Paternalism and Rule-Making: Three Experimental Studies in Political Economy

I n a u g u r a l d i s s e r t a t i o n

zur

Erlangung des Doktorgrades

der

Wirtschafts- und Sozialwissenschaftlichen Fakultät

der

Universität zu Köln

Vorgelegt von

Max Rainer Pascal Großmann, Master of Science

aus Darmstadt

2025

Referent:
Prof. Dr. Axel Ockenfels

Korreferent:
Prof. Dr. Frederik Schwerter

Vorsitzender der Prüfungskommission:
Jun.-Prof. Dr. Arno Apffelstaedt

Datum der Disputation:
28. April 2025

Diese Inauguraldissertation wurde von der Wirtschafts- und Sozialwissenschaftlichen Fakultät der Universität zu Köln angenommen.

meinen Eltern

Contents

List of Figures

List of Tables

Introduction

Modern societies are characterized by an abundance of formal rules. Such rules are expressions of rule-makers' subjective preferences, knowledge, beliefs, and perceived constraints. Investigating how policymakers employ purposeful human design to order society is fundamental to understanding the exercise of governmental authority.

The work in this thesis uses experiments to study rule-making—an issue at the heart of political economy and non-market decision-making—while focusing on paternalistic rules. Paternalism in government policy has become widespread, with governments worldwide implementing sin taxes, banning organ markets, prostitution, and public smoking, conducting a "War on Drugs," and instituting mandatory Social Security. While many motives can underlie support for such policies, paternalism is a central determinant. Dworkin's early definition of paternalism defined it—"roughly"—as the "interference with a person's liberty of action justified by reasons referring exclusively to the welfare, good, happiness, needs, interests or values of the person being coerced" (Dworkin, 1972, p. 65). However, paternalism can be understood in nuanced ways (e.g., Dworkin, 2020), and we touch upon these important definitional issues in each Chapter.

Recent normative assessments have ranged from arguments in favor of increased paternalism (e.g., Conly, 2012) to arguments against (e.g., Snowdon, 2017). Nobel Prize-winning economist Richard Thaler and highly-cited legal scholar Cass Sunstein have argued in favor of "nudges," softly paternalistic interventions that allow opting out (Sunstein, 2014; Sunstein and Thaler, 2003; Thaler and Sunstein, 2008). The present work attempts to use normative statements only in the formulation of testable predictions. The focus lies on an empirical investigation of paternalistic rule-making: Which situational factors cause rule-setters to intervene in decision-makers' personal decisions? What

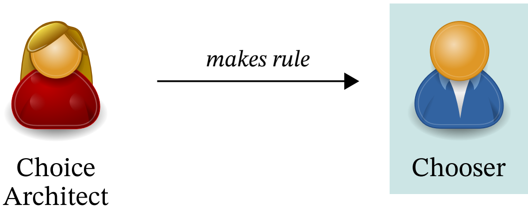


Figure 1: Experimental design used in this thesis

do these policymakers know? To what extent can personal preferences predict paternalistic action towards others? How do paternalistic policy tools interact?

These investigations are inherently linked to the economics of freedom. Freedom, or liberty, is essential for human flourishing (Hayek, 1944, 1960, 1988). But what is liberty? Russian-British philosopher Isaiah Berlin drew the following distinction: Negative liberty refers to the absence of restrictions, while positive liberty relates to the actual ability of decision-makers to act (Berlin, 1958; Carter, 2022). One could summarize these standpoints as follows: negative liberty is about what one *may* do, while positive liberty is about one *can* do. This thesis is exclusively concerned with research on grants of negative liberty: When are restrictions placed on the set of actions that decision-makers can choose from?

I use controlled experiments to *make* subjects into policymakers ("Choice Architects") who determine what decision-makers ("Choosers") may do. This basic study design (Figure 1) is used to investigate the behavioral foundations of paternalistic rule-making. The experimental method provides ample opportunity for this line of investigation. A. E. Roth (1986, p. 246) described three main purposes for experimentation in economics. First, such studies can be used to "speak to theorists." By providing a test bench where situational features or stimuli can be enabled or disabled, randomly allocated treatment conditions allow us to distinguish between theoretical models of human behavior. This application of experiments is of extraordinary significance to the work contained in this thesis. In Chapter 1, we demonstrate that many Choice Architects intervene in Choosers' decision to share personal data by setting minimum prices for that sharing of data. These minimum prices are highly correlated with but systematically smaller than Choice Architects' own Willingness-to-Accept (WTA). That Choosers are left some space to express their preferences conforms well with a theory of projective paternalism introduced by Ambuehl, Bernheim, and Ockenfels (2021). Experiments can also vary the Choice Architect's or the

Chooser's choice ecology and beliefs. Chapter 2 strongly resonates with this approach to experimental work. I use John Stuart Mill's *On Liberty* (1869) to inform a parsimonious microtheory of paternalistic intervention. This theory is subsequently empirically qualified using experiments. Such findings allow the creation of more predictive and insightful theories of human behavior (Kagel and A. E. Roth, 1995, p. 22).

Not all features of reality can be known ex ante. A. E. Roth (1986, p. 255) noted that researchers can use the laboratory to "search for facts" that are so far unrecognized in theory. If empirical regularities are observed through experimentation, those regularities can be used to develop a better understanding of reality. This purpose also enables researchers to incorporate psychological ideas into economic research. In the field of paternalism, this class of experiments is valuable as it can establish basic features of interventions. For example, Chapter 1 finds that beliefs about others' behavior are biased: Choice Architects tend to underestimate others' willingness to share data. Similarly, Chapter 3 investigates the mental model of Choice Architects, revealing that increased Chooser deliberation causes an optimistic shift in beliefs. In both of these cases, mental models are not modeled explicitly—rather, beliefs are analyzed using reduced-form methods. Findings such as these can be a stepping stone to a systematic understanding of the cognitive underpinnings of political economy.

Experiments can also be useful to "whisper in the ears of princes" (A. E. Roth, 1986, p. 261): to inform and motivate regulatory advice. On the one hand, experiments in rule-making can be used to inform policy debates and to shape the development of regulations that best reflect popular preferences and preserve freedom of choice: Chapter 2 highlights the tensions between the subjective preferences of Choice Architects and how they are distinctively combined with Chooser preferences if the latter are known. On the other hand, the empirical study of paternalism hints at opportunities for institutional reform. For example, Chapter 3 shows how waiting periods—a well-known softly paternalistic policy—are add-on restrictions, increasing the regulatory burden placed on Choosers. The "princes" of democratic polities—voters—will have to decide whether policymakers should face additional constraints in the continuously more sophisticated rules employed to order society.

In Chapter 1, we investigate experimentally how and to what extent individuals intervene in others' sharing of personal data. Much research has been done on a "privacy paradox" through which individuals state a high concern for privacy but do not act on it (e.g., Barnes, 2006). A newer literature has criticized these findings on methodological grounds (Solove, 2021). We investigate an "interpersonal" privacy paradox to ascertain whether people want more privacy for others than themselves. This research question relates decisively to the demand for privacy regulation. Moreover, this change in the object of analysis allows us to overcome the inherent conceptual difficulty in correlating stated and revealed privacy preferences. We conduct an experiment in which Choosers can set a WTA for the publication of data about themselves. The data that we collect are answers to personal questions; individually, they are harmless, but their inner product is revelatory. Choice Architects can impose minimum and maximum prices that may prevent publication even if a Chooser would have agreed to it. We collect beliefs about the relative frequency of WTA intervals.

We correlate the decision to intervene with Choice Architects' own behavior in an experiment similar to that of Choosers. We find that 30 % of Choice Architects set a minimum price, but only few set a maximum price. Despite us giving Choosers full control of when their data may be shared, many Choice Architects intervene. The WTA of Choice Architects who intervene is systematically higher than the WTA of non-interventionist Choice Architects. Moreover, when we correlate minimum price to interventionists' own WTAs, we find a strong positive relationship. Nonetheless, Choice Architects don't merely impose their own WTA on Choosers; they grant Choosers some liberty to express their own preferences. This relates to previous findings that Choice Architects act like "ideals-projective" paternalists (Ambuehl, Bernheim, and Ockenfels, 2021) and simultaneously demonstrates—in our experiment—the absence of an interpersonal privacy paradox. We find that beliefs over WTAs are systematically biased. The frequency of low WTAs is underestimated, while that of high WTAs is overestimated. We note that while WTAs correlate with the Concern for Internet Privacy instrument (Smith, Milberg, and Burke, 1996), two automatically and inconspicuously elicited measures of privacy-interested behavior (ad-blocker usage and time spent on our experiment's privacy policy) do not. This first study on what might be termed "privacy paternalism" sheds light on the political economy of privacy regulations. Choice Architects who demand more money for their data are more likely to impose a minimum price for others, but they do leave wiggle room for other preferences.

Paternalism has been justified on many grounds. One of the most important stated reasons for interfering with others' choices is asymmetric information.

Policymakers are often thought to possess more knowledge about a subject area. Even John Stuart Mill, whose *On Liberty* (1869) is often viewed as a statement against paternalism (Arneson, 1980), allowed exceptions to protect uninformed decision-makers. Chapter 2 uses these writings to inform a simple formal theory of liberty that conditions on Choosers' ability to make mistakes, for example if information is absent. This theory makes testable statements both about the extensive margin (fewer space for mistakes should lead to fewer interventions) and the intensive margin (if an intervention is to occur, it ought to reflect the Chooser's hypothetical full-information preference except if Choice Architects heavily weigh their own preference). In a first experiment, I induce asymmetric information between Chooser and Choice Architect. I vary the amount of ambiguity inherent in a binary lottery in a choice between a fixed amount of money and that lottery. Choice Architects, however, face no ambiguity. I find a strong effect of ambiguity on Choice Architects' intervention rates. A second experiment replicates these results. This second experiment also reveals that beliefs about Chooser behavior under risk have a small, but statistically significant association with the intensive margin, independent of the Choice Architects' own preference. Moreover, I let each Choice Architect decide for a Chooser whose hypothetical full-information preference is known to the Choice Architect. My results reveal that Choice Architects use information about the Chooser's type to determine the intensive margin. To some extent, then, Choice Architects "help" Choosers attain what they would have chosen had they known all the facts. However, Mill's postulate about the intensive margin is challenged empirically. Choice Architects may use their own preference to about the same degree as that of the Chooser in shaping interventions. Further analyses organize this result: both causally and by correlation, Choice Architects are simply more likely to impose the riskless option independently of their own preference. This is an important qualification to what Ambuehl, Bernheim, and Ockenfels (2021) termed "(ideals-)projective paternalism": paternalistic projection appears to be asymmetric. This suggests the existence of "cosmic ideals" or bliss points for intervention that are correlated with Choice Architects' own preference, but conceptually distinct from it. These cosmic ideals represent options that are simply more objectively "correct" and thus more easily imposed on others.

I also study Choice Architects' behavior regarding the provision of information. Mill taught that it is preferable to warn Choosers instead of intervening. I allow Choice Architects to resolve the ambiguity faced by Choosers by providing the latter with true information. Indeed, Choice Architects overwhelmingly provide information to Choosers. My study also reveals that Choice Architects generally do not strategically exploit the non-provision of information in order to license a subsequent intervention. Nonetheless, Choice Architects who do not provide information are much more likely to intervene. Moreover, as predicted by theory, if mistakes go into a Choice Architect's subjectively preferred direction, such a Choice Architect would be slightly less likely to provide information to the Chooser. In sum, there appears to exist a minority of Choice Architects that strategically exploits non-provision, but most Choice Architects do not fundamentally prefer to intervene if they can avoid it. These findings highlight the context-dependent and highly heterogeneous nature of paternalism when conditioned on knowledge—and how *policymakers'* knowledge matters, too.

Writers have drawn a variety of distinctions between various forms of paternalism. One major distinction is between "soft" and "hard" paternalism (e.g., Feinberg, 1989; Hanna, 2018; Kirchgässner, 2017; Schnellenbach, 2012). In the original formulation (Feinberg, 1989), soft paternalism refers to policies that attempt to prevent decisions made by those who are uninformed or impaired, but leave purposeful, deliberate choices intact. On the other hand, hard paternalism is the restriction of freedom of choice even if choices are entirely voluntary, informed and deliberate. If an experimental policymaker has the choice between *both* hard and soft tools, do they substitute one for the other? In general, does Chooser deliberation cause more respect for autonomy? Chapter 3 again relies on the experimental design in Figure 1. In my experiment, Choosers are to make an ultra-high-stakes decision that relates to risk–reward trade-offs and involves some temptation: Choosers can earn money by opening virtual boxes containing $20 each, but one box contains a "curveball" that eliminates all earnings.

I use a survey experiment with a sample representative of the United States to investigate rule-setting given different political economies of Choice Architects. First, to investigate the possible substitution between soft and hard rules, I let Choice Architects decide either *(i)* merely on a cap on the Chooser's behavior or *(ii)* on the cap *and* whether the Chooser has to wait one day before making his decision. A highly-powered general population experiment (Mutz, 2011) demonstrates that Choice Architects who impose the waiting period do not "go easy" on the cap, suggesting that mandatory waiting periods—a softly paternalistic policy that has found widespread use in healthcare, firearms regulation and the dissolution of marriage—are add-on restrictions that monotonically increase the regulatory burden. About 40% of Choice Architects whose political economy allowed the simultaneous imposition of caps and waiting periods did impose the waiting period.

In another set of treatments, I vary the temporal characteristics of the Chooser's decision exogenously. Does it matter whether Choosers must decide immediately, after one day, or with the right to revise an initial choice? I find that Choice Architects do not change the cap as Chooser deliberation increases. However, Choice Architects believe average Choosers to come closer with additional deliberation to Choice Architects' subjectively preferred bliss point. This finding is driven by a forecast reduction in risk-seeking behavior, combined with most Choice Architects believing that the average Chooser opens too many boxes. Yet, deliberation does not change the rules set by Choice Architects. One explanation consistent with my findings is that softly paternalistic policies such as waiting periods are predominantly aimed at Choosers with small norm deviations, while hard restrictions on behavior are intended to target extreme decision-makers.

This thesis contains early experimental investigations of paternalism and rule-making. While rules are constraints on behavior often related to economic action, their design has so far been rarely studied systematically. In contrast to the prevailing approach in economics, this thesis is not concerned with equilibrium notions of exchange in institutions, but with *choice between institutions*. As we demonstrate in Chapter 1, paternalism is prevalent across fields and, to some extent, predictable. Moreover, policymaker misconceptions abound. Chapter 2 represents theory-driven experimental work on the relationship between autonomy and knowledge. Knowledge is shown to be an important determinant of the social contract. Finally, Chapter 3 studies paternalistic rule-making when multiple policy tools are available, revealing that they do not offset each other. An increase in the number of available policy tools appears to imply an increasing total regulatory burden. I hope that these findings can inspire future research on the inherently subjective nature of policymakers' actions.

Chapter overview

Chapter 1

Grossmann, M. R. P. and Ockenfels, A. (2024). “Paternalism in Data Sharing”. Mimeo.

This Chapter was co-authored with Axel Ockenfels. Both authors contributed equally to the project. The research idea was jointly developed by Axel Ockenfels and me. We jointly designed the experiment, which was implemented and conducted by me. I carried out the statistical analyses and wrote the draft. Axel Ockenfels gave feedback on the draft.

Chapter 2

Grossmann, M. R. P. (2024). *Knowledge and Freedom: Evidence on the Relationship Between Information and Paternalism*. arXiv preprint. URL: `https://arxiv.org/abs/2410.20970`.

This Chapter was single-authored.

Chapter 3

Grossmann, M. R. P. (2025). *Paternalism and Deliberation: An Experiment on Making Formal Rules*. arXiv preprint. URL: `https://arxiv.org/abs/2501.00863`.

This Chapter was single-authored.

Chapter 1

Paternalism in Data Sharing

This Chapter is co-authored with Axel Ockenfels, Professor of Economics at the University of Cologne and Director at the Max Planck Institute for Research on Collective Goods in Bonn.

Abstract

The privacy paradox is concerned with an ostensible inconsistency between stated and revealed preferences for data sharing: while many people claim to be concerned with privacy, their actual behavior shows little correlation with that concern. The existence of this paradox has recently been questioned on methodological grounds. We study an interpersonal privacy paradox that overcomes these challenges. We test a claim of inconsistency directly: do people want more privacy for others than themselves? We conduct an experiment in which Choosers can state a Willingness-to-Accept (WTA) for the publication of data about themselves. Another group of subjects, impartial policymakers (Choice Architects) can intervene in the Chooser's decision by imposing minimum and/or maximum prices to prevent publication even beyond the WTA. We find that 30 percent of Choice Architects set a minimum price, but only few set a maximum price. The WTA of Choice Architects who set a minimum-price tend to exceed the WTA of non-interventionist Choice Architects. When we correlate minimum prices to interventionists' WTA, we find a strong relationship. Choice Architects don't merely impose their WTA on Choosers: they grant Choosers some liberty to express their own preferences. Choice Architects do not appear to act inconsistently: they are not stricter towards Choosers than towards them-

selves. We find that beliefs over WTAs are strongly and systematically biased. The frequency of low WTAs is underestimated, while that of high WTAs is overestimated.

Acknowledgments

We thank Arno Apffelstaedt and audiences at Cologne and the ECONtribute Workshop on Autonomy & Paternalism for valuable comments. We acknowledge funding by the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation program, as well as by the German Research Foundation (DFG) under Germany's Excellence Strategy. Excellent research assistance by Emma Brauckhoff, Kirsten Marx, Yero Ndiaye, Rosa Wolf and Lea Zaruchas is gratefully acknowledged.

Declarations

Funding was received from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation program under grant agreement number 741409, as well as from the German Research Foundation (DFG) under Germany's Excellence Strategy (EXC 2126/1–-390838866).

The experiment was approved by an Independent Ethics Advisor under the European Union's Horizon 2020 research and innovation program. See Section 1.2.6 for details.

◇

1.1 Introduction

Is there a disconnect between attitudes and actions? The "privacy paradox" suggests so (Acquisti and Grossklags, 2005). Its essence is that individuals behave inconsistently: stated preferences indicate substantial privacy concerns, which are, however, not revealed in action (Barnes, 2006).

A large number of surveys have yielded contradictory and nuanced findings on the relationship between stated preferences and actual behavior (Barth and De Jong, 2017; Kokolakis, 2017). In general, privacy preferences appear to be highly heterogeneous and context-dependent (e.g., Feri, Giannetti, and Jentzsch, 2016; Frik and Gaudeul, 2020; Taddicken, 2014). In a comprehensive critique of the privacy paradox, Solove (2021) notes that discussions about the privacy paradox have tended to overvalue stated preferences and undervalue revealed preferences (or vice versa), but that these measures are inherently incongruent concepts and should not be easily correlated. He writes: "Individual risk decisions in particular contexts indicate little about how people value their own privacy, which is distinct from how people value privacy in general" (p. 5). Solove approvingly cites legal scholar Michael Froomkin on "privacy myopia:" the idea that people simply cannot quite grasp the value of their own data (p. 44). Finally, he argues in favor of solutions that do not rely on people's self-management of their data sharing (p. 50).

We introduce a change in the object of analysis. Several previous studies have dealt with peoples' privacy behaviors given various institutional setups. We concern ourselves instead with decisions *for* others. Our setup is simple: we elicit personal data from a Chooser. The Chooser can state a willingness-to-accept (WTA) for the publication of this data (Hermstrüwer and Dickert, 2017). We augment the classic Becker–DeGroot–Marschak (BDM) mechanism by enabling an impartial, policymaker-like subject—the Choice Architect (CA)—to intervene in the publication of data by setting minimum and/or maximum prices. These limits constrain data sharing beyond the WTA stated by the Chooser. We also elicit CAs' *own* WTA for publication of their data and beliefs about WTAs. Moreover, we have CAs work on Smith, Milberg, and Burke's (1996) Concern for Information Privacy instrument (the CFIP)—a standard survey measure—and we inconspicuously elicit whether they use an ad-blocker and how much time they spent on our experiment's privacy policy.

We can correlate CAs' interventions with their own behavior. This circumvents the conceptual incongruity uncovered by Solove (2021): instead of correlating stated preferences with revealed preferences, we correlate revealed preferences towards oneself with revealed preferences towards a Chooser. By investigating whether people are stricter towards others than they are towards themselves, we can test an allegation of inconsistency similar to the privacy paradox: an *interpersonal* privacy paradox.

We have three core results and contributions to the literature. First, we provide evidence that ideals-projective paternalism can be replicated in the context of data sharing. CAs project their own privacy preferences onto others—what

Ambuehl, Bernheim, and Ockenfels (2021) called "ideals-projective paternalism." In so doing, CAs leave space for the Chooser to express himself. We reject the idea of an interpersonal privacy paradox. Second, our findings suggest that beliefs about others' preferences are systematically biased: the prevalence of high WTAs is overestimated, while that of low WTAs is underestimated. This aspect can be linked to a recent literature on misperceptions (see Bursztyn and Yang, 2022, for a review), and their impact on behavior and attitudes (e.g., Bursztyn, González, and Yanagizawa-Drott, 2020). Finally, we show that CFIP is associated with WTA, but that these measures do not correlate with automatically measured revealed-preference actions. Privacy preferences are evidently highly context-dependent.

This paper proceeds as follows: Section 1.2 presents our experimental design. Section 1.3 develops hypotheses. Section 1.4 tests these hypotheses and discusses our results. Finally, in Section 1.5, we conclude.

1.2 Study design

One class of subjects (Choice Architects, CAs) is enabled to make a rule for a Chooser. The Chooser will be able to state a WTA for the publication of data about himself.[1] CAs can intervene by imposing minimum and maximum prices that constrain data sharing even beyond the Chooser's WTA.

Our experiment consists of three separate sessions. In the first session, subjects were asked to work on a preliminary survey to enable the incentivization of beliefs. The second session constitutes the main part of the experiment and it concerns CAs making rules for Choosers. Finally, the third session serves to implement the decisions of CAs among Choosers. All participants partook in exactly one session.

1.2.1 Generating and sharing data

A common design element in all three sessions is the collection and possible sharing of personal data. All subjects participate in a questionnaire with ten highly variegated questions (HVQs). These questions are available in Appendix A.3. We include HVQs in our study to generate data that can be subsequently shared with other participants and visitors to our department's website.

[1] In this paper, we use male pronouns to refer to the Chooser, and female pronouns to refer to the CA.

Any one answer to a question says little about an individual. However, the inner product of answers may reveal a substantial amount of information. Like Rockenbach, Sadrieh, and Schielke (2021) and others, we employ data that is fundamentally not verifiable. Subjects received a flat payment of €4 for answering these questions. In the next step, we asked subjects whether they would like to donate €3 to UNICEF. This serves to elicit a proxy for altruism as well as to augment the shareable data by a choice variable that affects the subject's payoff.[2]

Besides participating in the CFIP survey instrument, subjects were asked to state a WTA for the sharing of these data. These WTAs were used in a modified BDM mechanism, similar to the approach by Hermstrüwer and Dickert (2017). Simply put, the BDM mechanism draws a price $p \in R$, where $R = [0, 0.01, \dots, 4]$. If the WTA is z, then data sharing takes place if and only if $p \geq z$. All subjects were allowed to indicate a WTA of "more than €4." Since the random prices in our mechanism were restricted to be between €0 and €4, this ensured that no sharing could take place.

1.2.2 Preliminary survey

In session 1, we collected initial data on WTAs to enable the incentivization of beliefs in session 2. We advised 50 % of subjects in this session against sharing data. In total, 20 subjects participated in this session.

1.2.3 Interventions

Our BDM mechanism is modified in one fundamental way: We allow a particular class of subject in session 2 (Choice Architects, CAs) to make rules for future subjects in session 3 (Choosers). As we describe in Section 1.3, this is to test for paternalistic intervention in the sharing of data.

