60.

Andreoni, James, "An experimental test of the public-goods crowding-out hypothesis," *The American Economic Review*, 1993, pp. 1317–1327.

_ **, Christina Gravert, Michael A Kuhn, Silvia Saccardo, and Yang Yang**, "Arbitrage Or Narrow Bracketing? On Using Money to Measure Intertemporal Preferences," *NBER working paper*, 2018.

Bartling, Björn, Ernst Fehr, and Holger Herz, "The intrinsic value of decision rights," *Econometrica*, 2014, *82* (6), 2005–2039.

Basu, Kaushik, "The economics and law of sexual harassment in the workplace," *The Journal of Economic Perspectives*, 2003, *17* (3), 141–157.

_ , "Coercion, Contract and the Limits of the Market," *Social Choice and Welfare*, 2007, *29* (4), 559–579.

Benartzi, Shlomo, John Beshears, Katherine L Milkman, Cass R Sunstein, Richard H Thaler, Maya Shankar, Will Tucker-Ray, William J Congdon, and Steven Galing, "Should governments invest more in nudging?," *Psychological Science*, 2017, *28* (8), 1041–1055.

Bennyhoff, Donald G and Francis M Kinniry Jr, "Advisor's Alpha," *Vanguard research paper (December). https://advisors. vanguard. com/iwe/pdf/ICRAA. pdf*, 2011.

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

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

Bernheim, B Douglas and Dmitry Taubinsky, "Behavioral public economics," *NBER working paper*, 2018.

Besley, Timothy, "A simple model for merit good arguments," *Journal of Public Economics*, 1988, *35* (3), 371–383.

Bisin, Alberto, Alessandro Lizzeri, and Leeat Yariv, "Government policy with time inconsistent voters," *American Economic Review*, 2015, *105* (6), 1711–37.

Boven, Leaf Van, David Dunning, and George Loewenstein, "Egocentric empathy gaps between owners and buyers: misperceptions of the endowment effect.," *Journal of Personality and Social Psychology*, 2000, *79* (1), 66.

Carlin, Bruce Ian, Simon Gervais, and Gustavo Manso, "Libertarian paternalism, information production, and financial decision making," *The Review of Financial Studies*, 2013, *26* (9), 2204–2228.

Clemens, Michael, "Testing for Repugnance in Economic Transactions: Evidence from Guest Work in the Gulf," *unpublished*, 2017.

Cohen, Jonathan D, Keith Marzilli Ericson, David Laibson, and John Myles White, "Measuring time preferences," *NBER working paper*, 2016.

Cooper, David J and John H Kagel, "Other-regarding preferences," *The handbook of experimental economics*, 2016, *2*, 217.

Currie, Janet and Firouz Gahvari, "Transfers in cash and in-kind: Theory meets the data," *Journal of Economic Literature*, 2008, *46* (2), 333–83.

Delavande, Adeline, Xavier Giné, and David McKenzie, "Measuring subjective expectations in developing countries: A critical review and new evidence," *Journal of Development Economics*, 2011, *94* (2), 151–163.

Dohmen, Thomas, Armin Falk, David Huffman, and Uwe Sunde, "Are risk aversion and impatience related to cognitive ability?," *The American Economic Review*, 2010, *100* (3), 1238–1260.

Dworkin, Gerald, *Paternalism*, Wadsworth Publishing Company, 1971.

_ , "Paternalism," *the Monist*, 1972, pp. 64–84.

Elias, Julio J, Nicola Lacetera, and Mario Macis, "Markets and Morals: An Experimental Survey Study," *PloS one*, 2015, *10* (6).

Elias, Julio, Nicola Lacetera, and Mario Macis, "Sacred Values? The Effect of Information on Attitudes Toward Payments for Human Organs," *American Economic Review, Papers & Proceedings*, 2015, *105* (5), 361–365.

_ , _ , and _ , "Efficiency-morality trade-offs in repugnant transactions: A choice experiment," *unpublished*, 2016.

Engelmann, Dirk and Martin Strobel, "Deconstruction and reconstruction of an anomaly," *Games and Economic Behavior*, 2012, *76* (2), 678–689.