As stated above, our BDM mechanism was constrained to yield random prices between €0 and €4. Beyond that, CAs were granted the opportunity to indicate additional regions of R in which sharing was not to take place. More precisely, in session 2 of the experiment, CAs can set *minimum and maximum prices* to be applied in the BDM. These values do not change R, but they block

[2]The inner product of verifiable data is easily used to deanonymize individuals. One such instance is recounted in Kearns and A. Roth (2019, ch. 1): When the Commonwealth of Massachusetts released "anonymized" health care records, Sweeney (2000) was able to reidentify many individuals using census data and the inner product of zip code, birthdate and sex, the (only) personal information in the released medical records. To avoid getting subjects actually deanonymized, we use non-verifiable data.

the sharing of data.[3] This modification to the BDM mechanism implements minimum and maximum prices while not affecting the incentive compatibility of the mechanism. For example, suppose the CA set a minimum price of €0.75. The Chooser has a WTA of €0.40. If the random price chosen through our mechanism is between €0.40 and €0.74, no data would be shared although the price exceeds the Chooser's WTA. Simply put, the minimum price can be understood as the CA overriding a WTA below the minimum price.

After responding to the HVQs and making the donation decision, CAs have to correctly answer rigorous control questions about nuances of the above mechanism and its implementation before proceeding to the interventional stage. Firstly, we ask CAs whether they would like to make a recommendation to the future subject.[4] They can choose to make no recommendation, to recommend sharing data, or to advise against sharing. After so doing, they can intervene more directly in the future subject's behavior by setting a minimum and/or a maximum price. If they refrain from setting both minimum and maximum prices, the Chooser plays the original unconstrained BDM mechanism.

Despite findings by De Quidt, Haushofer, and C. Roth (2018) that experimenter demand might be a fringe phenomenon with limited influence, we go to great lengths to prevent anchoring or leading the behavior of subjects. In any case where a default had to be set by us, we set the defaults to be in line with a null hypothesis of libertarianism. This means that a subject must consciously change the default to intervene. Even if a CA wishes to impose minimum or maximum prices for sharing data, the CA actively needs to opt to show the respective form fields, contrary to being simply presented with the instruments to indicate minimum or maximum prices.

Subjects in session 1 and CAs (in session 2) state their WTA without an intervention (that is, they play the original unconstrained BDM mechanism).

1.2.4 Further privacy preferences elicited from CAs

Beyond the WTA, we elicited whether CAs use an ad-blocker, and we measure the time they spent on the experiment's privacy policy. Ad-blocker usage and time spent on the experiment's privacy policy are automatically measured during the experiment without the subject's intervention. Modern internet ads are well-

[3]We impose only that the maximum price is at least equal to the minimum price.

[4]CAs know that their decisions are only implemented with a certain probability because there are more CAs than future Choosers.

known to track users (Falahrastegar et al., 2014), and ad-blockers allow users to improve their privacy (Garimella, Kostakis, and Mathioudakis, 2017). We track the time spent on the experiment's privacy policy because it should correlate with subjects' latent concern for privacy.

CAs also work on the CFIP instrument (Smith, Milberg, and Burke, 1996). We use the translated version by Harborth and Pape (2018). The CFIP is a standard survey measure to gauge respondent's subjectively perceived importance of privacy. It is widely used and has been cited more than 3,200 times according to Google Scholar.

This results in seven measures of privacy preferences: *(i)* CAs' WTA, *(ii)* use of an ad-blocker, *(iii)* time spent on the experiment's privacy policy, *(iv)–(vii)* CFIP's "Collection," "Errors," "Unauthorized Secondary Use" (USU) and "Improper Access" (IA) scores.

We also ask CAs which proportion of subjects from session 1 fell into each of the following six categories with respect to their WTA: exactly 0, $[0.01, 1]$, $[1.01, 2]$, $[2.01, 3]$, $[3.01, 4]$, greater than 4. We imposed that the sum of proportions equal 100 percent. We incentivize beliefs using the binarized scoring rule by Hossain and Okui (2013).[5] The incentivization was based on data from session 1.

1.2.5 Session 3

This session implemented the minimum and maximum prices set by CAs. Since there were more CAs (132) than subjects in the implementation session (13), CAs were randomly drawn to have their decisions implemented.

1.2.6 Implementation

The experiment was conducted online in September and October 2021 at the Cologne Laboratory for Economic Research. Subjects were recruited with ORSEE (Greiner, 2015). The experiment was implemented in oTree (Chen, Schonger, and Wickens, 2016). English-language instructions for all sessions are available in Section A.5 in the Appendix. The design was approved by an Independent Ethics

[5]This belief elicitation is done twice for each CA: Once in an unrestricted, baseline, fashion and later given that subjects were advised against sharing data (the two treatments present in session 1). When stating the second set of beliefs, CAs were shown their baseline beliefs to allow them to move the probability mass as they thought proper. This allows the analysis of beliefs about the causal effect of advice. We compared these beliefs to baseline beliefs in Appendix A.2. CAs generally believe that advising against sharing data causes an increase in Chooser's WTA.

Advisor under the European Union's Horizon 2020 research and innovation program on July 7, 2021, and experiments were authorized on August 26, 2021 by the responsible European Research Council ethics officer. The experiment was preregistered at AsPredicted.org.[6]

We optimized our experiment to enable it to run in all JavaScript-enabled browsers and subjects were allowed to use a modern device of their choice.

All CAs were asked to participate at a time of their choosing between 10am and 6pm on a given day. The average completion time in session 2 was 23 minutes and the average payment was €9.53, including a €2.50 show-up fee. Subjects were paid using PayPal. Before making the central decisions in the experiment, CAs were asked to complete two sets of rigorous control questions. Only subjects who correctly answered all control questions were allowed to continue. 68.5 % of all subjects correctly answered at least one set of control questions on their first attempt, thereby earning a small bonus.

CAs could not increase the Chooser's payoff by setting a minimum price, and our final control question focused on exactly that nuance. We established a telephone hotline for those subjects who needed help to answer the control questions. We received only few calls, all of which could be resolved quickly.

Of 132 CA observations, 130 observations are *complete* as defined by the preregistration.[7] All subsequent analyses are solely concerned with complete observations, although the probability of implementation for the two excess CAs was strictly positive.

1.3 Hypotheses

Given the opportunity to impose both minimum and maximum prices for the sharing of data, what will CAs do? Note that most privacy regulations are about the implementation of minimum standards. Firms are virtually always able to offer more protection of personally identifiable information.[8] How does that translate to our setting? By setting a minimum price, CAs can provide a kind of

[6]The preregistration is available at `https://aspredicted.org/JFC_X2N`.

[7]Our preregistration stated: "[Completion] is defined as follows: A participant who stated their privacy attitudes using the CFIP by Smith et al. (1996). (This is the last measure collected in our experiment and one can only proceed to the CFIP if one has fulfilled all other parts of the experiment.)" We did not permit excess subjects to complete the CFIP, resulting in exactly 130 complete observations.

[8]One exception is that firms have to comply with requests by law enforcement. However, the amount of data that companies must keep about their customers depends on the jurisdiction.

paternalistic protection to Choosers if the latter are thought to underestimate the cost of revealing their data. On the other hand, a maximum price does not provide such protection. A limit on the amount earned by the Chooser could be justified based on social preferences. Moreover, monetary payments for certain transactions are often viewed as coercive or repugnant, perhaps because they induce decision-makers to transact in the first place (Ambuehl, 2024; Ambuehl, Niederle, and A. E. Roth, 2015). Hypothesis 1 is a null hypothesis that allows us to test the relative magnitude of these concerns.

Hypothesis 1. *CAs protect Choosers.* CAs are as likely to set minimum prices as they are to set maximum prices.

Moreover, we test whether Ambuehl, Bernheim, and Ockenfels's central finding can be replicated in a different setting. Our study design allows us to correlate minimum and maximum prices with CAs' WTAs. We also use a preregistered analysis that correlates generalized preferences for privacy with minimum prices.

Hypothesis 2. *Projective paternalism.* Interventions are positively associated with CAs' own preferences.

This hypothesis inherently relates to the distinction between what Ambuehl, Bernheim, and Ockenfels called *mistakes-projective* and *ideals-projective* paternalism (pp. 815–817). In the former case, interventions express that CAs' own choice is undesirable. In the latter case, CAs' own choice is optimal and interventions are used to meliorate Chooser decisions in the direction of CAs' preferences.

Ambuehl, Bernheim, and Ockenfels (2021) found that Choosers are generally left with some scope to express their own preferences. In other words, CAs do not intervene by imposing their own preference; rather, they only intervene "in the direction" of their own preference (see also their proposition 1, p. 814). By relating CAs' WTAs to minimum prices, we can directly test the idea of an "interpersonal privacy paradox" hinted at in Section 1.1. CAs may act hypocritically in their decision to intervene. In that case, a CA's minimum price would be systematically larger than her WTA. The following null hypothesis allows us to test for the relationship between minimum prices and CAs' WTAs.

Hypothesis 3. *Interpersonal privacy paradox.* Minimum prices are equal to CAs' WTA.

1.4 Results

We now test hypotheses 1–3 econometrically. We also report additional exploratory analyses on beliefs, recommendations and on the robustness of privacy preferences.

1.4.1 Projective paternalism

1.4.1.1 How do CAs intervene?

Of 130 CAs, 39 (30 %) set a minimum price.[9] The 95 % confidence interval for the proportion of CAs who set a minimum price is [0.223, 0.387]. Of 130 CAs, 13 (10 %) set a maximum price.[10] The resulting 95 % confidence interval for the proportion of CAs who set a maximum price is [0.054, 0.165]. A within-CA paired t-test on dummies for "set a minimum price" and "set a maximum price" reveals $t = 4.27$, $p < 0.001$. Thus, in general, CAs intervene to prevent prices that are too *low*—not too high.

Result 1. A significant proportion—30 %—of CAs intervenes; the proportion of CAs who set a maximum price is less than the proportion of CAs who set a minimum price. Hypothesis 1 is rejected.

We now proceed to investigate the relationship between privacy preferences and interventions.

1.4.1.2 Relating WTAs and minimum prices

Figure 1.1 plots CAs' WTAs against *all* minimum prices, including €0 if no minimum price was set (the natural lower bound in the BDM mechanism). 52 out of 130 CAs lie on the 45° line, indicating that these CAs' minimum price equals their WTA. While 64 CAs set a minimum price below WTA, only 14 set a minimum price above. On the sample level, this asymmetry confirms a prediction from Ambuehl, Bernheim, and Ockenfels's (2021) theoretical framework (their proposition 1). A Wilcoxon signed rank test confirms that minimum prices tend to be lower than CAs' WTAs ($V = 2846$, $p < 0.001$). For CAs with both a

[9] As preregistered, we define a minimum price as having been set if the CA set it to at least €0.01.

[10] Analogously to our definition of "set a minimum price", we define "set a maximum price" as the maximum price being less than €4.01, the default value.

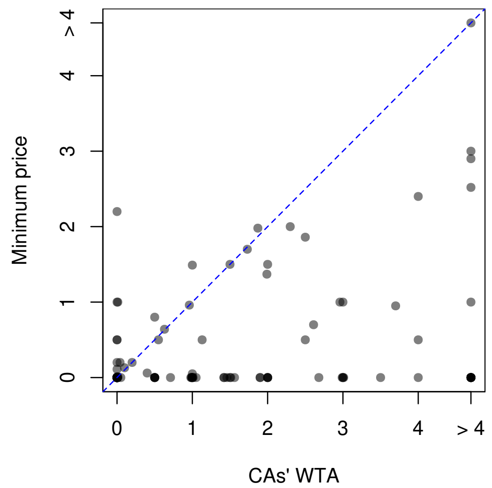


Figure 1.1: Minimum prices and CAs' WTAs

WTA *and* a minimum price of at most 4,[11] this is confirmed in a paired t-test, $t = 5.95$, $p < 0.001$. The general tendency is to grant some freedom of choice to Choosers.

Result 2. Minimum prices are generally lower than CAs' WTAs (rejecting hypothesis 3). CAs tend to leave some space to Choosers to express their own preferences. However, there is a significant minority of CAs that sets the minimum price at exactly their WTA.

In the parlance of Ambuehl, Bernheim, and Ockenfels (2021), CAs who set a minimum price above their WTA would be said to make a "mistake"—they know what is the optimal choice, but they cannot resist choosing a sub-optimal action. In our setting, only few CAs make a "mistake." We thus also find that there is no evidence for an interpersonal privacy paradox. In general, minimum prices are below CAs' WTA—Choosers are required by CAs to guard their data to a lesser extent than CAs guard themselves. Neither do CAs just impose their own preferences nor are their preferences just what they impose on others—but CAs give future participants some wiggle room to deviate from the former's precepts. While this indicates that minimum price and WTA are distinct concepts to CAs, we now proceed to demonstrate that they are highly correlated.

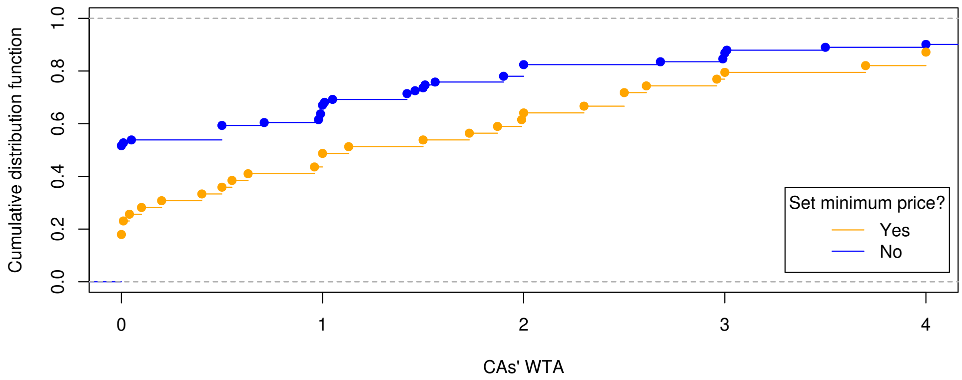


Figure 1.2: Distribution of CAs' WTAs conditional on interventions

1.4.1.3 Extensive margin

As specified in the preregistration, we apply a one-sided two-sample Mann–Whitney U test to our choice data. More precisely, we test the null hypothesis that the WTA of CAs who did not set a minimum price is equal to or greater than the WTA of CAs who set a minimum price. The alternative hypothesis is that CAs who did not set a minimum price have a lower WTA than those who did. Accepting the alternative hypothesis would constitute evidence in favor of ideals-projective paternalism. The null hypothesis is rejected at the 5 % level ($W = 1219$, $p = 0.002$). Among all CAs, median WTAs are €1.13 (set a minimum price) and €0 (did not), an economically significant difference.

Figure 1.2 shows that the distribution of WTAs for those who did not set a minimum price first-order stochastically dominates that of CAs who did. A two-sample Kolmogorov–Smirnov test rejects the null hypothesis of identical distributions ($p = 0.004$). We can confirm this association in a logistic regression. The outcome is whether the CA set a minimum price. In Table A.1 in the Appendix, we find a coefficient on WTA that is significantly different from 0 ($z = 2.57$, $p \approx 0.01$). Taken together, we conclude from this evidence:

Result 3. On the extensive margin, CAs' WTAs are positively associated with the decision of whether or not to set a minimum price.

Before we can relate these findings to hypothesis 2, we have to concern ourselves with the intensive margin of interventions.

[11]Among all CAs, only one elected to set a minimum price of more than 4, i.e., to ban publication completely.

1.4.1.4 Intensive margin

In this section, we investigate the specific minimum price that was set by the CA *if the CA did set a minimum price*. Recall that 39 of 130 CAs (30 %) set a minimum price.

Empirical strategy Ideally, we would wish to estimate

$$\text{minimum}_i = \beta_0 + \beta_1 \text{privacy}_i + \varepsilon_i,$$

where privacy_i are objective measurements of true privacy preferences.

First, note that the dependent variable—the minimum price—is censored from above as no minimum price could be set in excess of €4. Any minimum price that may have exceeded €4 was recorded as €4.01. However, the minimum price is not censored from below; $\text{minimum}_i = 0.01$ represents a corner solution (Wooldridge, 2010, sec. 19.2.3) which is not caused by censoring but because this analysis is solely concerned with the intensive margin.

Moreover, note that privacy_i is not observed. It would be simple to relate the minimum price on the CA's WTA—as we did above—but this does not provide an answer to the research question of whether *privacy preferences* among those who intervene translate to a higher minimum price unless WTA and privacy preferences were perfectly correlated.[12] Privacy preferences are latent psychological constructs which cannot be elicited directly. To account for this measurement error, we apply the *Obviously Related Instrumental Variables* (ORIV) method introduced by Gillen, Snowberg, and Yariv (2019). This approach uses instrumental variables models to allow us to include multiple noisy measurements of a latent construct (such as privacy preferences) in regression models. In the simple case where two measurements of the same underlying construct are elicited, these measurements are first normalized. The dataset is then duplicated; in the upper half, the first measurement is designated as regressor and the second measurement is used as the instrument. In the lower half of the dataset, the second measurement is the regressor and the first measurement acts as the instrument. On this new dataset so constructed, an instrumental variables model is applied. Clearly, this method would not be valid if the standard errors were not adjusted to account for the duplication of data (any test on the coefficients would be

[12]In Table A.2 in the Appendix, we run a Tobit regression of minimum prices on CAs' WTA. Unsurprisingly, this model confirms the previous findings on the relationship between WTA and interventions.

	Coefficient	Clustered Std. Err.	z	p-value, two-tailed	p-value, one-tailed
(Intercept)	1.121	0.151	7.409	0.000	
privacy_i	0.854	0.471	1.814	0.070	0.035
Log Pseudolikelihood	−4489.922				
Num. obs.	1638				
Num. clusters	39				

Table 1.1: IV Tobit model on the intensive margin

misspecified). Hence, Gillen, Snowberg, and Yariv (2019) suggest to compute standard errors that are clustered on each genuine observation of the *dependent* variable.

As stated in our preregistration, we have seven measures of privacy preferences for each CA: (1) their WTA; (2) the time spent on the experiment's privacy policy; (3) whether subjects block ads in their web browser; (4)–(7) CFIP's Collection, Errors, Unauthorized Secondary Use and Improper Access scores. These measures are discussed in Section 1.2.4. After normalization, the ORIV dataset is constructed in a fashion analogous to the two-measurement case: Each of the seven measurements is used once as the regressor and, for each given regressor, every other measurement is used as an instrument.[13] The newly constructed dataset has $39 \cdot 7 \cdot (7-1) = 1638$ rows. In accordance with our preregistration, we apply a Tobit IV model, with standard errors clustered on each genuine observation of the dependent variable, minimum_i. The estimated model is shown in Table 1.1. We then run a preregistered one-tailed z-test of the coefficient on privacy_i; the null hypothesis is that this coefficient is zero or negative. Since $z = 1.814$, the null hypothesis is rejected at the 5 % level ($p = 0.035$).

Result 4. The privacy preferences of CAs are significantly and positively associated with the minimum price they set if they set a minimum price, i.e. on the intensive margin. Hypothesis 2 is confirmed.

Clearly, this result replicates previous findings of ideals-projective paternalism in CAs in a novel setting. Moreover, *generalized* privacy preferences correlate with the intensive margin. However, the statistical significance is much decreased from Table A.2 in the Appendix to Table 1.1. This is because there is

[13]We code CAs' WTAs that exceed €4 as €4.01. Coding these cases as €10, €100 or €1000 does not substantially change the following conclusions.

only a small correlation between our seven measures of privacy preferences. We discuss this issue further below, in Section 1.4.4.

1.4.2 Beliefs

In this Section, we conduct exploratory analyses to compare actual WTAs against beliefs about WTAs. Models 1–2 of Table 1.2 present Ordinary Least Squares models of the following regression:

$$\text{Belief} = \sum_{j=1}^{6} \beta_j D_j + \varepsilon, \tag{1.1}$$

where D_j are indicators for our six WTA categories. Model 1 presents sample average beliefs about the prevalence of each WTA category. Hence, the outcomes are CAs' estimates about the proportion of WTAs in any given category. For each CA, we obtain six estimates, giving a total sample size of $6 \times 130 = 780$. Each of these estimates turns on one of the indicators on the right-hand side of Equation 1.1. We provide cluster-robust standard errors on the CA level. On average, $r = 0$ is thought to be the modal category. However, CAs with a high WTA (defined as an above-median WTA) believe $r > 4$ to be most popular, followed by the second-highest category.

We now compare CAs' beliefs against the true distribution of CAs' WTA. We augment the regression in Model 1 by adding a one-hot encoding of actual WTAs: a binary indicator indicating in which category each CA's WTA falls. We can then estimate the following regression equation:

$$Y = \sum_{j=1}^{6} \beta_j D_j + \sum_{j=1}^{6} \gamma_j \Delta D_j + \epsilon, \tag{1.2}$$

where ΔD_j is an indicator for data on beliefs and Y is a belief or an indicator of whether a WTA belongs to any of the six categories. This results in estimates for γ_j that are the differences between the true distribution of WTAs and beliefs. A $\gamma_j < 0$ indicates that CAs underestimate the prevalence of a particular category. Similar to Equation 1.1, the β_j in Equation 1.2 represent the proportion of WTAs that fall into each category. In model 3 in Table 1.2, we find that CAs tend to underestimate the prevalence of low WTAs, and overestimate the prevalence of high WTAs. For example, 41.5% of CAs have a WTA of exactly €0, but the

	Model 1	Model 2	Model 3
WTA = 0	0.320***	0.181***	0.415***
	(0.025)	(0.026)	(0.044)
0 < WTA ≤ 1	0.119***	0.115***	0.200***
	(0.009)	(0.011)	(0.035)
1 < WTA ≤ 2	0.129***	0.142***	0.154***
	(0.008)	(0.012)	(0.032)
2 < WTA ≤ 3	0.143***	0.173***	0.077**
	(0.009)	(0.012)	(0.024)
3 < WTA ≤ 4	0.139***	0.192***	0.046*
	(0.011)	(0.017)	(0.019)
WTA > 4	0.150***	0.197***	0.108***
	(0.015)	(0.026)	(0.027)
Δ [WTA = 0]			−0.096*
			(0.038)
Δ [0 < WTA ≤ 1]			−0.081*
			(0.035)
Δ [1 < WTA ≤ 2]			−0.025
			(0.031)
Δ [2 < WTA ≤ 3]			0.066**
			(0.025)
Δ [3 < WTA ≤ 4]			0.093***
			(0.020)
Δ [WTA > 4]			0.042
			(0.022)
Coefficients represent	Beliefs	Beliefs	True proportion
Subset	—	High WTA	—
Standard errors	Clustered HC3	Clustered HC3	Clustered HC3
R^2	0.569	0.581	0.336
Adj. R^2	0.566	0.574	0.331
Num. obs.	780	372	1560

$^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$

Table 1.2: Beliefs and true proportion of WTA categories

average estimate of the proportion having such a WTA is 32.0%, implying a difference of $\gamma_1 \approx -9.6\%$.

We can also construct a lower bound on the conditional expectation implied by beliefs.[14] Appendix A.4 describes this in more detail. The mean lower bound is 1.029 (95% confidence interval: [0.908, 1.149]). However, the mean conditional WTA is only 0.839 (95% confidence interval: [0.632, 1.046]). While we cannot reject the null hypothesis that these random variables have the same mean (two-sample t-test, $t = -1.52$, $p = 0.12$), that *even the lower bound* directionally tends to exaggerate the conditional mean supports the idea that beliefs are, in general, too large. This is confirmed by a Mann–Whitney U test in the Appendix, but it should be noted that we only elicited forecasts about proportions, not (conditional) means.

Result 5. Beliefs about WTAs are systematically biased upward.

This result hints at the existence of systematic misconceptions about others' privacy behavior (e.g., Bursztyn and Yang, 2022). Such misconceptions have been shown to be a driver of attitudes and behavior. For example, in a study of pluralistic ignorance, Bursztyn, González, and Yanagizawa-Drott (2020) informed Saudi men about the proportion of other Saudi men accepting women's work outside the home. This correction of overly pessimistic beliefs caused a change in attitudes and also led to more women seeking employment outside the home. Similar mechanisms could also be an important determinant of demand for privacy regulation. This idea represents a promising avenue for future research.

Nonetheless, these results should be viewed as exploratory: First, note how CAs stated beliefs about *Chooser* behavior, not about CA behavior. We view it as acceptable to compare both quantities as both CAs and Choosers come from an identical subject pool. It is not immediately apparent why beliefs would be biased against CAs and not vis-á-vis Choosers. Second, however, there may be secular factors that lead subjects to reveal biased beliefs. For example, we directly elicited beliefs about proportions, not relative frequencies (Price, 1998; Schlag, Tremewan, and Van der Weele, 2015). Subjects' numeracy can influence the elicitation of beliefs.

[14]That is, conditional on the WTA being at most €4.