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

Falk, Armin, Anke Becker, Thomas Dohmen, David Huffman, and Uwe Sunde, "The preference survey module: A validated instrument for measuring risk, time, and social preferences," 2016.

Faravelli, Marco, Priscilla Man, and Randall Walsh, "Mandate and paternalism: a theory of large elections," *Games and Economic Behavior*, 2015, *93*, 1–23.

Fedyk, Anastassia, "Asymmetric Naiveté: Beliefs about Self-Control," *unpublished*, 2017.

Fehr, Ernst and Urs Fischbacher, "Why social preferences matter–the impact of non-selfish motives on competition, cooperation and incentives," *The Economic Journal*, 2002, *112* (478), C1–C33.

_ , Holger Herz, and Tom Wilkening, "The lure of authority: Motivation and incentive effects of power," *The American Economic Review*, 2013, *103* (4), 1325–1359.

Foerster, Stephen, Juhani T Linnainmaa, Brian T Melzer, and Alessandro Previtero, "Retail financial advice: does one size fit all?," *The Journal of Finance*, 2017, *72* (4), 1441–1482.

Frederick, Shane, George Loewenstein, and Ted o'Donoghue, "Time discounting and time preference: A critical review," *Journal of Economic Literature*, 2002, *40* (2), 351–401.

Gangadharan, Lata, Philip J Grossman, and Kristy Jones, "Donor types and paternalism," *unpublished*, 2015.

Gigerenzer, Gerd, "Moral intuition = fast and frugal heuristics?," in "The cognitive science of morality: intuition and diversity," MIT Press, 2008, pp. 1–26.

Gillen, Ben, Erik Snowberg, and Leeat Yariv, "Experimenting with measurement error: techniques with applications to the Caltech cohort study," *Journal of Political Economy*, forthcoming.

Glaeser, Edward L, "Paternalism and psychology," *NBER working paper 11789*, 2005.

Gorber, S. Connor, M. Tremblay, D. Moher, and B. Gorber, "A comparison of direct vs. self-report measures for assessing height, weight and body mass index: a systematic review," *Obesity Reviews*, 2007, *8* (4), 307–326.

Gul, Faruk and Wolfgang Pesendorfer, "Temptation and self-control," *Econometrica*, 2001, pp. 1403–1435.

Holt, Charles A. and Susan K. Laury, "Risk Aversion and Incentive Effects," *American Economic Review*, 2002, *92(5)*, 1644 – 1655.

Ifcher, John and Homa Zarghamee, "Behavioral Economic Phenomena in Decision-Making for Others," 2018.

Jacobsson, Fredric, Magnus Johannesson, and Lars Borgquist, "Is altruism paternalistic?," *The Economic Journal*, 2007, *117* (520), 761–781.

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

Kataria, Mitesh, M Vittoria Levati, and Matthias Uhl, "Paternalism with hindsight: do protégés react consequentialistically to paternalism?," *Social Choice and Welfare*, 2014, *43* (3), 731–746.

Krawczyk, Michal and Lukasz P Wozny, "An experiment on temptation and attitude towards paternalism," *unpublished*, 2017.

Laibson, David, "Private Paternalism, the Commitment Puzzle, and Model-Free Equilibrium," *American Economic Review, Papers & Proceedings*, 2018, *108*, 1–21.

Leider, Stephen and Alvin E Roth, "Kidneys for sale: Who disapproves, and why?," *American Journal of Transplantation*, 2010, *10* (5), 1221–1227.

Linnainmaa, Juhani T, Brian Melzer, and Alessandro Previtero, "The misguided beliefs of financial advisors," *Journal of Finance*, forthcoming.

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

Loewenstein, George and Emily C Haisley, *The economist as therapist: Methodological ramifications of 'light' paternalism*, Vol. 1 of *Handbook of Economic Methodologies*, Oxford University Press, 2007.

Lübbecke, Silvia and Wendelin Schnedler, "Don't patronize me! An Experiment on Rejecting Paternalistic Help," *unpublished*, 2018.