1.4.3 Recommendation

68 out of 130 CAs recommend to their Chooser to share data, 4 recommend against it, and 58 do not send a recommendation. As Table A.5 in the Appendix indicates, CAs with a higher WTA are less likely to make this recommendation. However, the proportion of recommenders is not associated with the propensity to intervene through setting a minimum price.

How can these findings be reconciled? On the one hand, not sending a message—or even advising Choosers not to share data—can be a substitute for intervention. We observe no empirical support for the latter theory, but future work can causally investigate the substitutional relationship between various paternalistic interventions. On the other hand, setting a minimum price guarantees that the Chooser will never sell his data for less than the minimum price. If CAs wish to protect Choosers from prices that are too low, the implementation of a minimum price makes sharing safer—and thus perhaps *less* objectionable.

1.4.4 Other measures of privacy preferences

Table A.3 in the Appendix relates measures *(ii)*–*(vii)* to CAs' WTA. CFIP scores are robustly but only to a small degree correlated with WTA. Note from the intercepts in models 3–6 that baseline CFIP scores in our sample are substantial. Given the bounded nature of CFIP scores, the extent to which WTA can matter is inherently limited.

On the other hand, ad-blocker usage and time spent on the privacy policy do not appear to correlate with WTA. Moreover, Table A.4 in the Appendix reveals that CFIP scores are unrelated to ad-blocker usage and time spent on the privacy policy (even at a joint test of significance). We could only speculate why this is the case. However, these findings paint a nuanced picture: measures that directly prompt respondents to reveal their preferences seem to elicit similar latent constructs. On the other hand, such measures, while attractive to implement in surveys and experiments, show little external validity to unobtrusively and automatically ascertained measures of revealed privacy-protecting behavior.

1.5 Conclusion

In this study, we made the first foray into the issue of paternalism regarding others' decisions to share personal data. Other forms of paternalism might manifest as size limits for sugary drinks or taxes on undesirable "junk food," but paternalism can also manifest in the area of privacy.

We allowed CAs to set minimum prices for the publication of personal data by future participants. 30% of CAs set a minimum price. Both on the extensive and the intensive margins, this tendency was significantly associated with CAs' own privacy preferences. Furthermore, we found that CAs appear to underestimate the willingness to share data. Echoing a concern in recent methodological critiques of the privacy paradox, our findings suggest that common measures of privacy preferences have limited external validity.

Most importantly, our study reveals that CAs set minimum prices that are systematically smaller than their own WTA. This is in line with "ideals-projective paternalism" (Ambuehl, Bernheim, and Ockenfels, 2021) and is evidence against the existence of an interpersonal privacy paradox. Subjects do not show inconsistency in their interventions. Future research can further investigate the existence of beliefs about privacy preferences and the apparent prevalence of misconceptions—and how such misconceptions shape the demand for policy.

Chapter 2

Knowledge and Freedom: Evidence on the Relationship Between Information and Paternalism

Abstract

When is autonomy granted to a decision-maker based on their knowledge, and if no autonomy is granted, what form will the intervention take? A parsimonious theoretical framework shows how policymakers can exploit decision-maker mistakes and use them as a justification for intervention. In two experiments, policymakers ("Choice Architects") can intervene in a choice faced by a decision-maker. We vary the amount of knowledge decision-makers possess about the choice. Full decision-maker knowledge causes more than a 60% reduction in intervention rates. Beliefs have a small, robust correlation with interventions on the intensive margin. Choice Architects disproportionately prefer to have decision-makers make informed decisions. Interveners are less likely to provide information. As theory predicts, the same applies to Choice Architects who believe that decision-maker mistakes align with their own preference. When Choice Architects are informed about the decision-maker's preference, this information is used to determine the imposed option. However, Choice Architects employ their own preference to a similar extent. A riskless option is causally more likely to be imposed, being correlated with but conceptually distinct from

Choice Architects' own preference. This is a qualification to what has been termed "projective paternalism."

Acknowledgments

We thank Arno Apffelstaedt, Simon Brandkamp, Marius Gramb, Vitali Gretschko, Susanna Grundmann, Sophia Hornberger, Moritz Janas, Sander Kraaij, Axel Ockenfels, Martin Ricketts, Paula Scholz, Frederik Schwerter, Itai Sher, Sven A. Simon, and Robert Slonim, audiences at Buckingham, Cologne, Maastricht, the IMEBESS in Lisbon, the ASFEE in Montpellier and the ESA World in Lyon for valuable comments. All remaining errors are our own. We thank Martin Strobel for support in running the experiment. We acknowledge funding by the Center for Social and Economic Behavior (C-SEB) at the University of Cologne and the German Research Foundation (DFG) under Germany's Excellence Strategy (EXC 2126/1–390838866).

Declarations

Financial support was received from the Center for Social and Economic Behavior (C-SEB) at the University of Cologne and the German Research Foundation (DFG) under Germany's Excellence Strategy, EXC 2126/1–390838866. The author has no competing interests to declare that are relevant to the content of this article.

Data availability, IRB approval

All software and data used for this paper are freely and publicly available. See Section B.1 in the Appendix for details.

The experiments including key analyses were preregistered and IRB approval was granted. For details, see Sections 2.3.4.3 and 2.3.5.2 in the main text.

⋄

2.1 Introduction

To successfully act on our preferences, we need to know what we are doing. Government regulation often conditions freedom of choice on decision-maker knowledge, and decision-maker ignorance has long been understood as a justification for intervention. Many legal rules protect decision-makers and others. For example, driver's licenses restrict the use of high-powered vehicles on public roads to decision-makers who possess a basic understanding of rules and common road practices. Other such regulations target only the decision-maker. For example, many countries restrict the offering of advanced financial products to the general public.[1] Parents' everyday accounts conform to this story: a recent survey found that American parents' overriding concern about tattoos is that their child may regret it later (C. S. Mott Children's Hospital, 2018). The possibility of decision-maker regret also appears to play an important role in the opposition to organ markets (e.g., Ambuehl, Niederle, and A. E. Roth, 2015; Clemens, 2018; A. E. Roth, 2007). Regret is possible because we are ignorant about our future values and experiences. We lack access to these states of mind. These examples share a common thread: individual freedom can be curtailed based on a perceived lack of knowledge.

Policymakers can exercise their authority to restrict the choices of decision-makers. They can also choose to provide decision-makers with information central to their decision-making. Moreover, policymakers' own knowledge can shape the rules that they impose. Economists have tended to focus on equilibrium behavior in institutions, without much consideration for causal determinants of institutional design. However, policy is invented through conscious and purposeful human design. What determines elements of this design? We use experiments and simple economic theory to investigate how and why experimental policymakers *actually govern* over decision-makers in the presence of asymmetric information on either side.

Does decision-makers' knowledge cause them to obtain more autonomy from impartial policymakers? We first formulate a parsimonious formal theory of optimal paternalism with and without decision-maker mistakes. Chooser's

[1] The European Union's MiFID II mandates that investors prove some experience or knowledge about financial markets before they can be offered certain products. Under the U.S. Securities Act of 1933, Rule 506 of Regulation D establishes rigorous requirements to become an "accredited investor," ensuring that only those with tremendous financial resources can invest in unregistered securities.

mistakes can arise from a lack of knowledge. In contrast to the standard approach in welfare economics, we allow policymakers ("Choice Architects", CAs) to be non-neutral—to prefer that the Chooser choose the CA's subjectively preferred option. The model has only two parameters: one for the weight CAs place on their own preference, and a belief about the Chooser's preference. CAs can elect to have the Chooser have his own choice, or to impose one option on him.[2] Our theory shows that Chooser mistakes can make freedom of choice less attractive if CAs are sufficiently neutral. However, there can be a strategic advantage to Chooser mistakes. The model nests John Stuart Mill's (1869) idea of respecting fully informed choices and a preference for information provision over imposition.

We conduct two simple experiments to test the influence of Chooser knowledge on the autonomy they are granted. In both experiments, Choosers choose between a potentially ambiguous lottery and a safe amount. A CA is matched to each Chooser. She can leave the Chooser free to decide or impose an option on him. Interventions are made conditional on the Chooser's knowledge. We vary the amount of knowledge that Choosers have about the probability in the lottery. However, CAs face no ambiguity.[3] Both experiments demonstrate that CAs respect increased Chooser knowledge. Few CAs intervene without regard to knowledge. Across both experiments, full information causes a more than 60% reduction in intervention rates. As in previous work, intervening CAs tend to impose their own preference between the options.

Knowledge can be enhanced or diminished by policymakers. If policymakers can send true information to decision-makers, is this opportunity used strategically? Since Mill, information provision has long been recognized as an alternative to prohibitions. Blackwell (1953) proved that true information must make a decision-maker better off according to his own subjective utility. Thus, to most economists, decision-maker mistakes are inherently linked to the correctness of beliefs. For example, under this criterion, it may well be the case that some individuals should smoke *more* if they overestimate the risks from smoking (e.g., Caplan, 2011). Some evidence for such overestimation has been found (e.g., Viscusi, 1990). In such cases, economists may argue for a correction of false beliefs. On the other hand, policymakers can engage in "paternalism by

[2] We follow the convention that the Choice Architect is female and the Chooser male.

[3] Experiment 1 uses a version of partial ambiguity (e.g., Camerer and Weber, 1992; Chew, Miao, and Zhong, 2017; Chipman, 1963; Gigliotti, 1996): Choosers observe the lottery k times before choosing. In experiment 2, Choosers face a fully ambiguous lottery or no ambiguity whatsoever.

omission" by deliberately keeping decision-makers uninformed if that leads to policymakers' preferred option being more likely to be chosen. Our theory shows that mistakes are not created equal: departures from true preferences that go in the direction of policymakers' subjectively preferred option can benefit the CA, and in turn mute the provision of true information.

In experiment 2, we allow CAs to construct the information sent to a Chooser. Overall, about 80% of CAs do provide information to Choosers. An additional analysis demonstrates that non-provision of information is indeed associated with a belief that Chooser mistakes go in the direction of CAs' subjectively preferred option. The presence of this combination of beliefs and CA preferences is associated with a 13.4 p.p. reduction in information provision. Our theoretical and empirical study of strategic information provision is related to Grossman's (1981) and Milgrom's (1981) early conceptualizations of strategic communication. In these models, a sender can withhold information but not misrepresent it because it is fundamentally verifiable (Kartik, 2009). In the present paper, verifiability arises because experimental Choosers cannot be deceived.

Policymakers can, in general, use both information provision and intervention in choices. As outlined above, if it is certain that the Chooser will get his choice, more information must make him better off. However, if intervention can follow information provision, non-provision can be employed strategically to provide a justification for intervention. In a treatment, we randomly enable some CAs to intervene in the resulting choice simultaneous to deciding on information provision. In general, we find that whether or not CAs can intervene makes no difference regarding the information they provide to Choosers. However, intervening CAs are 29.7 p.p. less likely to provide information to Choosers. A novel finding is that only a small but highly statistically significant proportion of subjects—about 3.3% of CAs—appear to provide information strategically in this way.

So far, we have discussed instances of policymakers simply knowing more than a decision-maker, and asking how the former react to this asymmetric information under varying political economies (imposition *vs.* information provision). However, policymakers had to rely on their own beliefs or preferences when deciding what to impose onto the decision-maker. Suppose now that the decision-maker is ignorant, but the policymaker knows about his (counterfactual) full-information preference. That is, the policymaker actually knows what the decision-maker would have done had he been in possession of all information. In such cases, policymakers can unambiguously resolve the problem of intervention by just implementing the decision-maker's fully informed preference for him. Mill highlighted that this behavior complies with notions of classical liberalism, since the intervention is based on what the decision-maker actually wants. How do policymakers react to this information?

We experimentally provide information about another Chooser's fully informed preferences to CAs. This approach also helps us disentangle two causal models of the intensive margin: CAs' own preferences may directly cause the intensive margin (e.g., Ambuehl, Bernheim, and Ockenfels, 2021), or bias beliefs and in turn determine the intensive margin. Our treatment shocks CAs' beliefs: CAs know *with certainty* what option the Chooser would have preferred had he been fully informed. We find that CAs significantly "help" the Chooser by imposing the hypothetically preferred option. But this result is to be qualified: first, the null hypothesis that the CA's own preference and the Chooser's type matter to a similar degree cannot be rejected. Second, we establish a novel finding that CAs do not simply intervene in the direction of their own preference:

Exploratory analyses reveal that a riskless option is causally more likely to be implemented *even when controlling for the CA's own preference*. When both the CA and the Chooser prefer the riskless option, the CA imposes it in 79.6% of cases. Yet when both prefer the lottery, the CA imposes the lottery in only 52.2% of cases. A similar pattern emerges by correlation for the baseline experiment, in which CAs had to rely on their beliefs about Choosers. CAs are less likely to match Choosers with the risky option than the safe option.

This novel result suggests that CAs are able to conceive a "cosmically" ideal intervention that may not always coincide with their personal preference, providing an important qualification to what has been termed "projective paternalism" (Ambuehl, Bernheim, and Ockenfels, 2021). As we discuss below, our result of a cosmic ideal in intervention may be driven by interpersonal regret: in expected value terms, the lottery used in our experiment is just barely better than the safe option.[4] This design feature allows us to identify cosmic ideals for the first time.

There is a small, significant association between beliefs and what option is imposed. This association is independent of CAs' own preference (although that preference robustly biases beliefs). We construct a random-utility binary choice model of our formal theory, and estimate its parameters from data. This model organizes our results by revealing that CAs consistently attach a weight

[4]The expected value of the lottery is €16, in contrast to a safe option of €15.

of about one-third to their own preference. Consistent with our results, this implies that CAs actually care a lot about what the Chooser would do if he were able to choose informedly. This is true for both the baseline experiment (where we utilized incentivized beliefs) as well as for the treatment in which CAs actually knew the hypothetical choice of a Chooser (where beliefs were exogenously provided). The binary choice model of our formal theory performs well in predicting the intensive margin despite its parsimony; indeed, it performs about as well as full-fledged linear probability models and logistic regressions.

This paper contributes to the intersection of public policy, political economy and experimental economics. The importance of knowledge in decision-making is universally recognized in economics. Pareto taught in his *Manual* that only repeated choices reveal a person's preferences (Pareto, 2014, p. 72). Hayek (1945) highlighted that markets—unlike central planners—can efficiently aggregate the dispersed knowledge of decision-makers using prices as signals. Bernheim (2016) distinguished between direct judgments (underlying preferences) and indirect judgments (e.g., choices). Both the former and, if *correctly informed*, the latter should not be questioned by the analyst. In the field of public policy, Musgrave (1956) introduced the notion of merit goods such as education and healthcare. Policymakers are assumed to possess more knowledge about these goods, justifying their provision to an uninformed or myopic citizenry (Head, 1966; Kirchgässner, 2017; Musgrave, 1959). The same intuition holds for the examples in the beginning of this introduction: less knowledge means more intervention.

We add to a nascent experimental literature on paternalism (e.g., Chapters 1, 3; Ackfeld and Ockenfels, 2021; Ambuehl, Bernheim, and Ockenfels, 2021; Bartling et al., 2023; Buckle and Luhan, 2023; Kiessling et al., 2021) that investigates foundations of public policy and political economy. We, too, focus on actions whose outcomes affect only the Chooser, not a third party. Much paternalism (Dworkin, 1972, 2020) is motivated by a lack of knowledge on the side of decision-makers. Recent empirical work on paternalism has emphasized that even if the consequences of some action are precisely defined, CAs intervene. For example, a CA might remove impatient options from a choice menu to enforce a minimum of patience in intertemporal choice. CAs tend to intervene in the direction of their own preference while leaving some space for Choosers to express themselves (Ambuehl, Bernheim, and Ockenfels, 2021, Chapter 1).

As we take no stance on the normative assessment of governmental policy, our approach is purely descriptive or positive (Friedman, 1953).

Our finding of cosmic ideals modifies and qualifies our understanding of "projective paternalism" as introduced by Ambuehl, Bernheim, and Ockenfels (2021). Their conceptualization of paternalism as projection suggested two anchors that shape CAs' interventions: *(i)* CAs impose in the direction of their own preference while *(ii)* leaving some space so that Choosers can partially express their preferences. Because our experiment features a relatively unattractive lottery in a binary choice between that lottery and a safe option, we can identify a third anchor that appears to influence CAs: *(iii)* some options may simply be more "objectively correct" independent of CAs' own preferences. This implies that the projection embedded in paternalistic action is somewhat asymmetric.

A study related to our own is by Bartling et al. (2023). They conduct a study of paternalism in the United States, and vary the feature of the choice ecology through which a Chooser makes a mistake. They show that few CAs restrict freedom of choice, but that a substantial share of CAs provides information to Choosers. However, their design only allowed information provision or intervention as substitutes, not the joint use of both tools. We replicate their approach and their results in cases where CAs can both intervene *and* provide information. We characterize the nature of non-providers theoretically and empirically, and we highlight the possibility for a strategic use of information provision if CAs are not neutral.

Another related study is by Buckle and Luhan (2023), where CAs impose their risk preference despite knowing the Chooser's preferences. Our results also reveal that both CA and Chooser preferences matter. CAs mix their own preference distinctly with that of the Chooser. We add a recognition that some options are more likely to be implemented beyond CA preferences. We argue that these cosmic ideals represent uncontroversially imposed options—perhaps those without the possibility of regret—that are correlated with, but conceptually distinct from, CA preferences.

We contribute to the literature by investigating central motives of experimental policymakers, with increased Chooser knowledge leading to more respect for autonomy. This paper demonstrates that freedom of choice is strikingly contingent on knowledge. This core result has important implications for the design of norms and institutions, and sheds light on central behavioral foundations of political economy. CA behavior is nuanced yet systematic and partially strategic.

Interventions go beyond CAs' preferences, hinting at the existence of cosmic ideals that are just more likely to be imposed. Most CAs provide Choosers with crucial information, but non-providers can be empirically and theoretically characterized according to their own preferences, actions, and beliefs. CAs' knowledge also matters. Our findings suggest that knowledge fundamentally shapes the social contract between them and policymakers. Informational asymmetries are resolved through deliberate institutional responses. Knowledge is a decisive determinant of the boundary between individual autonomy and control by policymakers and others.

The remainder of this paper proceeds as follows: in Section 2.2, we formulate a theory of optimal self-interested paternalism. This theory is informed by other authors' normative and positive conceptions of paternalism. We use these insights to design and discuss experimental investigations of the relationship between knowledge and freedom. In Section 2.3, we present the experimental designs used in our two experiments. Section 2.4 discusses our results. Finally, we conclude.

2.2 Theoretical framework

In this paper, we have a CA govern over a Chooser: the CA decides whether to have the Chooser have his choice between two options, or to impose an option on him. This Section builds a formal model of optimal but possibly self-interested paternalism.

Our framework is most closely related to the models by Manski and Sheshinski (2023) and Bartling et al. (2023). In contrast to these approaches—and in line with Ambuehl, Bernheim, and Ockenfels (2021)—we allow a CA to include her own preference in a measure of welfare. We investigate the implications for interventions, choices made in place of the Chooser and information provision. The Section subsequently informs model parameters with expectations from the literature in order to sketch hypotheses for experimental research.

2.2.1 A formal model of paternalism with and without mistakes

Consider a choice between two options: x and y. CA and Chooser are two agents endowed with a strict preference over x, y. While the CA does not make an interpersonal utility comparison (Binmore, 2009; Hausman, 1995; Kolm, 1993),

	CA prefers x	CA prefers y
Chooser prefers x	$W(x) = 1$ $W(y) = 0$	$W(x) = 1 - \phi$ $W(y) = \phi$
Chooser prefers y	$W(x) = \phi$ $W(y) = 1 - \phi$	$W(x) = 0$ $W(y) = 1$

Table 2.1: W, given the Chooser's and CA's possible types

she does recognize the possibility of the Chooser's preferences disagreeing with her own. The CA's utility function is $U : \{x, y\} \rightarrow \{0, 1\}$, and the Chooser's utility function—as perceived by the CA—is $V : \{x, y\} \rightarrow \{0, 1\}$.

The CA aggregates these binary utilities using a welfare function, $W : \{x, y, \{x, y\}\} \rightarrow [0, 1]$. The argument to W indicates the menu of choices available to the Chooser. She has to decide whether to impose one of these options or to let the Chooser have his choice. In this model, freedom is instrumental; i.e., it is merely a means to achieve an end. For $z \in \{x, y\}$, W is defined as follows:

$$W(z) = \phi U(z) + (1 - \phi) V(z), \tag{2.1}$$

where $\phi \in [0, 1]$ is the weight placed by the CA on her own preference. In this model, self-interest may arise from whatever connection a CA feels to the Chooser's choice. ϕ may thus be a reflection of CAs' social preferences (e.g., Fehr and Charness, 2024), paternalistic projection (Ambuehl, Bernheim, and Ockenfels, 2021), a means of enforcing norm compliance (e.g., Traxler and Winter, 2012) or an attempt to change the norm by bringing others' behavior in line with the CA's own (Khalmetski and Ockenfels, 2024). While our experiments do not create any explicit link between CAs and the Chooser's decision,[5] a connection may still be present because of prominent behavioral phenomena.

The value of $W(\{x, y\})$ depends on how Choosers actually choose; we define it in the following Sections based on conditional expectations of Equation 2.1. In contrast to the usual approach in welfare economics, policymakers may be self-interested. As we demonstrate in Section 2.2.5, John Stuart Mill implied that ϕ ought to be nil. In that case, the CA assumes the Chooser's preference. However, as shown empirically (e.g., Ambuehl, Bernheim, and Ockenfels, 2021), interventions are correlated with CAs' own preferences. In that case, ϕ may be

[5]Most importantly, CAs' payment is independent of the Chooser's decision.

strictly positive. Table 2.1 gives all possible values of Equation 2.1 if Chooser preferences are known to the CA exactly. In that case, CAs partially project their own tastes on Choosers.

2.2.1.1 Beliefs

In the following—without loss of generality—we restrict our analysis to an x-preferring CA. A Chooser's type may not be known exactly, or a CA may need to evaluate W for a distribution of types. Let $\tilde{q} \in [0, 1]$ denote the CA's belief about the proportion of Choosers preferring x. By mixing over Chooser types in Table 2.1, we find

$$W_x(x) = \phi + \tilde{q}(1 - \phi), \tag{2.2}$$

$$W_x(y) = (1 - \tilde{q})(1 - \phi), \tag{2.3}$$

with the subscript indicating the CA's type. In Section 2.4.3, we use Equations 2.2–2.3 (and those corresponding to $W_y(x)$ and $W_y(y)$) and incentivized or provided data on beliefs to estimate ϕ.

2.2.1.2 Interventions without Chooser mistakes

Consider first the instrumental value of liberty if Choosers choose *perfectly* according to their type. Assume that Choosers choose the option implied by their true preferences (for critical viewpoints of this concept, see Špecián, 2019; Sugden, 2022). That is, all Choosers with a preference for option x choose option x. Similarly, all Choosers with a preference for option y choose y.

In the absence of mistakes, $\tilde{q}$ in Equations 2.2–2.3 will equal 1 for those who choose x and 0 for those who choose y.[6] Let $\pi \in (0, 1)$ be the proportion of Choosers opting for option x. We define freedom of choice as the Chooser being able to choose from the menu $\{x, y\}$. Then, by mixing over W, for all CA types $\theta \in \{x, y\}$,

$$W_\theta\left(\{x, y\}\right) = \pi W_\theta\left(x \mid \tilde{q} = 1\right) + (1 - \pi)\, W_\theta\left(y \mid \tilde{q} = 0\right). \tag{2.4}$$

It is straightforward to compute that $W_x\left(\{x, y\}\right) = 1 - \phi(1 - \pi)$. The CA will compare this value of Equation 2.4 against Equations 2.2–2.3, with $\tilde{q} \equiv \pi$.

[6] Recall how $\widetilde{q}$ is defined as a belief about preferences, not choices. However, if choices are noiseless, both concepts coincide.

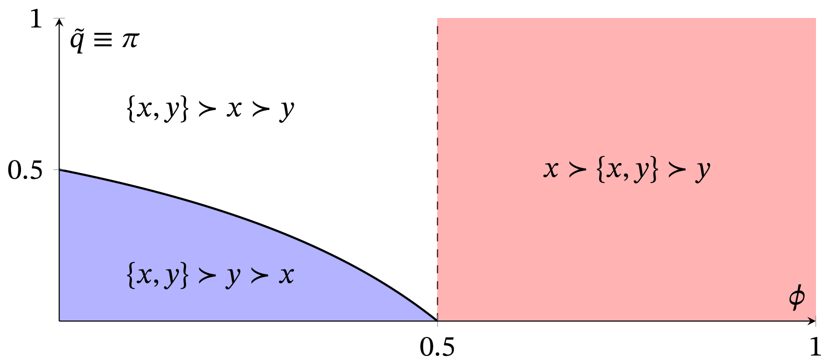


Figure 2.1: Optimal CA decisions

This ternary comparison of welfares determines the CA's governance over the Chooser. In this model, the comparison between Equations 2.2–2.3 (i.e., what to impose) is independent of the comparison between Equations 2.2–2.3 and Equation 2.4 (i.e., whether to intervene, and if so, how). This *independence of irrelevant alternatives* arises because Equations 2.2–2.3 are independent of actual choice proportions—they refer only to beliefs about true preferences. This finding is important for empirical analysis. It suggests that we can separately analyze the extensive and intensive margins of paternalistic intervention.

Let us reconsider the relationship between Equations 2.2–2.4 if Choosers choose correctly. It follows from Equations 2.2–2.3 that an x-preferring CA will never impose y. This is because $W_x(\{x, y\}) - W_x(y \mid \tilde{q} = \pi) = \pi > 0$. (Similarly, a y-preferring CA will not impose x.)

Proposition 1. Without Chooser mistakes, a CA will (weakly) impose her own preference if and only if $\phi \geq 1/2$. Otherwise, she does not intervene at all.

Proof. Use Equations 2.2 and 2.4, setting $\tilde{q} = \pi$, and solve for ϕ. □

In other words, if CAs weigh their own preference heavily enough, CAs implement it. This reproduces the core empirical result of Ambuehl, Bernheim, and Ockenfels (2021) in a simpler model.

Under choice without mistakes, CAs cannot impose their non-preferred option. However, it is not true that x always prevails over y; rather, freedom of choice prevails over y. Figure 2.1 reveals why. For an x-preferring CA, there are two corridors at which the ranking of optimal governance changes. At very low $\tilde{q}$, y would be chosen over x if the choice were only between x and y, not x,

y and $\{x, y\}$. However, at the points where y could be imposed, ϕ is too low to cause an intervention.

Moreover, note how the non-utilitarian nature of the model produces a majoritarian standard in the case of $\phi = 0$. If a CA were forced to intervene, theory predicts a clear trade-off between ϕ and $\tilde{q}$, with the more popular option being imposed at $\phi = 0$. Section 2.2.5.3 discusses this majoritarian standard further. Both $W_x(\{x, y\})$ and $W_x(x)$ increase in $\tilde{q}$ for all ϕ.

2.2.1.3 Mistakes and intervention

We now relax the assumption that Choosers choose correctly. In a choice between two options, mistakes can be incorporated easily. Let $\epsilon_x, \epsilon_y \in [0, 1]$. We say that a proportion of $\pi' = \pi - \epsilon_x + \epsilon_y$ Choosers chooses x. Consider the following contingency table for Chooser preferences and decisions:

	Choose x	Choose y	**Sum**
Prefer x	$\pi - \epsilon_x$	ϵ_x	π
Prefer y	ϵ_y	$1 - \pi - \epsilon_y$	$1 - \pi$
Sum	π'	$1 - \pi'$	1

An absence of mistakes implies $\epsilon_x = \epsilon_y = 0$. We now allow arbitrary off-diagonal elements. Out of π', only a proportion of $(\pi - \epsilon)/\pi'$ Choosers actually prefers x. Similarly, a proportion of $\epsilon_x/(1 - \pi')$ Choosers prefers x (but they choose y). Equations 2.2–2.3 are unchanged, as they only refer to $\tilde{q}$, i.e., beliefs about true preferences. However, Equation 2.4 must be adjusted to account for a change in conditional proportions.

$$W_\theta'(\{x, y\}) = \pi' W_\theta\left(x \,\middle|\, \tilde{q} = \frac{\pi - \epsilon_x}{\pi'}\right) + (1 - \pi') W_\theta\left(y \,\middle|\, \tilde{q} = \frac{\epsilon_x}{1 - \pi'}\right). \tag{2.5}$$

The following Theorem establishes the welfare attained under freedom of choice.

Theorem 1. *Freedom with Chooser mistakes.*

$$W_x'(\{x, y\}) = 1 - \phi\left[1 - \pi - 2\epsilon_y\right] - \epsilon_x - \epsilon_y. \tag{2.6}$$

Proof. Use Equations 2.2–2.3 in Equation 2.5. □

Let us briefly reflect on this Theorem. Recall that the error implied by ϵ_x is worse, to an x-preferring CA, than that implied by ϵ_y. In the case of ϵ_x, Choosers who would have preferred x erroneously choose y, the option not preferred by the CA. On the other hand, ϵ_y reflects a proportion of Choosers who now choose x (although their true preferences are better reflected by y). This highlights that freedom of choice is purely instrumental in this model. $\partial W_x' (\{x, y\})/\partial \epsilon_x = -1$, implying that an increase in the error disfavored by the CA always leads to an absolute decrease in the welfare accorded to liberty. However, $\partial W_x' (\{x, y\})/\partial \epsilon_y = -1 + 2\phi$. Not only is the negative effect on welfare attenuated by an increase in the error favored by the CA, but if x is heavily weighed by the CA, more errors in x's direction actually enhance the attractiveness of liberty.

Recall that the ranking of x and y on the intensive margin is unaffected by mistakes. This implies that we can concern ourselves exclusively with the effect of variations to Choosers' choice ecology on the extensive margin. The following Corollary makes one statement about such an effect.

Corollary 1. *Chooser mistakes make freedom less attractive for sufficiently neutral CAs.* Suppose a CA does not intervene where Choosers choose without the possibility of mistakes. She may intervene where Choosers have the possibility of making mistakes.

Proof. If the CA does not intervene if Choosers cannot make mistakes, this reveals her $\phi \leq 1/2$ (Proposition 1). Consider the following object:

$$\begin{aligned} \Delta &= W_x' (\{x, y\}) - W_x (\{x, y\}) \\ &= -\epsilon_x - \epsilon_y [1 - 2\phi] . \end{aligned}$$

Since $\phi \leq 1/2$ and $\epsilon_x, \epsilon_y \geq 0$, $\Delta \leq 0$. □

This reduction in the welfare perceived by the CA is nonzero if $\pi \neq \pi'$. It may cause $\{x, y\}$ to become less desirable than the imposition of x or y.

Chooser mistakes can cause a "peeling back" of the dashed frontier in Figure 2.1. If the area of intervention is increased, Choosers may face the imposition of x or y. Above, we showed that if no Chooser mistakes are possible, the intensive margin is restricted to CAs' own type. This result does not obtain here.

In some instances, CAs can choose between $W_x' (\{x, y\})$ and $W_x (\{x, y\})$. For example, Corollary 1 suggests that if information can be costlessly provided to Choosers to help them decide without mistakes, CAs can raise welfare by doing so if $\phi \leq 1/2$. Corollary 1 thus also reproduces Mill's idea that providing information to Choosers is preferable over imposition (Section 2.2.4). Moreover, the Corollary implies that some CAs may consciously choose to have Choosers make mistakes by having him decide uninformedly—because those mistakes benefit the CA.

2.2.2 Mill on paternalism

We draw on the work of English economist[7] and philosopher John Stuart Mill (1806–1873) to inform economic theory and to create predictions for positive research (Friedman, 1953) on paternalism. Mill's "On Liberty" (1869) emphasizes that *(i)* people are, in general, best left free to pursue their lives unless *(ii)* they have insufficient knowledge about the consequences of their actions. In this case, interventions must be based on a counterfactual assessment of true preferences.

2.2.3 Extensive margin: The "if" of interventions

Freedom of choice allows a diverse group of individuals to satisfy every extant preference (Konrad, 2024). In a society with wide-ranging preferences and viewpoints, liberty will increase satisfaction compared with imposition (Mill, 1869, ch. 1).

Nonetheless, Mill provides some exceptions to this general rule. Of the five identified by Mabsout (2022), arguably the most relevant for real-life policy is harm to others. We focus instead on a lack of knowledge. At least two authors (Arneson, 1980; Scoccia, 2018) have attempted to modify the definition of paternalism so that interventions motivated by a lack of knowledge on the side of decision-makers are only paternalistic if they conflict with their full-information preference (Mabsout, 2022, fn. 6). We discuss below the relevance of this counterfactual assessment in Mill.[8] Nonetheless, as hinted at in the introduction, it

[7]See Cowen (2023, ch. 6).

[8]At this point, let us offer two arguments in response to Arneson and Scoccia. First, real-life policy governs over broad swathes of the population that is differentially informed. Some of these Choosers may possess all relevant information and still decide to act in a specific way. To that extent, the policy regains a paternalistic character. Second, this argument begs the question of what relevant knowledge is. Policymakers must unavoidably make subjective judgments about choice situations and whether they feel able to improve upon them. Whether decision-makers have all objectively required knowledge is not a helpful benchmark because what information

is important to note that one does not need to accept Dworkin's definition of paternalism to recognize the profound role that Chooser knowledge plays in policy.

Mill viewed it acceptable to never intervene. However, he gives the following well-known example:

> If either a public officer or any one else saw a person attempting to cross a bridge which had been ascertained to be unsafe, and there were no time to warn him of his danger, they might seize him and turn him back without any real infringement of his liberty; for liberty consists in doing what one desires, and he does not desire to fall into the river. *(Mill, 1869, 172f.)*

Here, a CA is neutral and reacts only to informational disadvantages: if the pedestrian knew what he was doing, it would not be justifiable to intervene. The pedestrian lacks knowledge and warning him is not possible, *ipso facto* intervention is justified. In the language of our formal model, the pedestrian makes a mistake: there is a discrepancy between preferences and choices (Section 2.2.1.3).

Below, we experimentally vary the amount of knowledge available to Choosers in a choice between two options. This tests the fundamental conditionality of liberty on knowledge that Mill so eloquently put forward: in a situation where knowledge is fixed and cannot be provided, do more interventions take place if the Chooser is less informed? This aspect of our research deals with the *extensive margin* of paternalism: whether or not interventions take place. As demonstrated above, theory provides a justification for the separation of the extensive and intensive margins.

2.2.4 Information provision as a substitute to intervention

Mill argued that knowledge should be provided instead of intervening, but if that is not possible, it is legitimate to intervene even if the harms accrue only to the decision-maker. Corollary 1 explains why: the appeal of liberty is reduced if decision-maker mistakes are possible. (That is, if the CA is Millian and does not weigh her own preference too heavily.)

is useful is merely another subjective judgment, one that has often given rise to abuse (Berlin, 1958). In acting upon this subjective assessment, policymakers unavoidably accept to place their judgment upon others; they prioritize their own understanding of a situation.

It has long been recognized in economics (Blackwell, 1953) that an expected-utility decision-maker can better his position by relying on more accurate information. Indeed, Bartling et al. (2023) found experimentally that a vast majority of CAs transmit true information to Choosers. However, information can play a strategic role for self-interested policymakers. First, if (a lack of) information leads to Chooser mistakes in the CA's subjectively preferred direction, information may be withheld from Choosers. Corollary 1 illustrates this point formally. This argument hints at why policymakers do not correct the widespread pessimistic misconceptions about health risks to smoking (Viscusi, 1990): these misconceptions go in policymakers' subjectively preferred direction (fewer people smoke). By not correcting these misconceptions, policymakers engage in "paternalism by omission."

Second, if it is certain that the Chooser obtains his choice after information is provided, the CA may help to effectuate the Chooser's preferences by providing information. In this sense, the provision may activate a utilitarian system in CAs, where they may disagree with what the Chooser does, but where they keep him informed so that he may at least achieve his ends. However, if an opportunity to intervene presents itself to the CA, any prior provision of information may be used "behaviorally." As an example, consider that Choosers may be grateful for intervention if they are uninformed. If a CA suspects that this may be the case, she can launder her intervention by keeping the Chooser uninformed, hiding that she controlled the information received. Far from enabling a utilitarian system in CAs, the possibility of a follow-on intervention changes the first-stage calculus of providing information. We test this idea below: in the baseline, CAs can only provide information (or not), similar to Bartling et al. (2023). In a treatment, they may simultaneously intervene. From a theoretical perspective, the treatment tests the independence of irrelevant alternatives: in the baseline, CAs choose between $W_\theta(\{x, y\})$ and $W'_\theta(\{x, y\})$. In the treatment, $W_\theta(x)$ and $W_\theta(y)$ are additionally available (both with and without information provided to Choosers).

2.2.5 Intensive margin: The "how" of interventions

Mill makes a normative statement about the intensive margin should an intervention ever occur. It relates to what is still supposedly intended by many interventions: the implementation of true preferences. Mill makes an inference on the hypothetical preference of the "person" *viz.* Chooser. The intervener is to implement the Chooser's counterfactual full-information preference, not her own personal view. If the pedestrian's goal was to get wet, intervention would not have been legitimate (Arneson, 1980). This implies that $\phi = 0$ in Equations 2.2–2.3, which in turn requires CAs to implement the Chooser's true preference if it is known.

If a CA is not informed about the Chooser's hypothetical full-information preference, what will they rely upon to form an intervention? Equations 2.2–2.3 suggest that beliefs about counterfactual choice proportions ought to be used, and perhaps a majoritarian standard. Moreover, we test how CAs use exact information on a Chooser's type. Let us now distinguish a number of positive and normative theories on how to intervene.

2.2.5.1 Utilitarianism

In the absence of hypothetical full-information choices, no information about utilities is available. Furthermore, no method for eliciting a prediction of utilities is known. Thus, this theory cannot be tested. It is not necessary at this point to delve into the manifold issues with utilitarianism and utilitarian calculation in general (e.g., Kolm, 1993).

2.2.5.2 Projective paternalism

Previous studies on paternalism and intervention (e.g., Chapter 1; Ambuehl, Bernheim, and Ockenfels, 2021) have emphasized a profound tendency for CAs to impose "in the direction" of their own preference as well as a similar bias in beliefs. For example, Chapter 1 shows how CAs that share their data tend to believe that others also want to share their data. False consensus bias has a long history in psychology; see Ross, Greene, and House (1977) for a magisterial exposition. More recently, the idea that individuals project their tastes onto others has garnered renewed attention in experimental economics. For example, Bushong and Gagnon-Bartsch (2024) show that workers in a real-effort task who were more (less) tired believed other workers to be less (more) willing to work.

We can test this theory by correlating the CA's own preference with the intensive margin. Ideally, we would shock the CA's own preference to provide causal estimates. However, we leave that to future work.

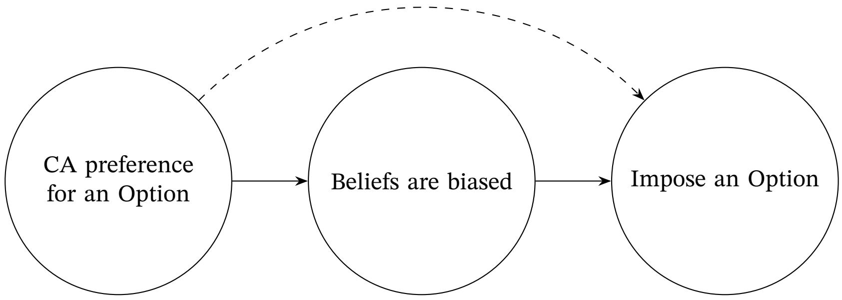


Figure 2.2: Two causal models of interventions based on beliefs

2.2.5.3 Majoritarian counterfactuals

In the absence of information on utilities—as in Section 2.2.1—CAs can rely on their beliefs in selecting the intensive margin. One simple theory is that the CA implements what she believes to be the majority choice under full information. Such an intervention could be justified by reference to the median voter theorem (Black, 1948; Downs, 1957), as this voter would prevail in a simple majority election if preferences are single-peaked. In our theoretical framework, use of a majoritarian standard implies $\phi = 0$.

It is important to note that beliefs may be incorrect (Ambuehl, Bernheim, and Ockenfels, 2021; Ross, Greene, and House, 1977), highlighting the subjective nature of rule-making. For this reason, it is essential to disentangle two potential causes of an intervention on the intensive margin (Figure 2.2). First, a CA's own preference could lead a majoritarian CA to believe that an Option is more popular at full information than it really is, leading to a systematic but unintended imposition of the subjectively preferred Option. Second, a CA's own preference could *directly* cause the imposition of the preferred Option, without Mill's normative inference.

We interpret the indirect path as purely *statistical or majoritarian intervention* and the direct path as *taste-based intervention*. In Becker's classic discussion (1957), similar language is used to distinguish two fundamental motives of discrimination. In the modern reading of this perspective, beliefs can be biased (e.g., Bohren et al., 2023), but statistical discrimination relies on perceived differences in productivity, whereas taste-based discrimination is concerned with non-objectifiable personal decider biases. The same fundamental distinction applies here. As Figure 2.2 demonstrates, CA tastes can enter interventions on the intensive margin on two paths. Majoritarian intervention demands that the intensive margin is selected from beliefs about the most popular full-information counterfactual choice, i.e., without the dashed path. On the other hand, real CAs may intervene on taste-based grounds. The tendency to—inadvertently—impose one's own preference must be accounted for in the analysis by including both beliefs and the CA's own preference. Otherwise, the estimates for the weight of the CA's own preference may be biased upward. Our formal model makes precise statements about the trade-off between ϕ and π which we use in estimation (Section 2.4.3).

2.2.5.4 Individual counterfactuals: Testing Mill's postulate

Mill's idea of implementing what the Chooser *would have done given full information* can be tested directly. CAs can be informed about the Chooser's type. Ambuehl, Bernheim, and Ockenfels (2021) revealed the self-image of some Choosers to CAs, which significantly shifted the intensive margin of interventions. More recently, Buckle and Luhan (2023) conducted an experiment in which decisions in an investment game could be overridden by financial advisors. These authors found that advisors use both their own preferences and those of the investor to decide on the intensive margin.

In the context of paternalism, focusing on a single individual also resolves the issues of utilitarian calculation mentioned above. Beliefs about true preferences are shocked to $\pi = 0$ or $\pi = 1$. If CAs are provided with information about what the Chooser would have done had they possessed full information, Mill's postulate about the intensive margin—that $\phi = 0$—can be separately tested.

2.2.6 Aggregating degrees of knowledge

A related literature on positive welfare economics has investigated how CAs can work to aggregate the preferences of a group's members. In Ambuehl and Bernheim (2024), group members have to work on a task. Each member possesses an individual preference rankings about the available tasks; an experimental social planner is informed about these rankings to assign tasks to each member of the group. Separately, these authors study how planners direct donations to Swiss political parties based on similar rankings.

Both preferences and degrees of knowledge can vary widely in society. Below, we study a related case: how degrees of knowledge are aggregated when a single person can either be well informed or not informed at all. A CA concerned with a Chooser who is in one of several states has to weigh the relative importance of each state. We conduct trials to explain interventions for this scenario using previous interventions in which the amount of information provided to the Chooser was known to CAs in advance.

2.3 Experimental design

In our experiments we vary the amount of information obtained by a Chooser in a binary choice and let CAs govern for the Chooser. CAs face no monetary incentives whatsoever from the Chooser's actual choice.

2.3.1 Operationalizing partial knowledge

Consider the following binary choice faced by a Chooser:

Option 1		**Option 2**	
Prob.	Outcome	Prob.	Outcome
1	z	p	y'
		$1-p$	y''

with $y' \leq z \leq y''$. Option 1 is a certain amount of money, while Option 2 is a simple lottery. The degree of knowledge that Choosers obtain about Option 2 can be varied. We consider only an ambiguity variant in which knowledge about p is varied, but not information about z, y', y''.

In our experiments, we will consider three cases: *(i)* the Chooser will know the value of p exactly (there is no ambiguity); *(ii)* the Chooser will observe exactly k draws from Option 2 (knowing a prior for p); or *(iii)* the Chooser obtains no information whatsoever about p, not even the prior. In case *(i)*, the Chooser is fully informed, while in *(iii)* he decides in a scenario of sheer ignorance. Case *(ii)* is an intermediate position. Experiment 1 will focus on the comparison between cases *(i)* and *(ii)*, while experiment 2 uses the contrast between cases *(i)* and *(iii)*. In all experiments, given this information structure, the Chooser decides between the Options under rules constructed by a CA.

2.3.2 Ambiguity in the Estimation Game

Option 2 can be made partially ambiguous by having the Chooser observe k draws from Option 2 before deciding. This is what we call the "Estimation Game."

This Game lends itself well to experiments with asymmetric information. Consider the following setup: a computer draws p in Option 2 from a distribution. The decision-maker (Chooser) enters the experiment. The Chooser obtains k draws from Option 2 and the prior. $n \leq k$ draws show y'. A CA may be matched to the Chooser; this CA knows the exact value of p. CAs may intervene in the Choosers' choice between Options 1 and 2 (or let the Chooser have his choice). From an economic viewpoint, Choosers estimate the utility of Option 2 to decide whether to choose it over Option 1 (see Gigliotti, 1996, for an early example of partial ambiguity). From a statistical viewpoint, the Chooser attempts to estimate p. By having CAs decide for varying values of k, we can disentangle the effect of Chooser knowledge on their autonomy. That is the setup of experiment 1 below (Section 2.3.4).

Clearly, $n \sim \text{Binomial}(k, p)$. For any n, k, and the uniform distribution from which p is drawn, the Bayesian posterior for p is $\text{Beta}(n+1, k-n+1)$. Marginalizing over n, we obtain a marginal posterior of p with the following density function:

$$f_k(x) = \sum_{n=0}^{k} \frac{x^n(1-x)^{k-n}}{B(n+1, k-n+1)} \binom{k}{n} p^n(1-p)^{k-n} \tag{2.7}$$

$$= (k+1)(1-p)^k(1-x)^k \, {}_2F_1\left(-k, -k, 1, \frac{px}{(p-1)(x-1)}\right), \tag{2.8}$$

where ${}_2F_1$ is the hypergeometric function. This is the distribution an expected-utility CA would expect an expected-utility Chooser to work with if the CA knows p and the Chooser is to obtain k draws from Option 2, but the precise n observed by the Chooser is not yet known.

Suppose that CAs decide for the case of $p = 0.2$, $z = €15$, $y' = €0$ and $y'' = €20$ (as below). Table B.1 in the Appendix presents summary statistics of the marginal posterior for $p = 0.2$. All of these statistics converge in k, i.e., a higher k carries superior information. These measures are important beyond Blackwell's (1953) order, as they highlight the value of an increased k even for non-expected-utility Choosers. As k increases, Choosers' inference on p is robustly improved because a higher k is monotonically more informative. The

Estimation Game is similar to the well-known “balls and urns” paradigm, but the samples from Option 2 are drawn with replacement.

2.3.3 Choice Architects’ information structure

In all experiments, CAs obtain information about the Chooser’s decision scenario. They know z, y', y'', p and they know what Choosers know: z, y', y'', k and that n will be drawn once from Binomial(k, p). CAs may then impose one of the Options or have the Chooser have his choice. All experiments were free of deception.

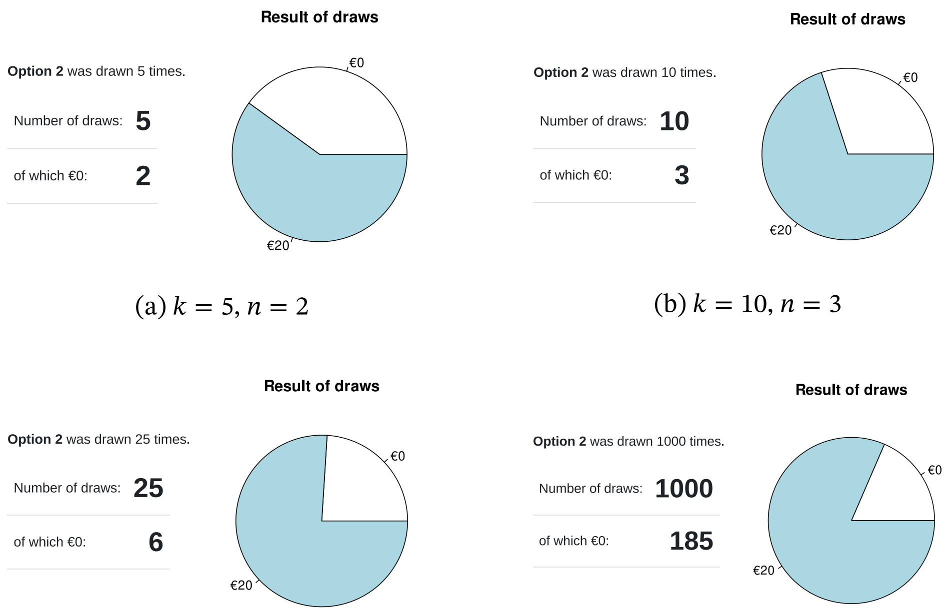
In experiment 1, CAs know that their decision for the Chooser can only be implemented if p takes on the value 0.2.[9] In experiment 2, CAs know that p will certainly take the value $p = 0.2$ for Choosers. Their decision can only be implemented if they are randomly selected.