Lusk, Jayson L, Stephan Marette, and F Bailey Norwood, "The paternalist meets his match," *Applied Economic Perspectives and Policy*, 2013, *36* (1), 61–108.

Mill, John Stuart, *On liberty*, Longmans, Green, Reader, and Dyer, 1869.

Moffitt, Robert, "Welfare work requirements with paternalistic government preferences," *The Economic Journal*, 2006, *116* (515), F441–F458.

Moffitt, Robert A, "The negative income tax and the evolution of US welfare policy," *Journal of Economic Perspectives*, 2003, *17* (3), 119–140.

Mulligan, Casey B and Tomas J Philipson, "Merit motives and government intervention: Public finance in reverse," *National Bureau of Economic Research Working Paper*, 2000, *7698*.

National Institutes on Alcohol Abuse and Alcoholism, "Recommended Alcohol Questions," https://www.niaaa.nih.gov/research/guidelines-and-resources/recommended-alcohol-questions July 2018.

Nawaz, Haq, Wendy Chan, Mustapha Abdulrahman, David Larson, and David L Katz, "Self-reported weight and height: implications for obesity research," *American Journal of Preventive Medicine*, 2001, *20* (4), 294–298.

O'Donoghue, Ted and Matthew Rabin, "Doing it now or later," *American Economic Review*, 1999, *89* (1), 103–124.

Owens, David, Zachary Grossman, and Ryan Fackler, "The control premium: A preference for payoff autonomy," *American Economic Journal: Microeconomics*, 2014, *6* (4), 138–61.

Pikulina, Elena and Chloe Tergiman, "Preferences for Power," *unpublished*, 2018.

Ross, Lee, David Greene, and Pamela House, "The "false consensus effect": An egocentric bias in social perception and attribution processes," *Journal of Experimental Social Psychology*, 1977, *13* (3), 279–301.

Roth, Alvin E, "Repugnance as a Constraint on Markets," *Journal of Economic Perspectives*, 2007, *21* (3), 37–58.

Saint-Paul, Gilles, "Policy consequences of limited cognitive ability," *Spanish Economic Review*, 2004, *6* (2), 97–105.

Schroeder, Juliana, Adam Waytz, and Nicholas Epley, "Endorsing Help For Others That You Oppose For Yourself: Mind Perception Alters the Perceived Effectiveness of Paternalism.," *Journal of Experimental Psychology: General*, 2017.

Sheffer, Lior, Peter John Loewen, Stuart Soroka, Stefaan Walgrave, and Tamir Sheafer, "Nonrepresentative representatives: an experimental study of the decision making of elected politicians," *American Political Science Review*, 2018, *112* (2), 302–321.

Sobell, Linda C and Mark B Sobell, "Alcohol consumption measures," *Assessing alcohol problems: A guide for clinicians and researchers*, 1995, *2*, 75–99.

Stango, Victor and Jonathan Zinman, "We are all behavioral, more or less: Measuring and using consumer-level behavioral sufficient statistics," *NBER working paper*, 2019, *25540*.

Sunstein, Cass R, "Moral heuristics," *Behavioral and Brain Sciences*, 2005, *28* (4), 531–541.

Thaler, Richard H and Cass R Sunstein, "Libertarian paternalism," *American Economic Review*, 2003, *93* (2), 175–179.

_ **and Hersh M Shefrin**, "An economic theory of self-control," *Journal of Political Economy*, 1981, *89* (2), 392–406.

Uhl, Matthias, "Do self-committers mind commitment by others? An experiment on weak paternalism," *Rationality, Markets and Morals*, 2011, *2*, 13–34.

Vesterlund, Lise, "Using experimental methods to understand why and how we give to charity," *Handbook of Experimental Economics*, 2016, *2*, 91–151.

West, Robert, Witold Zatonski, Krzysztof Przewozniak, and Martin J Jarvis, "Can we trust national smoking prevalence figures? Discrepancies between biochemically assessed and self-reported smoking rates in three countries," *Cancer Epidemiology and Prevention Biomarkers*, 2007, *16* (4), 820–822.

Zamir, Eyal, "The efficiency of paternalism," *Virginia Law Review*, 1998, pp. 229–286.