2.3.4 Experiment 1

Experiment 1 investigates the effect of Choosers’ partial ambiguity (Section 2.3.2) on the freedom they are granted.

[9]Instead of a continuous distribution for p, we used a discrete uniform distribution from 0% to 100%, inclusive, in 1% increments, to draw from p.

2.3.4.1 The Choice Architect's view



(a) $k = 5, n = 2$ (b) $k = 10, n = 3$

(c) $k = 25, n = 6$ (d) $k = 1000, n = 185$

Figure 2.3: Examples of how the Chooser's view was presented to CAs, translated to English

CAs were able to simulate sample draws from Option 2; these draws are visualized using pie charts, similar to the approach in Harrison et al. (2018, fig. 1). n was drawn from the corresponding binomial distribution, given k and p. See Figure 2.3 for examples. For any fixed k, a new draw was made about every 2 seconds. This allowed CAs to obtain a thorough understanding of the distribution of n to dispel incorrect beliefs, such as in the "law of small numbers" (e.g., Tversky and Kahneman, 1971).[10]

Note that CAs do not know the actual value of n shown to Choosers. As in the discussion on the marginal posterior in Section 2.3.2, we decided to keep n random, as we believed any other design to be difficult to implement without deceiving either CAs or Choosers. Moreover, keeping n random while fixing k reduces statistical noise arising from differential intervention given the many

[10]As our results indicate, subjects made much use of the animations. Only 12% of CAs never watched an animation; the median number of animations shown to a CA was 24.5.

possible values of n. In sum, CAs only learn the distribution of n for all possible k.

2.3.4.2 Treatments

CAs were informed that p will originally be drawn from a discrete uniform distribution that includes the end points 0% and 100%, but that their decision can only be implemented if p will randomly take the value 0.2 (20%). Hence, the probability of implementation was low. Each CA was presented—in random order (Charness, Gneezy, and Kuhn, 2012)—with the following values of k: 0, 1, 2, 5, 10, 25, 50, 1000 and ∞. $k = \infty$ represents the condition that the Chooser learns the value of p exactly, as in case *(i)* of Section 2.3.1. Since all CAs participated in all of these treatments, our experiment has *within* characteristics.

2.3.4.3 Procedures

The experiment was conducted at the Cologne Laboratory for Economic Research in Germany in March 2023. Recruitment was done using ORSEE (Greiner, 2015). All participants identified as students. Participants were not selected based on major or any demographic variable. English-language instructions are available in Appendix B.5.1. IRB approval was granted on January 30, 2023 by the WiSo Ethics Review Board at the University of Cologne (Reference 230005MG). The experiment was preregistered at AsPredicted.[11]

Subjects were invited to participate in the online experiment at a date of their choosing. On that date, they were free to start the experiment at any point between 2pm and 6pm. Subjects were only allowed to intervene after passing a comprehensive set of comprehension checks. While we did not restrict the number of attempts, we offered a bonus for passing the comprehension checks on one's first attempt. A number of participants were unable to pass the comprehension checks during the time limit allocated to the page, and some participants withdrew at any point. All in all, 368 CAs and 2 Choosers participated. On average, CAs earned €5.14 and spent 22 minutes in the experiment. 301 CAs are "complete" as defined by the preregistration (they participated in all parts of the experiment).

[11]The preregistration can be viewed at `https://aspredicted.org/Y68_8JW`, accessed January 2, 2025.

Chooser	Block	Information	Type	Reference
1	I	None	Unknown to CA	Section 2.3.1, case *(iii)*
2		Full	Unknown to CA	Section 2.3.1, case *(i)*
3	II	Uncertain	Unknown to CA	Section 2.2.6
4		Decided by CA	Unknown to CA	Section 2.2.4
5		None	Known to CA	Section 2.2.5.4

Table 2.2: Choosers in experiment 2

As prescribed by the preregistration, only the data of the first 300 CAs is used in the analysis. No further exclusions are applied.

2.3.5 Experiment 2

Experiment 2 exploits cases *(i)* and *(iii)* of Section 2.3.1 to investigate further aspects of the relationship between paternalism and knowledge. Experiment 2 is a follow-up to experiment 1, designed after learning the results of experiment 1. As in experiment 1, CAs are allowed but not required to intervene in the decision faced by Choosers.

2.3.5.1 Chooser scenarios and treatments

In experiment 2, CAs made decisions for a total of five Choosers in two blocks (see Table 2.2). The order of Choosers within blocks was randomized, but block II always followed block I. Block I attempted to replicate the findings of experiment 1 in a setting where Choosers did not receive information about the prior for p (p is not random, but indeed fixed at 0.2).

Block I Chooser 1 obtained no information (not even a prior), as in case *(iii)* of Section 2.3.1.[12] As in case *(i)*, Chooser 2 received full information.

Block II CAs are told that Chooser 3 can either be uninformed or fully informed. Either of these states of nature can occur with 50% probability, as in Section 2.2.6. At the time of CA rule-making, it is yet unknown which state is the true one.

[12]Note how in Screen 9 of Section B.5.2 in the Appendix, the value of p is essentially blacked out. The value of p was not available through any other means, including navigating to the page's source code.

Chooser 4's degree of knowledge is determined by the CA (Section 2.2.4). For this Chooser, CAs were randomly allocated to a baseline or the treatment Plus. Our treatment relates to the institutional setup of information provision for Chooser 4: In the baseline, CAs were only allowed to provide information to Choosers, as in Bartling et al. (2023). In Plus, they were enabled to intervene in the resulting Choice in addition to providing information. Simply put, in both treatments the CA can choose between cases *(i)* and *(iii)* of Section 2.3.1; in Plus they may also add an intervention in the resulting choice. On the other hand, in the baseline, it is a given that the Chooser's own choice is implemented after Chooser 4 receives the information decided upon by the CA. Both information provision and—in Plus—the intervention for the Chooser took place on the same screen.

Chooser 5 is uninformed, but there is information about his counterfactual choice (Section 2.2.5.4). For this Chooser, CAs were randomly allocated to the Chooser's hypothetical full-information choice: what Option the Chooser preferred in the full-information counterfactual. CAs were told that Chooser 5 decided between the Options for several possible values of p, not knowing the true value. CAs were guaranteed that their decision could only be implemented if Chooser 5 did, in fact, prefer the given Option at $p = 0.2$. Simply put, CAs can "help" Chooser 5 get what he would obtain if he were fully informed.

2.3.5.2 Procedures

Experiment 2 was developed with uproot (Grossmann and Gerhardt, n.d.) and conducted at the Cologne Laboratory for Economic Research with students from Cologne and Maastricht in mid-2024. Participants were not selected based on major or any demographic variable. Recruitment was done using ORSEE (Greiner, 2015). English-language instructions are available in Appendix B.5.2. IRB approval was granted by the Gesellschaft für experimentelle Wirtschaftsforschung e.V. on May 13, 2024 (Approval ID RNnxiot5), a German nonprofit association providing services to experimental economists.[13] The experiment was preregistered at AsPredicted.[14]

[13] The ethics certificate can be viewed at `https://gfew.de/ethik/RNnxiot5`, accessed January 2, 2025.

[14] The preregistration can be viewed at `https://aspredicted.org/625_QRC`, accessed January 2, 2025.

Subjects were invited to participate in the online experiment at a date of their choosing. On that date, they were free to start the experiment at any point between 10am and 6pm.[15] All in all, 610 CAs started the experiment. 603 CAs are "complete" as defined by the preregistration (they participated in all parts of the experiment). On average, these CAs earned €3.60 and spent 10 minutes in the experiment.

As prescribed by the preregistration, only the data of the first 600 CAs is used in the analysis. No further exclusions are applied.

2.3.6 Predictions and research questions

Given the considerations of the value of k in Section 2.3.2, we predict that, for $p = 0.2$, there will be fewer interventions as k is increased. Since—to some extent—experiment 2 is a *conceptual replication* (or "many-designs replication") of experiment 1 (Derksen and Morawski, 2022), we predict that a similar result can be obtained for the extreme cases *(i)* and *(iii)* of Section 2.3.1. The following prediction also follows directly from our formal model of Section 2.2.1.

Prediction 1. *Knowledge and freedom. Experiment 1*: There will be fewer interventions under high k than under low k. *Experiment 2*: There will be fewer interventions under full knowledge (Chooser 2) than under no knowledge (Chooser 1).

Furthermore, we predict that beliefs about behavior at $k = \infty$ (or under full knowledge) will be systematically biased (see Section 2.2.5.3):

Prediction 2. *False consensus bias*. CAs will believe Option 1 to be more popular if they themselves prefer Option 1.

In research question 1, we assess the majoritarian and taste-based extents of interventions on the intensive margin.

Research Question 1. *Intensive margin*. When controlling for the CAs' own preferences, will those CAs who intervene systematically impose the more popular option full information, according to their own beliefs?

Research question 2 relates to Chooser 5.

[15]Students from Maastricht were allowed to participate at any time. They were recruited by email through a lecture.

Research Question 2. *Mill's intensive-margin postulate*. Is knowledge of the Chooser's hypothetical perfect-information choice able to overcome projective paternalism?

Prediction 3 is about the information provided to Chooser 4 by CAs.

Prediction 3. *Knowledge provision*. CAs will provide less information to Chooser 4 when they can both inform and intervene in the Chooser's decision (as in treatment Plus), compared to when CAs can only inform without being able to intervene.

Research question 3 relates to Chooser 3.

Research Question 3. *Aggregating degrees of knowledge*. Which is more predictive of the choice to intervene for Chooser 3: the choice to intervene for Chooser 1 or that for Chooser 2?

2.4 Results

In this Section, we discuss our experiments' results. For both experiments, we follow the preregistrations exactly (except for a minor correction to the preregistration of experiment 2, see Section B.4.4.1 in the Appendix). We reference our preregistered analyses when it comes to the evaluation of predictions and research questions. For reasons of exhibition, however, the presentation in the main text focuses on Ordinary Least Squares (OLS) regressions with HC3 standard errors used in all models (unless otherwise noted). The findings of the preregistered analyses are not contradicted.

2.4.1 Chooser knowledge increases freedom

How do CAs behave in experiment 1? Figure 2.4 demonstrates that there is a clear trend towards fewer interventions as k rises. We can also evaluate Prediction 1 econometrically. In accordance with the preregistration, we code all values of k by their rank. Table B.3.1 in the Appendix describes the transformation. As suggested by the preregistered analysis in Section B.3.2 in the Appendix as well as robustness checks (Table B.3 in the Appendix), k causes a highly statistically and economically significant reduction in the amount of intervention (two-tailed z-test with standard errors clustered on the subject level, $z = -9.95$, $p < 0.001$).

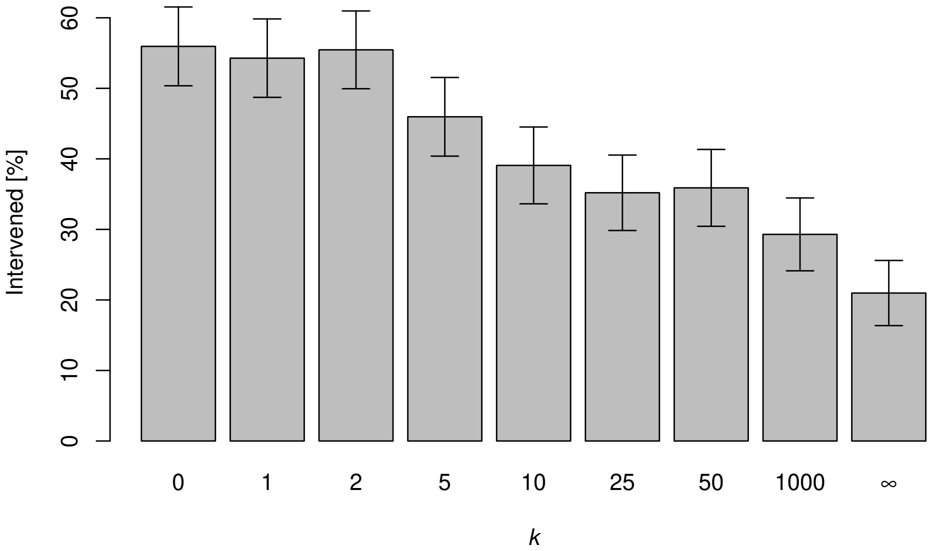


Figure 2.4: Percentage of interventions by k

Error bars span the 95% confidence interval (clustered by subject), calculated from the marginal effects in a logistic regression with treatment dummies, round number and demographic controls (Section B.3.2 in the Appendix).

It is also possible to evaluate the null hypothesis without relying on the transformation of k by using Page's L test (1963). This non-parametric test tests for a trend given repeated measurement. The null hypothesis of no trend is again rejected ($L = 63922.5$, $\chi^2_1 = 94.8$, $p < 0.001$).

Experiment 2 is able to replicate this pattern. Models 1–2 in Table 2.3 regress CA interventions on knowledge. Noteworthily, the baseline rate of intervention is much lower than in experiment 1. Recall from Section 2.3.5.1 that the order of Choosers 1 and 2 was randomized between subjects. Thus, model 3 restricts the analysis to the first Chooser that was seen. Once again, the coefficient on Full Knowledge is highly significant. This implies that knowledge enhances autonomy both in within (experiment 1, experiment 2: models 1–2 and Section B.4.4.3 in the Appendix) and between (experiment 2: model 3) analyses. From this evidence, we conclude:

Result 6. The greater knowledge in experiments 1 and 2, the greater the probability of autonomy being granted. Prediction 1 is confirmed.

	Choosers 1 and 2		Chooser 1 or 2	Chooser 1	
	Model 1	Model 2	Model 3	Model 4	Model 5
Intercept	0.220^{***}	0.289^{***}	0.199^{**}	0.488^{***}	0.657^{***}
	(0.017)	(0.062)	(0.069)	(0.077)	(0.181)
Full Knowledge	-0.143^{***}	-0.143^{***}	-0.103^{***}		
	(0.020)	(0.020)	(0.024)		
π (standardized)		-0.010	-0.019	0.107^{*}	0.163^{*}
		(0.017)	(0.019)	(0.044)	(0.065)
$\pi > 0.5$		-0.012	0.014		-0.198
		(0.057)	(0.066)		(0.193)
Block I order		0.034			
		(0.020)			
CA prefers 1		-0.095^{**}	-0.080^{*}	0.263^{**}	0.261^{**}
		(0.030)	(0.038)	(0.096)	(0.096)
Outcome	Intervened	Intervened	Intervened	Imposed 1	Imposed 1
Subset	—	—	Round 1	Intervened	Intervened
Standard errors	Clustered	Clustered	HC3	HC3	HC3
R^2	0.041	0.060	0.048	0.164	0.172
Adj. R^2	0.040	0.056	0.041	0.151	0.153
Num. obs.	1200	1200	600	132	132

$^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$

Table 2.3: Regressions of extensive and intensive margin outcomes on Chooser knowledge, CA beliefs and preferences

2.4.2 Testing Mill's intensive-margin postulate

What influences the intensive margin of interventions? That is, given that an intervention is to take place, what determines the choice between Options 1 and 2? As discussed in Section 2.2.5, Mill offered a solution that stands in stark contrast with the literature on projective paternalism (Ambuehl, Bernheim, and Ockenfels, 2021): the intervention ought to be based on an assessment of counterfactual full-information choices. We test this idea here.

2.4.2.1 Beliefs and the intensive margin

In both experiments, we asked CAs to estimate the number of people opting for one of the Options under full information.[16] In experiment 2, we also elicited unincentivized beliefs about behavior under ambiguity (that is, case *(iii)* of Section 2.3.1). For the models in Table B.2 in the Appendix, we put these data on a common scale (proportion of decision-makers thought to prefer Option 1). As in Section 2.2.1, we call this outcome π. In the models dealing with beliefs about fully informed behavior under risk, our results confirm classic findings on false consensus bias (e.g., Ross, Greene, and House, 1977). CAs believe that their subjectively preferred Option is more popular.

Result 7. Prediction 2 is confirmed (although beliefs about behavior under ambiguity cannot be explained by CAs' own preferences).

Evidently, the reported beliefs differ strongly between experiments, with beliefs in experiment 1 implausibly small.[17] Moreover, CAs predict ambiguity aversion (model 5). Note how we did not elicit data on the direction of mistakes made under ambiguity (Section 2.2.1.3). However, it is easy to see that if $\pi' > \pi$, $\epsilon_y > \epsilon_x$: to a CA who prefers Option 1, a belief in ambiguity aversion will augment mistakes in the subjectively preferred direction. To a CA who prefers Option 2, the same belief goes into the non-preferred direction. We exploit this fact below.

As outlined in Section 2.2.5.3, the presence of false consensus bias leads to an identification problem in motives. Under the majoritarian standard, CAs may

[16] In experiment 1, we elicited unincentivized beliefs about how many people out of 1,000 would prefer Option 2. In experiment 2, we used data from experiment 1 to elicit beliefs about the preferences of 300 subjects for Option 1. These beliefs were incentivized using the binarized scoring rule (Hossain and Okui, 2013).

[17] In experiment 1, beliefs were elicited at various values of p using sliders (see Screen 4 of Section B.5.1 in the Appendix).

wish to implement the Option they believe to be more popular, but if that belief is biased, beliefs *and* CA preferences must be accounted for in the analyses. As can be gleaned from the standard errors in Table B.4 in the Appendix, experiment 1 was not sufficiently powered to reliably identify the influences of $\pi > 0.5$ and π on intensive-margin interventions.

In experiment 2, beliefs have a significant, but small association with this outcome. Model 4 in Table 2.3 shows that a one standard deviation increase in π is correlated with a 10.7 p.p. increase in the probability of imposing Option 1. Model 5 demonstrates that there is no association with $\pi > 0.5$ (i.e., beliefs about majority behavior). As is clear from models 4–5, the CA's own preference also has a significant correlation with the intensive margin (Ambuehl, Bernheim, and Ockenfels, 2021). We estimate that a CA who prefers Option 1 is about 26 p.p. more likely to impose Option 1. Model 4 implies that beliefs that are $(0.263)/(0.107) \approx 2.458$ standard deviations removed from the mean are equivalent to the CA's own preference in their effect on the intensive margin. Section B.4.1 in the Appendix assesses the relative importance of beliefs and CA preferences in imposing Option 1 formally. We conclude from these investigations that beliefs are far weaker predictors of intensive-margin interventions than CAs' preferences.

Result 8. Experiment 2 shows that beliefs about majority behavior do not correlate on the intensive margin. Beliefs generally matter significantly, but much weaker as predictors than CAs' preferences.

2.4.2.2 Providing information about Chooser type

Even if we allow for biased beliefs—as in the previous Section—the CA's own preference appears to be an important predictor for the intensive margin.

When deciding for Chooser 5 in experiment 2, CAs were informed that the Chooser they were deciding for *certainly* preferred either Option 1 or Option 2 (the Chooser's "type"). Models 1 and 2 in Table 2.4 reveal that providing information about the Chooser's type does significantly influence the intensive margin selected by the CA. However, the preregistered Wald test in Section B.4.4.5 in the Appendix on the relative influence of CA and Chooser preferences reveals no significant difference between the coefficients on CA and Chooser preferences in a logistic regression ($p = 0.21$). It is important to note that while CA preference has no significant effect on the intensive margin for Chooser 5,

	Chooser 5				Chooser 1
	Model 1	Model 2	Model 3	Model 4	Model 5
Intercept	0.368^{***}	0.360	0.421^{***}	0.589^{**}	0.531^{**}
	(0.084)	(0.185)	(0.084)	(0.192)	(0.189)
π (standardized)		0.045		0.049	0.001
		(0.054)		(0.055)	(0.070)
$\pi > 0.5$		0.040		−0.185	0.049
		(0.184)		(0.186)	(0.198)
CA prefers 1	0.127	0.085	0.141	0.134	0.203^{*}
	(0.085)	(0.089)	(0.085)	(0.090)	(0.101)
Chooser prefers 1	0.254^{***}	0.263^{***}	0.192^{**}	0.200^{**}	
	(0.070)	(0.070)	(0.070)	(0.071)	
Outcome	Imposed 1	Imposed 1	Matched Chooser	Matched Chooser	Matched CA
Subset	Intervened	Intervened	Intervened	Intervened	Intervened
R^2	0.070	0.081	0.047	0.054	0.051
Adj. R^2	0.060	0.060	0.037	0.033	0.029
Num. obs.	188	188	188	188	132

$^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$

Table 2.4: Intensive margin regressions when Chooser type is known

this is not evidence of absence. The standard error on CA preference is merely large enough so that the hypothesis that *both* influences matter equally cannot be rejected.

Result 9. The Chooser's type significantly influences the intensive margin, but this influence is not significantly different from the influence of the CA's preference (research question 2).

To investigate this issue further, we estimated models 3–4 in Table 2.4. Here, the outcome variable is whether the CA actually implemented the Chooser's counterfactual choice. Note that about a majority of CAs "match" the Chooser. We detect a significant influence of the *Chooser's* preference. That is, if the Chooser prefers Option 1, that Option is about 20 p.p. likely to be implemented for the Chooser—ceteris paribus—than if he prefers Option 2, even when controlling for the CA's own preference.[18] The correlational analysis in model 5 reveals that the same pattern emerges for Chooser 1.

[18]We can verify this finding using the following highly conservative robustness check: among all CAs who intervened for Chooser 5, we can condition on their own preference and that of the Chooser. 54 of all intervening CAs preferred Option 1 and were presented with a Chooser preferring Option 1. In that case, Option 1 was implemented 43 times (79.6%). 23 of all intervening CAs preferred Option 2 and intervened for a Chooser preferring Option 2. 12 of these CAs imposed Option 2 (52.2%). The difference is significant (two-sided test of equal proportions, $\chi^2 = 4.69$, $p = 0.03$).

This result stands in contrast to findings concerning decision-making for others under risk (Polman and Wu, 2020), where decisions for others are more risky. Similarly, a literature has characterized the influence of social context on risk-taking (e.g., Bolton, Ockenfels, and Stauf, 2015; Schwerter, 2024) or the role of social preferences (Bolton and Ockenfels, 2010). The present finding highlights the nuances between policymakers' own preferences and the intensive margin. Some choices may be viewed as more respectable or objectively correct beyond the CA's own preference and thus more likely to be implemented. Future work can seek to disentangle CA preferences from such cosmic ideals or "bliss points." One possible explanation revolves around regret and guilt: any imposition in some sense makes the CA the Chooser and thus responsible for the outcome. Option 2 did yield €0 with probability 0.2, a significant possibility. Moreover, Option 2's expected value is only €1 higher than the safe amount, making it relatively unattractive. In that case, a preference for the safe Option 1 may result from an "interpersonal" kind of regret. Corbett, Feeney, and McCormack (2021) discusses pro-social risk-taking and its relation to regret and guilt further. Our own results do not allow us to make a statement about any underlying cognitive mechanisms.

Result 10. The riskless Option 1 is significantly more likely than Option 2 to be implemented. CAs do not simply implement the full-information counterfactual choice of the Chooser or their own preference.

2.4.3 Connecting formal theory with data

Our theory of Section 2.2.1 allows us to estimate ϕ using only three observable parameters: beliefs about fully informed preferences (π), CAs' own preferences and intensive margin interventions. Recall that, on the intensive margin, option x will be implemented if $W_\theta(x) > W_\theta(y)$ (e.g., Equations 2.2–2.3 for $\theta = x$). We can introduce unobserved and independent error terms on both sides of this inequality to yield a "random utility" (or Fechner-type) model. Let $\xi_x \sim N(0, \sigma^2_{\xi_x})$, $\xi_y \sim N(0, \sigma^2_{\xi_y})$. In that case, option x will be implemented on the intensive margin if

$$
\begin{aligned}
& W_\theta(x) + \xi_x > W_\theta(y) + \xi_y, \\
\Leftrightarrow \quad & \underbrace{\xi_y - \xi_x}_{\sim N(0,\sigma)} < W_\theta(x) - W_\theta(y).
\end{aligned}
$$

Call Φ the standard-normal cumulative distribution function and let I_i^1 be an indicator of whether CA i implemented Option 1. Using information on each CA's type, $\theta_i \in \{1, 2\}$, the following log-likelihood function can be maximized with respect to ϕ, σ to yield consistent estimates of these parameters:

$$\sum_i \left[I_i^1 \log \Phi\left(\frac{W_{\theta_i}(1) - W_{\theta_i}(2)}{\sigma}\right) + (1 - I_i^1) \log\left(1 - \Phi\left(\frac{W_{\theta_i}(1) - W_{\theta_i}(2)}{\sigma}\right)\right)\right]. \tag{2.9}$$

When we maximize Equation 2.9 for intensive-margin decisions vis-á-vis Chooser 1, we find $\widehat{\phi} = 0.294$ (95% confidence interval: $[0.081, 0.508]$) and $\hat{\sigma} = 0.809$ (95% confidence interval: $[0.504, 1.114]$). All confidence intervals in this Section are calculated from the Fisher information matrix.

The same can be done for Chooser 5, where CAs had certain knowledge about the Chooser's type. Here, we find $\widehat{\phi} = 0.430$ (95% confidence interval: $[0.221, 0.640]$) and $\hat{\sigma} = 1.767$ (95% confidence interval: $[0.840, 2.694]$). Similar results for ϕ can be obtained for Chooser 3.

In sum, ϕ is reasonably similar between Choosers 1, 3 and 5 and, in each case, significantly different from 0. As we describe in Appendix B.4.2, these binary-choice models fit the data about as well as model 5 in Table 2.3 and model 2 in Table 2.4. However, they do not work well for Chooser 2 (few CAs intervened) and Chooser 4 (few CAs could intervene and few did).

Result 11. The formal model of Section 2.2.1 performs well in describing intensive-margin behavior: ϕ is estimated at about one-third for Choosers 1, 3 and 5. $H_0 : \phi = 0$ is rejected.

2.4.4 Knowledge uncertainty

CAs' decision-making for Chooser 3 involved a scenario in which the Chooser could be in either of two states: fully informed or not informed. Both states were equally likely. Table 2.5 gives descriptive statistics, grouped by the decisions regarding Choosers 1 and 2. The data follow a plausible pattern: those who intervened more before will intervene more for Chooser 3. However, comparing rates of intervention (model 1 in Table 2.6 gives the average) with those in models

	Int. for Chooser 1	Not Int. for Chooser 1
Int. for Chooser 2	(12) **0.833** [0.516, 0.979]	(34) **0.412** [0.246, 0.593]
Not Int. for Chooser 2	(120) **0.400** [0.312, 0.493]	(434) **0.267** [0.226, 0.312]

Table 2.5: Intervention rates for Chooser 3 conditional on behavior toward Choosers 1 and 2

Chooser 1 is uninformed, Chooser 2 is fully informed (Section 2.3.5.1). The table shows, small and in parentheses, a cross-tab of occurrences; in bold, the rate of intervention for Chooser 3 conditional on behavior for Choosers 1 and 2; the rate's 95% confidence interval.

	Chooser 3			
	Model 1	Model 2	Model 3	Model 4
Intercept	0.313^{***}	0.262^{***}	0.221^{*}	0.524^{**}
	(0.019)	(0.021)	(0.103)	(0.177)
Int. for Chooser 1		0.158^{***}	0.168^{***}	
		(0.047)	(0.049)	
Int. for Chooser 2		0.219^{**}	0.210^{**}	
		(0.075)	(0.076)	
π (standardized)			0.000	0.043
			(0.030)	(0.047)
$\pi > 0.5$			0.081	−0.049
			(0.098)	(0.168)
Block I order			−0.037	
			(0.038)	
CA prefers 1			−0.018	0.288^{**}
			(0.052)	(0.093)
Outcome	Intervened	Intervened	Intervened	Imposed 1
Subset	—	—	—	Intervened
R^2	0.000	0.037	0.041	0.091
Adj. R^2	0.000	0.033	0.031	0.076
Num. obs.	600	600	600	188

$^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$

Table 2.6: Regressions explaining behavior toward a Chooser in uncertain state of knowledge

1–3 of Table 2.3 reveals that rates of intervention for Chooser 3 vastly exceed even those for Chooser 1.[19] We can only speculate why this is the case.

Result 12. Chooser 3 attains vastly higher intervention rates than Choosers 1 and 2.

Moreover, models 2–3 in Table 2.6 and the preregistered analysis of Section B.4.4.4 in the Appendix reveal that the decision to intervene for Chooser 3 is about equally well explained by each of the decisions to intervene for Choosers 1 and 2. Both of these prior decisions are highly significant in predicting behavior vis-á-vis Chooser 3. Model 4 once more highlights the predictive power of CAs' own preference in shaping the intensive margin.

Result 13. The decisions for Choosers 1 and 2 are about equally predictive for behavior towards Chooser 3 (research question 3).

2.4.5 Providing information

CAs were enabled to communicate to Chooser 4 the value of p (0.2) in the choice between Options 1 and 2.[20] In addition, some CAs—those in the treatment Plus—were able to intervene as well.[21] Sections 2.2.1.3, 2.2.4 and 2.3.5.1 motivated this design choice. Simply put, real-life policymakers are not restricted from using multiple policy tools simultaneously (e.g., Chapter 3) and they can strategically use information provision to achieve their ends.

We can test whether information is actually strategically provided. As stated in Section 2.2.4, if it is a given that the Chooser's decision will be implemented, the CA's calculus is fundamentally different then when she can combine both intervention and information provision. Model 1 in Table 2.7 and the preregistered analysis in Section B.4.4.2 in the Appendix demonstrate that not only do CAs not exploit strategic information provision, but CAs in the Plus treatment slightly exceed the degree of information provision observed in the baseline. This difference is not significant, but standard errors are very small. We thus

[19]We can use the binary indicators of intervention for Choosers 1 and 3 to test whether the within-subject difference is significantly different from zero: two-sided paired t-test with unequal variances, $t = -3.52$, $p < 0.001$.

[20]In our design, CAs had to deliberately choose whether to reveal p or not, as in Bartling et al. (2023). We, too, made sure that CAs' involvement in providing information is not revealed to the Chooser.

[21]In this Section, we use the word "intervene" only to refer to an intervention in the choice between Options 1 and 2, although some authors view information provision as an intervention (e.g., Bartling et al., 2023; Mabsout, 2022).

	Chooser 4		
	Model 1	Model 2	Model 3
Intercept	0.803***	0.974***	0.942***
	(0.023)	(0.105)	(0.154)
Plus	0.011	0.012	
	(0.032)	(0.032)	
π		−0.056	−0.056
		(0.114)	(0.169)
π'		−0.085	−0.007
		(0.121)	(0.195)
$\pi > \pi'$		−0.087	−0.062
		(0.079)	(0.108)
CA prefers 1		0.048	0.021
		(0.066)	(0.076)
Mistakes benefit CA		−0.134*	−0.094
		(0.062)	(0.071)
Intervened			−0.297**
			(0.103)
Outcome	Info provided	Info provided	Info provided
Subset	—	—	Plus
R^2	0.000	0.017	0.064
Adj. R^2	−0.001	0.007	0.044
Num. obs.	600	600	290

***$p < 0.001$; **$p < 0.01$; *$p < 0.05$

Table 2.7: Regressions explaining CAs' information provision

replicate the finding by Bartling et al. (2023) that a vast majority of CAs provide information to Chooser 4—even in the treatment.

Result 14. CAs do not strategically employ information provision (prediction 3).

This result suggests that independence of irrelevant alternatives is indeed satisfied (Section 2.2.1.3).

2.4.5.1 Who doesn't provide information?

This Section reports additional, exploratory analyses to further investigate the relationship between intervention and information provision. As we noted in Section 2.2.1.3, not all Chooser mistakes are created equal. It can be shown that if $\pi' > \pi$—as in the case of ambiguity aversion (Section 2.4.2.1)—$\epsilon_y > \epsilon_x$. To a CA who herself prefers Option 1, that is actually a good mistake. Similarly, a CA with a preference for Option 2 would prefer $\pi' < \pi$, implying $\epsilon_y < \epsilon_x$. In other words: with self-interest, CAs benefit from Chooser mistakes that go into CAs' subjectively preferred direction.

Our experiment did not elicit data on ϵ_x and ϵ_y. However, under the conditions in Corollary 1, non-provision of information can be optimal for sufficiently high ϕ and CA-preferring Chooser mistakes. While we do not have individual-level estimates of ϕ, we can provide evidence that this consideration is very real. We define a new variable,

$$\begin{aligned}\text{Mistakes benefit CA} = (&\text{CA prefers Option 1} \land \pi' > \pi) \lor \\ &(\text{CA prefers Option 2} \land \pi' < \pi)\,,\end{aligned}$$

and add it to model 1 of Table 2.7. Moreover, we add the CA's own preference, raw beliefs and a dummy about predicted ambiguity aversion. Model 2 demonstrates that if Chooser ignorance benefits CAs, the latter are about 13.4 p.p. less likely to provide information.[22] However, the effect is likely too small to be detectable among CAs in the treatment group (model 3).

[22] A two-sided test of proportions confirms this pattern, $\chi^2 = 7.31$, $p = 0.007$. However, to ensure that this difference is not driven by secular differences in the distribution of beliefs, it is important to adjust for these confounding variables using a regression model. A test on the coefficient of "Mistakes benefit CA" in the logistic regression equivalent to model 2 in Table 2.7 reveals $p = 0.011$.

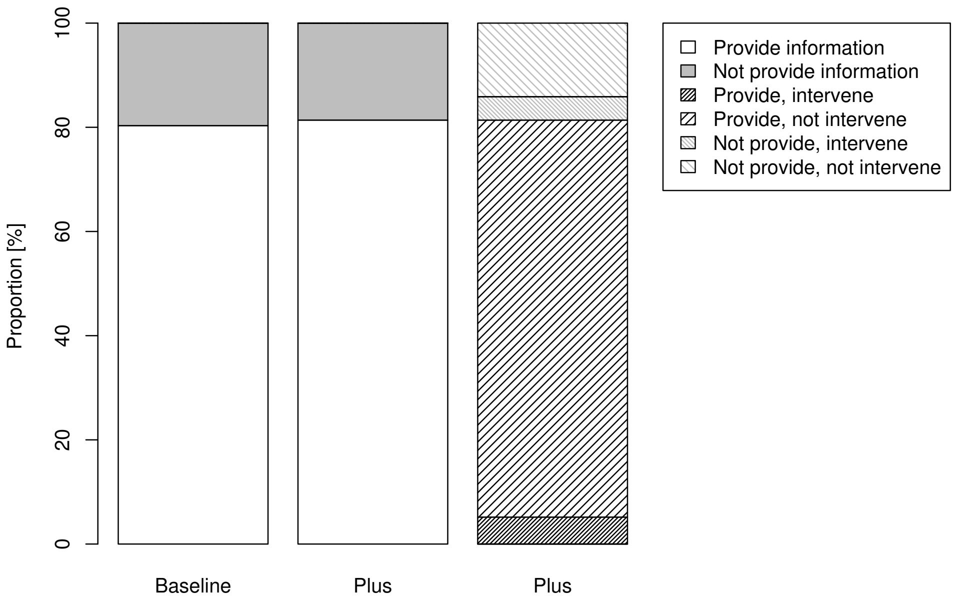


Figure 2.5: CA types in the population

Result 15. CAs are slightly less likely to provide information if Chooser mistakes increase the take-up of their subjectively preferred Option.

Model 3 in Table 2.7 restricts the analysis to CAs in the Plus treatment and regresses information provision on whether an intervention in Chooser 4's decision between Options 1 and 2 actually occurred. Although only 28 out of 290 CAs in the Plus treatment intervened, the difference in information provision between interveners and non-interveners is highly statistically and economically significant.

Result 16. CAs that actually intervene are much less likely to provide information to Chooser 4.

Model 8 in Table B.6 in the Appendix shows that those who intervened for Chooser 4 are especially likely to have intervened for other Choosers, too (see the discussion in Section B.4.3 in the Appendix). This hints at a hidden type of CA that is characterized by not providing information and intervening if possible. Out of the 54 CAs in Plus that did not provide information, 13 (24.1%) intervened. Out of the 236 CAs in Plus that did provide information, 15 (6.4%)

intervened.[23] Recalling Model 1 in Table 2.3, these values are almost identical to rates of intervention for Choosers 1 and 2.

The treatment thus helps us to uncover a minority of CAs that creates a lack of knowledge on the side of Choosers to intervene (Figure 2.5). This group of CAs makes up $(1 - [0.803 + 0.011]) \cdot (0.241 - 0.064) \approx 3.3\%$ of the sample. While small, this proportion is highly statistically significant.[24] This type of CA is not revealed by the baseline.

2.5 Conclusion

This work explored the role of knowledge in paternalism. We found across two experiments that more knowledge on the side of Choosers causes a vast increase in the autonomy they are granted by impartial Choice Architects (CAs). Information helps Choosers make the right decision and CAs overwhelmingly respect that. On the other hand, a *lack* of knowledge is taken by CAs as a right to intervene and prevent incorrect inference. Most CAs do not wish to override the Chooser's choice. They prefer to provide information, even when they would be able to obscure their intervention through the non-provision of knowledge. However, there is a minority of CAs that strategically abstains from providing information.

Paternalistic action is highly nuanced and context-dependent. If CAs do not know the Chooser's type, they rely on a proxy—their own preference—to select the intensive margin. If provided with the type, they use this information in conjunction with their own preference to arrive at the intensive margin. In both of these cases, a riskless Option was much more likely to be implemented. This hints at the existence of ideals for intervention that may be correlated with, but conceptually distinct from, CAs' own preferences. This is a qualification to our model and recent findings on "projective paternalism" (Ambuehl, Bernheim, and Ockenfels, 2021).

Policymakers' and decision-makers' beliefs, knowledge and preferences matter profoundly for regulation. Mill's arguments have stood the test of time. The idea that an intervention's intensive margin must reflect a full-information counterfactual underscores a central tenet of many classically liberal views of

[23] This difference is statistically different: two-sided test of equal proportions, $\chi^2 = 13.8$, $p < 0.001$.

[24] Bootstrapping with 250,000 replicates, we find a 99% confidence interval of $[0.005, 0.0666]$, $p < 0.002$.

governance: interventions can be justified based on a lack of knowledge on the extensive margin, but their precise embodiment on the intensive margin ought to be value-free. Mill's exposition does not leave much room for taste-based intervention. Yet, as we demonstrate experimentally, policymakers distinctly mix in their personal vision of what's right even when informed about what the Chooser would have done if he had full information. As polities consider the implementation of new paternalistic policies, understanding the subjective nature of regulation is an essential condition for creating legitimate institutions and laws.

Chapter 3

Paternalism and Deliberation: An Experiment on Making Formal Rules

Abstract

This paper studies the relationship between soft and hard paternalism by examining two kinds of restriction: a waiting period and a hard limit (cap) on risk-seeking behavior. Mandatory waiting periods have been instituted for medical procedures, gun purchases and other high-stakes decisions. Are these policies substitutes for hard restrictions, and are delayed decisions more respected? In an experiment, decision-makers are informed about an impending high-stakes decision. Treatments define when the decision is made: on the spot or after one day, and whether the initial decision can be revised. In a general population survey experiment, another class of subjects (Choice Architects) is granted the opportunity to make rules for decision-makers. Given a decision's temporal structure, Choice Architects can decide on a cap to the decision-maker's risk taking. In another treatment, Choice Architects can implement a mandatory waiting period in addition to the cap. This allows us to study the substitutional relationship between waiting periods and paternalistic action and the effect of deliberation on the autonomy afforded to the decision-maker. Our highly powered experiment reveals that exogenous deliberation has no effect on the cap. Moreover, endogenously prescribed waiting periods represent add-on restrictions that do not substitute for the cap. Choice Architects believe that, with time,

the average decision-maker will take less risk and—because of the distribution of Choice Architects' bliss points—come closer to Choice Architects' subjective ideal choice. These findings highlight the complementarity of policy tools in targeting various parts of a distribution of decision-makers.

Acknowledgments

We thank David Albrecht, Arno Apffelstaedt, James Druckman, Hannah Hamilton, Felix Kölle, Melisa Kurtis, Aenne Läufer, Axel Ockenfels and Rastislav Rehák, audiences at Berlin, Buckingham, Cologne, Maastricht and Riga and two anonymous reviewers at Time-sharing Experiments for the Social Sciences (TESS) for valuable comments. We thank TESS and the NORC at the University of Chicago for outstanding survey-related support. We are grateful to Ben Mrakovcic for his implementation of the Chooser experiment. Any errors are our own. We acknowledge funding by the Center for Social and Economic Behavior (C-SEB) at the University of Cologne and the German Research Foundation (DFG) under Germany's Excellence Strategy (EXC 2126/1–390838866).

Declarations

Financial support was received from the Center for Social and Economic Behavior (C-SEB) at the University of Cologne and the German Research Foundation (DFG) under Germany's Excellence Strategy, EXC 2126/1–390838866. In-kind support was also received from the U.S. National Science Foundation through TESS: Data collected by Time-sharing Experiments for the Social Sciences, NSF Grant 0818839, Jeremy Freese and James Druckman, Principal Investigators. The author has no competing interests to declare that are relevant to the content of this article.

IRB approval was granted by the WiSo Ethics Review Board at the University of Cologne and the NORC.

Data availability

All materials and data used in the analysis are freely available at `https://gitlab.com/gr0ssmann/undo`.

The experiment was preregistered at the Open Science Framework. See here for the preregistration: https://osf.io/m7526.

◇

> Leibniz never married; he had considered it at the age of fifty; but the person he had in mind asked for time to reflect. This gave Leibniz time to reflect, too, and so he never married. *(Bernard de Fontenelle)*

3.1 Introduction

The Cambridge Dictionary defines a rule as "an accepted principle or instruction that states the way things are or should be done, and tells you what you are allowed or are not allowed to do."[1] Formal rules such as man-made, positive laws have become abundant in many countries. However, the study of rule-making has received scarce attention in experimental economics. This paper studies how the inner content of formal rules is created and how rules interact with each other as policy tools. We use an exploratory, but minimal and precisely preregistered survey experiment to estimate the relationship between *(i)* limits to risky behavior and *(ii)* mandatory waiting periods. One of these rules is softly paternalistic, while the other is an expression of hard paternalism. We contrast these rules to empirically investigate the political economy of the link between policy tools. Moreover, we study the mental models that relate to rule-making: how do beliefs about others' behavior shift when more or less time to decide is granted to decision-makers? If rules aim to change behavior, and external factors like deliberation already do so, does this reduce the need for strict paternalistic rules?

A large literature has studied the role of deliberation on human decision-making. The idea that deliberative choices are different from, and perhaps superior to affective ones has reverberated through the ages. Behavioral economists have long conducted experiments or modeled behavior where the time available for deciding between courses of action is exogenously varied (e.g., Caplin, 2016; Caplin and Martin, 2016; Imas, Kuhn, and Mironova, 2022; Kocher and

[1] Source: https://dictionary.cambridge.org/dictionary/english/rule, accessed January 2, 2025.

Sutter, 2006). Loewenstein, O'Donoghue, and Bhatia (2015) describe models (or "systems") of human behavior that are distinguished by the immediacy of choice. A related literature has studied decision-making under "hot" and "cool" states. Metcalfe and Mischel (1999) explicitly describe the "cool system" as "contemplative" and the "hot system" as "impulsive" (p. 3).

The temporal structure of decision-making has also been purposely altered through government policy, often with the stated intention of meliorating choices. One example of such policy are mandatory waiting periods. Legislation has created mandatory waiting periods in various policy areas. While the U.S. federal government does not currently impose a waiting period for firearm purchases, several states do so (e.g., Edwards et al., 2018; Luca, Malhotra, and Poliquin, 2017). Jurisdictions in the U.S. and around the world require women seeking abortions to wait (e.g., Joyce et al., 2009; Lindo and Pineda-Torres, 2021). Sterilization (Rowlands and Thomas, 2020), marriage, divorce (Lee, 2013) and adoption are other examples of decision-making affected by governmentally imposed mandatory waiting periods in many jurisdictions. These laws are characterized by a temporal decoupling of choice and action. In other words, spontaneous behavior is curtailed by giving decision-makers additional time to back out of a decision. Similarly, the term "cooling-off periods" is used for laws that allow consumers to undo large purchases or contracts (Rekaiti and Van den Bergh, 2000; Sher, 1967; Sovern, 2013). The term is also used in negotiation protocols to avoid or resolve conflicts. Oechssler, Roider, and Schmitz (2015) use an online experiment to study the role of cooling-off periods in negotiations. Although these approaches are different from the present research, there appears to be a universal recognition that spontaneous behavior may be different from more contemplative behavior, perhaps to the detriment of some standard of decision-making quality.[2]

This paper uses this idea for dual purposes. First, we follow a recent line of investigation into behavioral determinants of the *supply* of paternalism (Ambuehl, Bernheim, and Ockenfels, 2021). In these studies, a decision-maker ("Chooser") is matched with a policymaker ("Choice Architect", CA). The CA can intervene in a decision faced by the Chooser. Some of the work in this area has modified the choice ecology of Choosers to ascertain the effect of situational factors on the degree of autonomy granted to Choosers. For example, Chapter 2

[2]Fudenberg, Strack, and Strzalecki (2018) modeled the joint distribution of choices and decision times, explaining empirical findings that agents that decide more quickly are more likely to choose correctly. However, in their model, decision times and choice quality are endogenous.

varies the knowledge the Chooser has about a choice between two options. The more knowledge the Chooser possesses, the fewer interventions take place. We, too, modify one crucial aspect of how people can decide: how much time they have to think about a decision, and whether they can revise it if necessary. We have Choosers participate in a simple risky decision: how many boxes to open (Crosetto and Filippin, 2013). In this experiment, one randomly chosen box out of twenty-five contains a "curveball" that eliminates all earnings. Because each non-curveball box earns the Chooser $20, the experiment has very high stakes. We vary the temporal structure of decision-making and let CAs intervene by setting a cap on the number of boxes that can be opened by the Chooser. This action effectively constrains Choosers' risk-taking. This allows us to test whether Chooser deliberation causes increased CA tolerance and what CAs believe about Choosers that have more time to think about their decision. This approach sheds light on the inherently subjective nature of rule-making and the mental models applied therein. Nonetheless, it should be noted that we shed light on but one particular determinant of rule-making. Paternalism itself can be understood in a variety of ways; for example, as a social preference that is strongly anchored in preventing others from making losses. Such losses are subjectively perceived by rule-makers (e.g., employers: Buchmann, Meyer, and Sullivan, 2024). However, other motives for intervention remain possible. That is also true for motives that would not usually be considered paternalistic.

Second, we recognize that waiting periods and caps are two rules that, in practice, may be deployed simultaneously. For example, political scientists and environmental economists have studied "policy mixes" (e.g., Bouma et al., 2019). This term is also used by macroeconomists to refer to the joint application of monetary and fiscal policy. By employing multiple instruments in a single policy area, policymakers' goals are thought to be attained more effectively. Our experiment contains one treatment to test for the relationship between the waiting period and the cap. In this treatment, CAs can implement both the cap *and* a waiting period. This is to test for substitution between rules: do those CAs who implement a waiting period relax the cap? If deliberation is thought to move behavior in a preferred direction, hard-nosed restrictions may become less attractive if CAs generally respect autonomous decision-making.

Gerald Dworkin's oft-cited definition of paternalism relates to an "interference with a person's liberty of action justified by reasons referring exclusively to the welfare, good, happiness, needs, interests or values of the person being coerced" (Dworkin, 1972, p. 65). However, even policies that do not involve coercion yet attempt to meliorate choices—such as nudges (Thaler and Sunstein, 2008)—are virtually universally viewed as inherently paternalistic. Indeed, Thaler and Sunstein speak of "libertarian paternalism" (e.g., Sunstein and Thaler, 2003; Thaler and Sunstein, 2003), even though nudges explicitly do not restrict choice sets. Rather, nudges are thought to guide behavior in autonomy-preserving ways (Thaler and Sunstein, 2008). Many authors (e.g., Kirchgässner, 2017; Schnellenbach, 2012) have made a distinction between "soft" and "hard" paternalism. This differentiation was originally introduced by Feinberg (1989). While some authors have attempted to motivate the difference between the concepts by referring to the "costs" imposed on a decision-maker (e.g., Sunstein, 2014) or, relatedly, the strength of intervention (Rizzo and Whitman, 2009, p. 687), Feinberg originally saw voluntariness as the deciding factor (Hanna, 2018). If a decision would be "voluntarily" made—e.g., by a competent adult—interference is an expression of hard paternalism. On the other hand, the prevention of ill-informed or incompetent choices is softly paternalistic. It has been argued that this distinction conforms well with John Stuart Mill's views on paternalism (Mabsout, 2022). The provision of information to uninformed decision-makers and even the prevention of uninformed decisions (e.g., Chapter 2) may be viewed as softly paternalistic measures.

We need not dwell here on the conceptual difficulties arising from Feinberg's distinction (e.g., Hanna, 2018). Waiting periods are inherently softly paternalistic policies. They are choice-preserving, yet impose a temporal delay between decision and action, allowing decision-makers to reflect on their course of action and potentially reverse it. The intention of these policies is to prevent impulsive, ill-considered decisions, appealing to Feinberg's delineation. It has to be noted that waiting periods impose costs on decision-makers (Loewenstein and Haisley, 2007, note 7) and—since they affect all who wish to take an action, irrespective of their pre-existing level of consideration—waiting periods are not asymmetrically paternalistic (Camerer, Issacharoff, et al., 2003). However, they allow decision-makers to take any action. To the extent that waiting is less costly than a ban, they fit Sunstein's (2014) definition, too: "Soft paternalism is weaker and essentially libertarian, in the crucial sense that it preserves freedom of choice" (p. 19). Similarly, Scoccia (2018, p. 18) writes: "[S]ome nudges, such as mandatory 'cooling off' periods for major financial or medical care decisions, may actually enhance autonomy [...]" On the other hand, laws that constrain behavior remove autonomy in a hard paternalistic manner (e.g., Rizzo and Whitman, 2009). Informed decision-makers face the cost of not being able to act on their true preferences, or even punishment for violations, indicating the hard paternalistic nature of such rules. Sunstein (2014, ch. 2) discusses some of these nuances.

To our knowledge, the relationship between “soft” and “hard” paternalism has never been studied empirically. Tor (2022) notes that nudges and similar behavioral intervention can be combined with and perhaps supplant classic regulation. In this paper, we conduct simple, exploratory experiments to investigate that relationship. A. E. Roth (1986) early recognized the potential of economic experiments for theory-building. This paper contributes to experimental political economy by highlighting the inherently subjective nature of rule-making. Our findings and the new hypotheses they inspire may add to our understanding of real-world issues in governance and the design of institutions. Moreover, we contribute to a literature that investigates the revision of choices. In economics (e.g., Benjamin, Fontana, and Kimball, 2020; Nielsen and Rehbeck, 2022), these studies have focused on disentangling intended but possibly nonstandard behavior from unintended mistakes. Secondly, a rich literature in social psychology (e.g., Gilbert and Ebert, 2002) has studied how behavior and attitudes are shaped by allowing people to change their minds after making an initial choice. The present paper addresses others’ responses to a decision-maker’s capability of revising a choice.

First, we establish that most CAs believe that average Choosers open too many boxes relative to CAs’ injunctive norm. Moreover, CAs tend to set the cap above their injunctive norm, replicating Chapter 1’s finding that CAs leave room for Choosers to express themselves. Second, our highly powered survey experiment reveals that exogenous Chooser deliberation has no effect on the cap. This result is confirmed when considering CAs who were enabled to use both the cap and the waiting period: neither the possibility of implementing the waiting period nor actually implementing it causes a change in the cap. Third, preregistered analyses reveal that deliberation causes CAs to become more optimistic about Chooser actions. This optimism is driven by a forecast reduction in the risk-taking by the average Chooser if he is given deliberation. Our results are consistent with a subjective theory of rule-making in which the cap is to constrain extreme Choosers, while the waiting period is thought to act on more moderate Choosers.

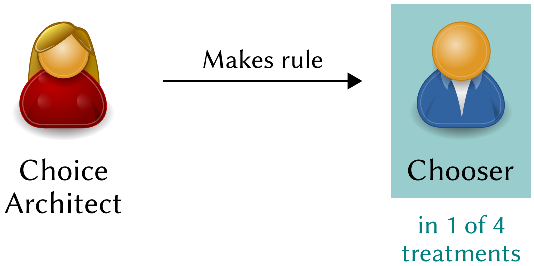


Figure 3.1: Experimental design used in this paper

The remainder of this paper proceeds as follows: Section 3.2 presents our experimental design. Section 3.3 develops testable hypotheses. Section 3.4 examines empirical results. In Section 3.5, we discuss our findings and conclude.

3.2 Experimental design

This paper relies on the basic experimental design used and popularized by Ambuehl, Bernheim, and Ockenfels (2021). There are two types of subject: Choosers and Choice Architects (CAs). A Chooser is put in one of four treatments, and a CA is matched to him. The CA is given an explanation regarding the Chooser's experiment (including the treatment); she can then create a rule for the Chooser, constraining Chooser actions (Figure 3.1).[3] The experiment has a between design and is free of deception. We now describe the game faced by Choosers.

3.2.1 Chooser experiment

Choosers participate in a two-day survey. At some point during these two days—as determined by treatment, see below—Choosers play the Bomb Risk Elicitation Task (BRET, Crosetto and Filippin, 2013) with the highest stakes ever reported in the literature. Our BRET works as follows: Choosers are faced with 25 boxes. Each box contains $20 to be collected by the Chooser. They can open whichever and as many boxes as they like, but one randomly selected box contains a "bomb."[4] If the one box containing the "bomb" is opened, all earnings

[3]In our experiment, CAs' decisions were stochastically implemented. There were four Choosers in total—one per treatment—and a CA was randomly selected for each of these Choosers to have their rule implemented. CAs knew that it was not certain whether their rule would be implemented, but all CAs had a strictly positive probability of making a rule for a real Chooser.

[4]In our experiment, the word "curveball" was used instead of "bomb," because the word "bomb" can carry negative associations. The Cambridge Dictionary lists the metaphorical use of

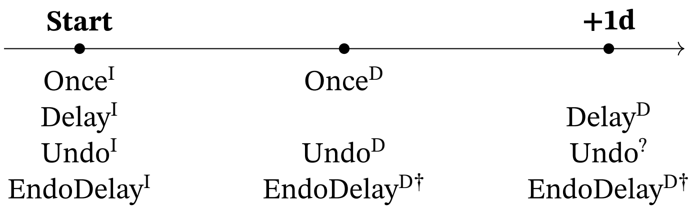


Figure 3.2: Treatments

[I]Chooser obtains BRET instructions. [D]Chooser decides. [?]Optional revision to decision. [†]CA decides when decision is made. *Note:* In all treatments, the Chooser experiment took two days.

are eviscerated, leaving the Chooser with no payment from the BRET. Choosers learn these rules on day 1 of the survey. While the BRET is ordinarily used to measure risk preferences, we use it to provide an intuitive game of risk taking.

3.2.2 CA experiment

We use a simple population-based survey experiment to test the hypotheses outlined below in Section 3.3.[5] In such a survey, respondent samples representative of a target population can be used to combine the internal validity of experiments with the generalizability and lack of self-selection of a diverse pool of respondents (Mutz, 2011). Our experiment contains four total treatments that are exogenously assigned to Choosers *viz.* CAs (see Figure 3.2).[6] Our treatments vary the temporal structure of the Chooser's decision-making. In Once, the Chooser decides straight after learning the rules of the game (Section 3.2.1). In Delay, the Chooser has to wait one day between learning the game and actually opening the boxes. In Undo, the Chooser decides straight after learning about the BRET, but he can revise the day-1 decision on day 2 of the survey. In EndoDelay, the CA decides whether the Chooser will be in Once or Delay. Since the Chooser experiment always takes two days, it is not possible for CAs to save Choosers time.

In *all* treatments, CAs may set a cap on the number of boxes that can be opened by the Chooser. However, in EndoDelay, CAs also choose when the

---

the word "curveball" as implying, in American English, "something such as a question or event that is surprising or unexpected, and therefore difficult to deal with" (`https://dictionary.cambridge.org/dictionary/english/curveball`, accessed January 2, 2025).

[5]In this paper, we focus on the decisions and the experiment faced by CAs.

[6]Since each CA decided for only one Chooser in one particular treatment, it is accurate to say that CAs were "assigned" to the treatment that their Chooser was assigned to. However, recall that only some CAs' decisions were implemented.

Chooser is to decide (on day 1 or day 2). In other words, the first three treatments allow the implementation of a hard paternalistic rule, while EndoDelay also grants access to a softly paternalistic policy.

The CA experiment consisted of two pages shown to respondents.[7] The first page explained the Chooser's decision using a simple visualization, elicited an injunctive norm, beliefs about the behavior of the *average* Chooser and beliefs about their happiness. Because respondents are not easily incentivized through our survey platform, we elicited only unincentivized beliefs. CAs received standard survey incentives (akin to a flat fee). Page 2 provided a treatment reminder and allowed CAs to make rules. CAs were told that to make no restrictions on the Chooser's behavior, they would have to enter "25," the maximum number of boxes a Chooser can open. Beyond the treatment reminder, page 2 differed only for CAs in EndoDelay, as they could choose when the Chooser was to make their decision. In all treatments except EndoDelay, four variables were elicited from CAs (injunctive norm, belief, happiness, cap). In EndoDelay, five were elicited (injunctive norm, belief, happiness, cap, when Chooser is to decide). This highlights the minimal nature of our otherwise exploratory experiment.

All instructions were optimized for general-population usage. The experiment was conducted in November 2023 in the English and Spanish languages using NORC AmeriSpeak through Time-sharing Experiments for the Social Sciences (TESS, `tessexperiments.org`). NORC AmeriSpeak is a renowned survey pool focusing on providing representative, high-quality samples of the United States population for research purposes. Unlike other survey pools, AmeriSpeak respondents are typically limited to participating in only a few surveys each month. TESS enables researchers to access NORC AmeriSpeak for free through a competitive grant scheme. IRB approval was granted by the WiSo Ethics Review Board at the University of Cologne and the NORC, and the experiment was fully preregistered. The Chooser experiment implemented the decisions of four CAs (one per treatment). It was conducted in October 2024 at the Cologne Laboratory for Economic Research. This experimental protocol included box rewards stated in U.S. dollars, exactly as indicated to CAs (Section 3.2.1). Official exchange rates by the European Central Bank were used to convert rewards into euros.

[7] English-language instructions are available in Section C.1 in the Appendix.

3.3 Hypotheses

We use the final project proposal (available in Grossmann, 2023, a web resource) to structure our research questions and hypotheses. The content (though not the language) of the following hypotheses were submitted and approved by TESS through a refereeing process before the experiment was conducted. Moreover, the proposal (Grossmann, 2023, appendix C) and the preregistration spelled out in detail how to test these hypotheses econometrically. This strategy serves to constrain the exploratory nature of this study and make inference more robust.

3.3.1 Endogenously imposed waiting periods

We can characterize rules as a mechanism to shift behavior on the side of Choosers. Several previous experiments on paternalism (e.g., Chapters 1, 2; Ambuehl, Bernheim, and Ockenfels, 2021) have considered situations where Choice Architects (CAs) could intervene in the decisions of Choosers without facing any material costs. Nonetheless, CAs do not simply impose their own preference; they leave space for Choosers to express themselves (Chapter 1). To some extent, autonomy appears to be respected.

Consider again the relationship between soft and hard paternalism. If a CA can choose between two equally effective means of achieving the same ends, she may well choose the less intrusive one. We certainly do not claim that soft and hard measures are equally effective. Nonetheless, marginal CAs may be shifted to a less interventionist rule when they can simultaneously employ less coercive policy tools. We can test whether the possibility to choose a soft intervention substitutes for hard imposition (Tor, 2022).

Hypothesis 4. The availability of a soft intervention substitutes for hard intervention.

Moreover, we can group together those CAs who intervene in *any* sense (hard or soft) if either provided or not provided the opportunity to intervene softly. We can test whether the proportion is identical. If so, the soft intervention could be viewed as a substitute.

Hypothesis 5. The proportion of CAs who intervene in *any* sense is independent of whether the soft intervention was available (that is, the soft intervention is a perfect substitute to hard intervention).

The contrast between treatments Once and EndoDelay allows us to investigate the role of the availability of an endogenously imposed waiting period and test hypotheses 4 and 5 directly.

3.3.2 Autonomy and deliberation

More generally, we can investigate how CA's rule-making is influenced by Choosers having additional time to contemplate their decision. Here, time is *exogenously* provided and the CA can only decide on a hard intervention. Below, we study three cases: *(i)* Choosers have no additional time (as in treatment Once); *(ii)* Choosers have one day to think about their decision (Delay); *(iii)* Choosers have to decide immediately, but they can optionally revise their decision until one day later (Undo). As above, CAs' hard intervention consists of a cap on risk-seeking behavior. First, we test whether deliberation causes CAs to "go easy" on the cap. In other words, does contemplation cause more respect for autonomy?

Hypothesis 6. The mean cap set in case *(ii)* is higher than in case *(i)*.

Second, researchers have long recognized status-quo bias in human decision-making (Samuelson and Zeckhauser, 1988). In case *(iii)*, Choosers may simply forget to correct an unfavorable initial choice. That alone would make it closer to case *(i)*. Hence, we test whether CAs recognize this possibility.

Hypothesis 7. The mean cap set in case *(ii)* is higher than in case *(iii)*.

Third, while case *(iii)* is perhaps inferior to pure deliberation, it does allow for correction.

Hypothesis 8. The mean cap set in case *(iii)* is higher than in case *(i)*.

In sum, hypotheses 6–8 predict a ranking of (mean) caps: the most restrictive caps are expected for case *(i)* and the least restrictive for case *(ii)*, with case *(iii)* in between the extremes. Our experimental treatments Once, Undo and Delay enable direct tests of these hypotheses.

3.3.3 Mental models

Our proposal stated two mechanisms that relate to how CAs think about Choosers (p. 5). We focus on the contrast between deciding now *vs.* later; for this reason, comparisons in this Section only comprise cases *(i)* and *(ii)*. First, does deliberation cause CAs to believe that Choosers come closer to CAs' subjectively preferred bliss point?

Hypothesis 9. The error[8] is lower in case *(ii)* than in case *(i)*.

Second, we elicit CAs' beliefs about Choosers' happiness.

Hypothesis 10. The mean forecast happiness (mean expected norm deviation) is higher in case *(ii)* than in case *(i)*.

These mechanism allow us to investigate how the mental models of CAs depend on Chooser deliberation.

3.4 Results

Below, we first discuss descriptive statistics from the experiment conducted with CAs through TESS. References to data sources and the analysis code are available in the Declarations. Subsequent sections present results from preregistered analyses in the Appendix and we report additional Ordinary Least Square regression models. The results of the preregistered analyses are not contradicted by these models. We also provide robustness checks and alternative specifications. All regression models use HC3 heteroskedasticity-consistent standard errors.

CAs were allowed to skip individual items in the survey, but the survey as implemented asked respondents to reconsider before submitting an empty response. We only report on CAs where data on the cap is available or, in Endo-Delay, data on the cap and the Chooser's day to decide. From an original sample size of 2,714 CAs who partook in the experiment, this wholly preregistered filtering procedure leaves data on 2,702 CAs. Attrition or selection are thus not significant concerns. However, whenever we report on any of the other three key variables (injunctive norm, belief, happiness), the sample size is slightly decreased. In 39 cases (1.44%), no data on the injunctive norm is available. In 41 cases (1.52%), no data on the belief about the average Chooser's behavior is available; in one case, no data on the average Chooser's happiness is available. Still, the vast majority of CAs gave full responses.

[8]The error is defined below (equation 3.1) as the absolute difference between a CA's injunctive norm and her beliefs about average Chooser behavior.

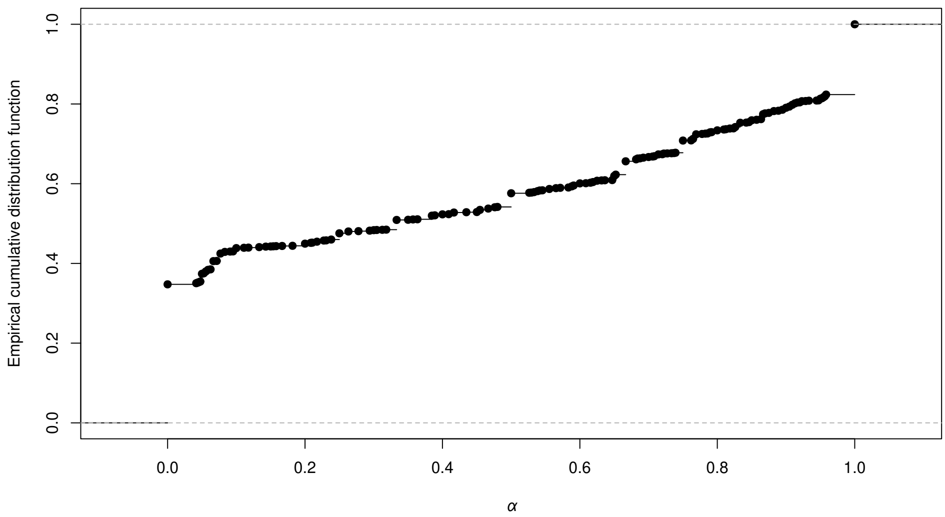


Figure 3.3: Distribution of α

This Figure is restricted to cases where α is defined and in $[0, 1]$.

3.4.1 Caps, norms and beliefs

Across all treatments, the mean cap is 17.5 boxes (std.dev.: 7.6). The mean (injunctive) norm is 9.6 boxes (std.dev.: 5.9). The mean belief about the number of boxes opened by the *average* Chooser is 11.4 (std.dev.: 5.7). 38.7% of CAs in EndoDelay imposed the 1-day waiting period.

In 92.8% of cases where the norm is available, the cap is weakly larger than the norm; in 74.8%, the cap is strictly larger. Moreover, in 77.2% of cases where both norm and belief are available, CAs believe that the average Chooser opens weakly more boxes than the CA perceives he should open. In 57.2% of cases, the inequality is strict. These descriptives suggest that the design choice to allow the implementation of an upper limit was sensible.

Result 17. CAs believe that the average Chooser opens too many boxes (relative to CAs' perceived optimal choice).

For each CA i, we can define α_i as the solution to the following equation:

$$\text{Cap}_i = \alpha_i \text{Norm}_i + (1 - \alpha_i) \cdot 25.$$

The distribution of α in the population of CAs captures how they mix their injunctive norm with imposing no restriction on the Chooser. In 89.6% of cases,

	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
Intercept	17.358^{***}	9.590^{***}	12.158^{***}	5.005^{***}	17.638^{***}	0.321^{***}
	(0.294)	(0.232)	(0.225)	(0.166)	(0.364)	(0.018)
Undo	−0.067	−0.283	-1.291^{***}	-0.743^{**}		−0.017
	(0.423)	(0.327)	(0.325)	(0.240)		(0.026)
Delay	0.329	0.252	-1.024^{***}	-0.607^{**}		0.032
	(0.409)	(0.323)	(0.306)	(0.230)		(0.026)
EndoDelay	0.333	0.016	−0.531	−0.290		0.034
	(0.412)	(0.324)	(0.315)	(0.229)		(0.026)
Waiting imposed					0.137	
					(0.598)	
Outcome	Cap	Norm	Belief	Error	Cap	Cap is 25
Subset	—	—	—	—	EndoDelay	—
R^2	0.001	0.001	0.007	0.005	0.000	0.002
Adj. R^2	−0.001	−0.000	0.006	0.003	−0.001	0.001
Num. obs.	2702	2663	2661	2649	690	2702

$^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$

Table 3.1: Outcomes by treatment

α is defined and in $[0, 1]$. Figure 3.3 shows the distribution of α in this subset. $\alpha = 0$ represents the modal choice (reflecting no cap), attained in about one third of cases, and $\alpha = 1$ is the second most common α. The median α is 0.333 and the mean is 0.42. In sum, while about two thirds of CAs intervene, Choosers are generally put on a long leash; this replicates a previous result from the literature.

Result 18. As in Chapter 1, CAs tend to leave room for Choosers to express themselves.

For the analysis below, we define one additional preregistered variable that relates to beliefs about Chooser behavior and the injunctive norm:

$$\text{Error}_i = |\text{Belief}_i - \text{Norm}_i| \,. \tag{3.1}$$

This variable captures the absolute expected norm deviation of the *average* Chooser. It is bounded from below by 0. The mean error is 4.6 (std.dev.: 4.2). Note that the notion of an "error" here entirely reflects the subjective judgment of CAs. In the Appendix, we also use $\mathbb{1}\,[\text{Error}_i = 0]$ as an outcome variable.

3.4.2 Hard paternalism and deliberation

Figure 3.4 suggests that the distribution of caps is not significantly different between two key treatments. This impression is confirmed for all treatments by model 1 in Table 3.1 as well as the preregistered analyses in Sections C.2.1,

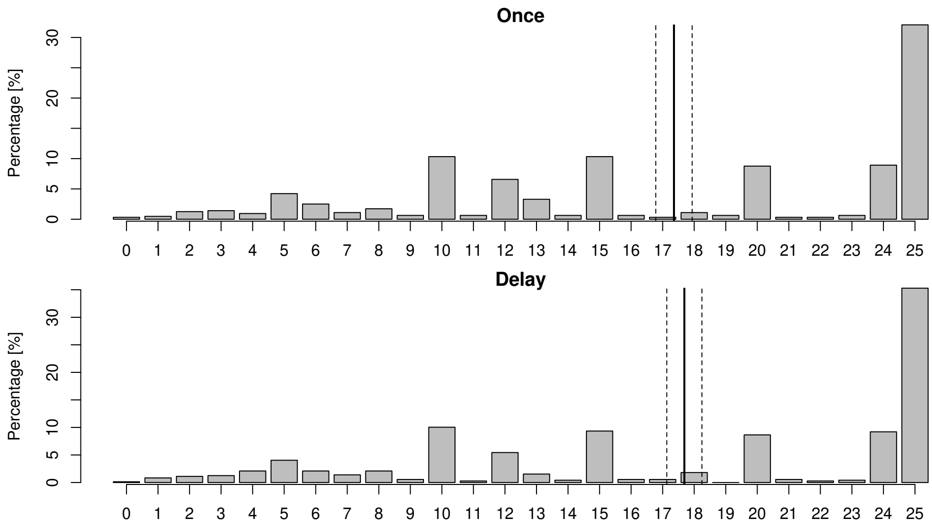


Figure 3.4: Caps in Once and Delay

Vertical lines show sample means. 95% confidence interval based on one-sample t-tests.

C.2.3, C.2.4 and C.2.5 in the Appendix. The standard errors on the treatment dummies suggest that, despite a substantial variation in caps, our experiment would have been able to identify differences in caps of about one box, a small degree of marginal risk. A robustness check in model 6 reveals that this null result obtains, too, for the binary outcome of whether the CA is completely libertarian. Moreover, the proportion of CAs not setting a cap does not change between Once, Undo and Delay (Section C.2.10 in the Appendix).

Result 19. Chooser deliberation does not cause a change in the cap. Hypotheses 4, 6, 7 and 8 are rejected (as null results at high power).

As stated above, we also elicited injunctive norms and beliefs. As should be expected, there is no difference in injunctive norms between the treatments (model 2). We discuss the findings on beliefs and error (models 3–4) below.

3.4.3 Waiting periods as add-on restrictions

Model 5 in Table 3.1 reveals that those CAs in the treatment EndoDelay who actually implemented the waiting period do not relax the cap. In a sense, this can be viewed to bolster result 19. Beyond that, this finding demonstrates how *even the conscious selection of a 1-day delay does not substitute for hard imposition.*

Since we cannot detect a substitutional relationship between these two policies, waiting periods represent an "add-on restriction" in our setting: they are an additional rule placed onto Choosers. The preregistered analysis in Section C.2.2 in the Appendix confirms this result as highly statistically significant ($p < 0.001$). Simply put, the only difference between Once and EndoDelay is that the latter also allows imposing the waiting period. 67.9% of CAs in Once imposed a cap of strictly less than 25. Almost exactly the same proportion, 64.5% of CAs in EndoDelay, imposed a cap of strictly less than 25. 38.7% of CAs in EndoDelay imposed the waiting period, with 25.2% implementing both the waiting period and a cap, while 13.5% *only* implemented the waiting period. 39.3% *only* implemented the cap. In sum, thus, 78.0% of CAs in EndoDelay intervened in *any* way. Since the cap does not give way to the waiting period, the possibility of imposing the waiting period merely increases the number of CAs that intervene at all, and represents a policy dimension that is not equivalent to the cap.

Result 20. Waiting periods are add-on restrictions. Hypothesis 5 is rejected.

Moreover, the preregistered analyses in Section C.2.9 in the Appendix suggest that CAs in EndoDelay, conditional on their choice between Once and Delay for the Chooser, behave similarly to CAs exogenously assigned to a Chooser in Once or Delay.

3.4.4 Mental models of deliberation

As suggested by Figure 3.5 and models 3–4 in Table 3.1, beliefs about average behavior and average errors (Equation 3.1) are causally reduced by deliberation, a result confirmed by the preregistered analysis in Section C.2.7 in the Appendix ($p = 0.009$). Section C.2.11 in the Appendix discusses issues of multiple hypothesis testing. Furthermore, model 1 in Table C.4 in the Appendix replicates this result for the binary outcome of the error being nil. Except for one particularly conservative adjustment of p-values, deliberation does appear to reduce errors. As expected, the injunctive norm is unchanged across treatments (model 2 in Table 3.1 and Section C.2.6 in the Appendix).

Result 21. CAs believe that Choosers, with deliberation, come closer to CAs' subjectively preferred choice. Hypothesis 9 (mechanism 1) is confirmed.

We now conduct exploratory analyses to distinguish mechanisms. We noted above that most CAs believe that the average Chooser opens too many boxes

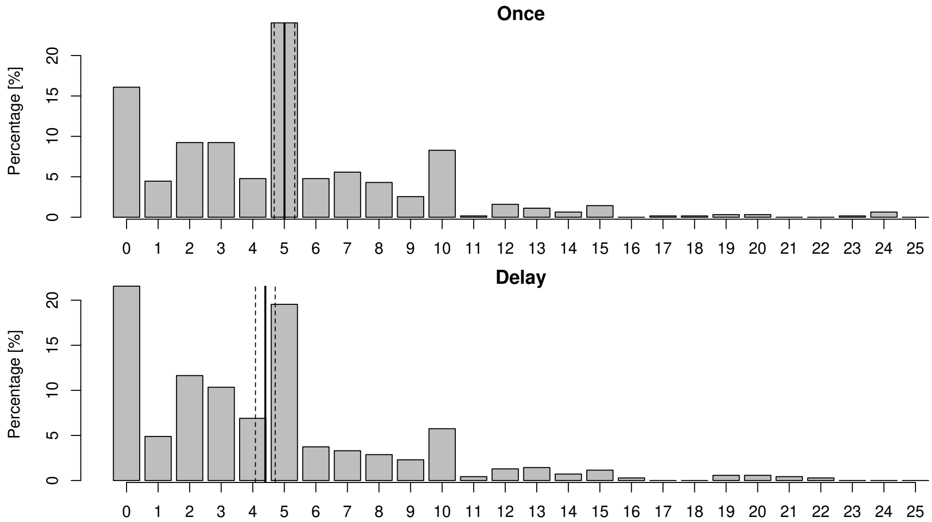


Figure 3.5: Expected error in Once and Delay

Vertical lines show sample means. 95% confidence interval based on one-sample *t*-tests.

(result 17). A pattern of reduced error can arise through at least two mechanisms by which CAs may become more optimistic by Choosers' deliberation: In what may be called "story 1," CAs believe that Choosers get closer to CAs' subjectively preferred bliss point. Here, no matter the level of the injunctive norm, deliberation causes the absolute distance between beliefs and the norm to decrease. Under deliberation, this hypothesis implies an *increase* in beliefs along with the norm level. It is easy to see why: If CAs generally believe that Choosers come closer to CAs' norms, Choosers should be believed to open more boxes under deliberation if the CA has a high norm level, but fewer boxes if the CA has a low norm level. Story 1 implies that norm levels moderate the effect of deliberation on beliefs, and an absence of moderation of the effect on errors.

On the other hand, CAs could believe that deliberation causes fewer boxes to be opened independent of the norm level. The reduction in error may be driven by CA's moderate norm levels and belief that Choosers open too many boxes. This leads to a statistical reduction in the forecast error, although CAs with high norm levels do not predict Choosers to open more boxes, given more time. We call this "story 2." Here, errors actually increase for CAs with a high injunctive norm. Story 2 implies that norm levels do not moderate the effect of deliberation on beliefs, and that there is such moderation of the effect on errors.

	Model 1	Model 2	Model 3	Model 4
Intercept	11.883^{***}	7.512^{***}	4.854^{***}	4.167^{***}
	(0.158)	(0.291)	(0.114)	(0.269)
Delay or Undo	-0.877^{***}	-0.926^{*}	-0.521^{**}	-1.309^{***}
	(0.221)	(0.410)	(0.164)	(0.381)
Norm		0.456^{***}		0.072^{**}
		(0.030)		(0.028)
Delay or Undo × Norm		0.006		0.082^{*}
		(0.042)		(0.039)
Outcome	Belief	Belief	Error	Error
R^2	0.006	0.229	0.004	0.032
Adj. R^2	0.005	0.228	0.003	0.031
Num. obs.	2661	2649	2649	2649

$^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$

Table 3.2: Heterogeneous treatment effects

To investigate this issue further, we fitted the models in Table 3.2. This Table groups the treatments Delay and Undo, and Once and EndoDelay, to increase power for the estimation of heterogeneous treatment effects.[9] In all models, the addition of CAs' injunctive norm adds explanatory power, a kind of false consensus effect (Ross, Greene, and House, 1977). The transition from model 1 to model 2 reveals no heterogeneity of treatment effects by norm on beliefs, despite extraordinarily high power. This implies that CAs, no matter their injunctive norm, forecast that deliberation simply causes a decrease in the number of boxes opened.

Models 3 and 4 provide further evidence for story 2. While CAs generally forecast a reduction in error that is caused by deliberation, that reduction is moderated by the norm ($p = 0.038$). *Ceteris paribus*, the higher a CA's injunctive norm, the lower the reduction in error through deliberation. Given the small treatment effects, the effect of deliberation on errors is strongly attenuated for plausible norm levels. A knife-edge result is that beyond a norm of about 16, deliberation even causes an *increase* in error. These findings suggest that CAs generally estimate that deliberation causes a reduction in the number of boxes opened by Choosers—and because most CAs believe that too many boxes would be opened without deliberation (Section 3.4.1), the treatment effect also

[9]Recall that Once and EndoDelay had an identical page 1 in the survey—where the norm and beliefs were elicited—and Delay and Undo share the crucial feature of Chooser deliberation. Table C.3 in the Appendix gives results for all treatments.

represents a reduction in the error, as in story 2. Model 2 in Table C.2 in the Appendix replicates this result for the binary outcome of the error being nil, although model 2 in Table C.4 in the Appendix shows that this result does not obtain without treatment grouping.

Result 22. Through deliberation, Choosers are thought to come closer to CAs' injunctive norm because deliberation is believed to reduce the number of boxes opened and CAs tend to believe that the average Chooser opens too many boxes (result 19).

Section C.2.8 in the Appendix reveals that forecast Chooser happiness is not significantly different between Once and Delay. We do not report these results in the main text.

3.5 Discussion and conclusion

We now face a puzzling empirical picture. CAs do not react to Choosers' deliberation when deciding upon limits to risk-seeking behavior. Similarly, they do not substitute between soft and hard paternalistic rules—but they believe that Choosers' behavior is changed by deliberation.

Note that our experimental design is silent on whether those CAs who *implemented* the waiting period are more optimistic about changes in behavior. However, we can use the results from treatments Once, Delay and Undo to form a prior expectation. Deliberation causes a shift in the behavior of the *average* Chooser, yet hard interventions remain unchanged. Most CAs believe that Choosers open too many boxes; and caps on Choosers' behavior are generally above beliefs about average behavior. These findings are consistent with a theory that soft and hard interventions in general *viz.* waiting periods and limits in particular aim to target different parts of the population. Soft interventions may target Choosers that are misguided but within a close-to-normal range of behavior. Hard interventions may target "extreme" Choosers. This would explain the simultaneous use of waiting periods and the cap and their non-substitutability.

To our knowledge, this point has never been explicitly recognized in the literature on paternalism, although effects that are heterogeneous with respect to the targeted outcome itself have been described. One example concerns the effect of alcohol taxes on drinkers' consumption depending on where they lie in the distribution of drinkers (Manning, Blumberg, and Moulton, 1995). Similarly, researchers have recently shown how heterogeneous decision-makers can be targeted if the effects of some policy on their individual behavior can be estimated (Lipman, 2024; Lipman et al., 2024; Opitz et al., 2024). The behavioral theory behind policy targeting in the context of the present work is simple: waiting may be thought to avert smaller mistakes, but tougher restrictions are required to prevent larger norm deviations. Future work can investigate this idea more thoroughly.

Recent work by Allcott et al. (2022) has emphasized that non-standard policies like nudges can enhance behavioral distortions. Given its limited scope, the experiment discussed here does not reveal much about the motives of those CAs in EndoDelay who chose to implement the waiting period (or not). It should be noted that CAs could not save Choosers time—as the survey always took exactly two days—and yet, a majority of CAs did not choose to implement the waiting period. Were they pessimistic about waiting periods' effects? What downsides do waiting periods have? Future work can shed light on the costs that CAs perceive for various kinds of rules.

The future will tell what happens with real-life mandatory waiting periods. The non-invasive, choice-preserving nature of waiting periods has historically been used as arguments in their favor. In *Planned Parenthood v. Casey* (1992), the U.S. Supreme Court established that no "undue burden" could be placed on abortion access, and waiting periods generally passed this test. Kalmanson (2016, p. 380) discusses the "undue burden" jurisprudence with respect to firearms regulation and restrictions on abortion. However, now that *Roe v. Wade* (1973) and *Casey* have been overturned, states can ban abortion outright without using roundabout policies that passed the "undue burden" test. In that case, hard restrictions may substitute for softer waiting periods.

While our results give some suggestions as to the mental models of CAs, we can only speculate as to the motivations of CAs in deciding between various policy instruments. If we are correct that softly paternalistic measures do not substitute for hard paternalism, this hints at an increase in the total regulatory burden caused by innovations in policy instruments—an ominous portend. Policymakers can simultaneously apply highly variegated rules to target specific parts of society even within the same policy area. Future research may investigate how this affects the efficacy of government action and to what extent policymakers correctly perceive the effects of these more complex policy bundles.

Bibliography

Ackfeld, V. and Ockenfels, A. (2021). “Do people intervene to make others behave prosocially?” *Games and Economic Behavior*, 128, 58–72.

Acquisti, A. and Grossklags, J. (2005). “Privacy and Rationality in Individual Decision Making”. *IEEE Security & Privacy*, 3(1), 26–33.

Allcott, H., Cohen, D., Morrison, W., and Taubinsky, D. (2022). “When do ”Nudges” Increase Welfare?” National Bureau of Economic Research.

Ambuehl, S. (2024). “An experimental test of whether financial incentives constitute undue inducement in decision-making”. *Nature Human Behaviour*, 1–11.

Ambuehl, S. and Bernheim, B. D. (2024). “Interpreting the will of the people: a positive analysis of ordinal preference aggregation”.

Ambuehl, S., Bernheim, B. D., and Ockenfels, A. (2021). “What motivates paternalism? An experimental study”. *American Economic Review*, 111(3), 787–830.

Ambuehl, S., Niederle, M., and Roth, A. E. (2015). “More money, more problems? Can high pay be coercive and repugnant?” *American Economic Review*, 105(5), 357–360.

Arel-Bundock, V., Greifer, N., and Heiss, A. (2024). “How to Interpret Statistical Models Using marginaleffects for R and Python”. *Journal of Statistical Software*, 111(9), 1–32. DOI: `10.18637/jss.v111.i09`.

Arneson, R. J. (1980). “Mill versus paternalism”. *Ethics*, 90(4), 470–489.

Barnes, S. B. (2006). “A privacy paradox: Social networking in the United States”. *First Monday*.

Barth, S. and De Jong, M. D. (2017). "The privacy paradox–Investigating discrepancies between expressed privacy concerns and actual online behavior–A systematic literature review". *Telematics and Informatics*, 34(7), 1038–1058.

Bartling, B., Cappelen, A. W., Hermes, H., Skivenes, M., and Tungodden, B. (2023). "Free to Fail? Paternalistic Preferences in the United States".

Becker, G. S. (1957). *The Economics of Discrimination*. University of Chicago Press.

Benjamin, D. J., Fontana, M. A., and Kimball, M. S. (2020). *Reconsidering Risk Aversion*. Working Paper 28007. National Bureau of Economic Research. DOI: 10.3386/w28007.

Berlin, I. (1958). *Two Concepts of Liberty*. Clarendon Press.

Bernheim, B. D. (2016). "The good, the bad, and the ugly: A unified approach to behavioral welfare economics". *Journal of Benefit-Cost Analysis*, 7(1), 12–68.

Binmore, K. (2009). "Interpersonal Comparison of Utility". In: *The Oxford Handbook of Philosophy of Economics*. Oxford University Press.

Black, D. (1948). "On the rationale of group decision-making". *Journal of Political Economy*, 56(1), 23–34.

Blackwell, D. (1953). "Equivalent comparisons of experiments". *The Annals of Mathematical Statistics*, 265–272.

Bohren, J. A., Haggag, K., Imas, A., and Pope, D. G. (2023). "Inaccurate Statistical Discrimination: An Identification Problem". *The Review of Economics and Statistics*, 1–45. DOI: 10.1162/rest_a_01367.

Bolton, G. E. and Ockenfels, A. (2010). "Betrayal Aversion: Evidence from Brazil, China, Oman, Switzerland, Turkey, and the United States: Comment". *American Economic Review*, 100(1), 628–633. ISSN: 0002-8282. DOI: 10.1257/aer.100.1.628.

Bolton, G. E., Ockenfels, A., and Stauf, J. (2015). "Social Responsibility Promotes Conservative Risk Behavior". *European Economic Review*, 74, 109–127. ISSN: 0014-2921. DOI: 10.1016/j.euroecorev.2014.10.002.

Bouma, J., Verbraak, M., Dietz, F., and Brouwer, R. (2019). "Policy mix: mess or merit?" *Journal of Environmental Economics and Policy*, 8(1), 32–47.

Buchmann, N., Meyer, C., and Sullivan, C. D. (2024). “Paternalistic Discrimination”.

Buckle, G. E. and Luhan, W. J. (2023). “Do as I Do: Paternalism and Preference Differences in Decision-Making for Others”. Univ. of Portsmouth.

Bursztyn, L., González, A. L., and Yanagizawa-Drott, D. (2020). “Misperceived Social Norms: Women Working Outside the Home in Saudi Arabia”. *American Economic Review*, 110(10), 2997–3029.

Bursztyn, L. and Yang, D. Y. (2022). “Misperceptions about others”. *Annual Review of Economics*, 14(1), 425–452.

Bushong, B. and Gagnon-Bartsch, T. (2024). “Failures in Forecasting: An Experiment on Interpersonal Projection Bias”. *Management Science*.

C. S. Mott Children’s Hospital (2018). *Tensions over Teen Tattoos*. Mott Poll Report. National Poll on Children’s Health, University of Michigan. URL: `https://mottpoll.org/sites/default/files/documents/082018_Tattoos_0.pdf`.

Camerer, C., Issacharoff, S., Loewenstein, G., O’Donoghue, T., and Rabin, M. (2003). “Regulation for Conservatives: Behavioral Economics and the Case for ”Asymmetric Paternalism””. *University of Pennsylvania Law Review*, 151(3), 1211–1254.

Camerer, C. and Weber, M. (1992). “Recent developments in modeling preferences: Uncertainty and ambiguity”. *Journal of Risk and Uncertainty*, 5, 325–370.

Caplan, B. (2011). *A.B.A.: Always Be Advising*. URL: `https://www.econlib.org/archives/2011/07/aba_always_be_a.html`.

Caplin, A. (2016). “Measuring and modeling attention”. *Annual Review of Economics*, 8, 379–403.

Caplin, A. and Martin, D. (2016). “The dual-process drift diffusion model: Evidence from response times”. *Economic Inquiry*, 54(2), 1274–1282.

Carter, I. (2022). “Positive and Negative Liberty”. In: *The Stanford Encyclopedia of Philosophy*. Ed. by E. N. Zalta. Spring 2022. Metaphysics Research Lab, Stanford University.

Charness, G., Gneezy, U., and Kuhn, M. A. (2012). “Experimental methods: Between-subject and within-subject design”. *Journal of Economic Behavior & Organization*, 81(1), 1–8.

Chen, D. L., Schonger, M., and Wickens, C. (2016). “oTree—An open-source platform for laboratory, online, and field experiments”. *Journal of Behavioral and Experimental Finance*, 9, 88–97.

Chew, S. H., Miao, B., and Zhong, S. (2017). “Partial ambiguity”. *Econometrica*, 85(4), 1239–1260.

Chipman, J. S. (1963). “Stochastic choice and subjective probability”. *Decisions, Values and Groups*, 1, 70–95.

Clemens, M. A. (2018). “Testing for repugnance in economic transactions: Evidence from guest work in the Gulf”. *The Journal of Legal Studies*, 47(S1), S5–S44.

Conly, S. (2012). *Against Autonomy: Justifying Coercive Paternalism*. Cambridge University Press. ISBN: 9781139176101. DOI: `10.1017/CBO9781139176101`.

Corbett, B., Feeney, A., and McCormack, T. (2021). “Interpersonal regret and prosocial risk taking in children”. *Cognitive Development*, 58, 101036.

Cowen, T. (2023). *GOAT: Who is the Greatest Economist of all Time, and Why Does it Matter?* URL: `https://econgoat.ai`.

Crosetto, P. and Filippin, A. (2013). “The ”bomb” risk elicitation task”. *Journal of Risk and Uncertainty*, 47(1), 31–65.

De Quidt, J., Haushofer, J., and Roth, C. (2018). “Measuring and bounding experimenter demand”. *American Economic Review*, 108(11), 3266–3302.

Derksen, M. and Morawski, J. (2022). “Kinds of Replication: Examining the Meanings of ”Conceptual Replication” and ”Direct Replication””. *Perspectives on Psychological Science*, 17(5), 1490–1505.

Divine, G. W., Norton, H. J., Barón, A. E., and Juarez-Colunga, E. (2018). “The Wilcoxon–Mann–Whitney procedure fails as a test of medians”. *The American Statistician*, 72(3), 278–286.

Downs, A. (1957). “An economic theory of democracy”. *Harper and Row*, 28.

Dworkin, G. (1972). “Paternalism”. *the Monist*, 64–84.

Dworkin, G. (2020). “Paternalism”. In: *The Stanford Encyclopedia of Philosophy*. Ed. by E. N. Zalta. Fall 2020. Metaphysics Research Lab, Stanford University.

Edwards, G., Nesson, E., Robinson, J. J., and Vars, F. (2018). “Looking down the barrel of a loaded gun: The effect of mandatory handgun purchase delays on homicide and suicide”. *The Economic Journal*, 128(616), 3117–3140.

Falahrastegar, M., Haddadi, H., Uhlig, S., and Mortier, R. (2014). *Anatomy of the Third-Party Web Tracking Ecosystem*. URL: `https://arxiv.org/abs/1409.1066`.

Fehr, E. and Charness, G. (2024). “Social Preferences: Fundamental Characteristics and Economic Consequences”. *Journal of Economic Literature*. Forthcoming.

Feinberg, J. (1989). *Harm to self*. Oxford University Press.

Feri, F., Giannetti, C., and Jentzsch, N. (2016). “Disclosure of personal information under risk of privacy shocks”. *Journal of Economic Behavior & Organization*, 123, 138–148.

Friedman, M. (1953). “The Methodology of Positive Economics”. In: *Essays in Positive Economics*. University of Chicago Press. Chap. 1.

Frik, A. and Gaudeul, A. (2020). “A measure of the implicit value of privacy under risk”. *Journal of Consumer Marketing*, 37(4), 457–472.

Fudenberg, D., Strack, P., and Strzalecki, T. (2018). “Speed, accuracy, and the optimal timing of choices”. *American Economic Review*, 108(12), 3651–3684.

Garimella, K., Kostakis, O., and Mathioudakis, M. (2017). “Ad-blocking: A study on performance, privacy and counter-measures”. In: *Proceedings of the 2017 ACM on Web Science Conference*, 259–262.

Gigliotti, G. (1996). “The testing principle: Inductive reasoning and the Ellsberg paradox”. *Thinking & Reasoning*, 2(1), 33–49.

Gilbert, D. T. and Ebert, J. E. (2002). “Decisions and revisions: the affective forecasting of changeable outcomes”. *Journal of Personality and Social Psychology*, 82(4), 503.

Gillen, B., Snowberg, E., and Yariv, L. (2019). “Experimenting with measurement error: Techniques with applications to the caltech cohort study”. *Journal of Political Economy*, 127(4), 1826–1863.

Greiner, B. (2015). “Subject pool recruitment procedures: organizing experiments with ORSEE”. *Journal of the Economic Science Association*, 1(1), 114–125.

Grossman, S. J. (1981). “The informational role of warranties and private disclosure about product quality”. *The Journal of Law and Economics*, 24(3), 461–483.

Grossmann, M. R. P. (2023). “Paternalism and Deliberation”. Proposal. Time-sharing Experiments for the Social Sciences. URL: `https://gitlab.com/gr0ssmann/undo/-/raw/master/tess/proposal_revised2.pdf`.

— (2024). *Knowledge and Freedom: Evidence on the Relationship Between Information and Paternalism*. arXiv preprint. URL: `https://arxiv.org/abs/2410.20970`.

— (2025). *Paternalism and Deliberation: An Experiment on Making Formal Rules*. arXiv preprint. URL: `https://arxiv.org/abs/2501.00863`.

Grossmann, M. R. P. and Gerhardt, H. (n.d.). “uproot”.

Grossmann, M. R. P. and Ockenfels, A. (2024). “Paternalism in Data Sharing”. Mimeo.

Hanna, J. (2018). “Hard and soft paternalism”. In: *The Routledge Handbook of the Philosophy of Paternalism*. Routledge, 24–34.

Harborth, D. and Pape, S. (2018). “German translation of the Concerns for Information Privacy (CFIP) construct”. *Available at SSRN 3112207*.

Harrison, G. W., Hofmeyr, A., Ross, D., and Swarthout, J. T. (2018). “Risk preferences, time preferences, and smoking behavior”. *Southern Economic Journal*, 85(2), 313–348.

Hausman, D. M. (1995). “The Impossibility of Interpersonal Utility Comparisons”. *Mind*, 104(415), 473–490. DOI: `10.1093/mind/104.415.473`.

Hayek, F. A. von (1944). *The Road to Serfdom*. Chicago, IL: University of Chicago Press.

— (1945). “The Use of Knowledge in Society”. *The American Economic Review*, 35(4), 519–530.

— (1960). *The Constitution of Liberty*. Chicago, IL: University of Chicago Press.

Hayek, F. A. von (1988). *The Fatal Conceit: The Errors of Socialism*. Ed. by William W. Bartley III. The Collected Works of F. A. Hayek. Chicago, IL: University of Chicago Press.

Head, J. G. (1966). "On Merit Goods". *FinanzArchiv/Public Finance Analysis*, 25 (1), 1–29.

Hermstrüwer, Y. and Dickert, S. (2017). "Sharing is daring: An experiment on consent, chilling effects and a salient privacy nudge". *International Review of Law and Economics*, 51, 38–49.

Holm, S. (1979). "A simple sequentially rejective multiple test procedure". *Scandinavian Journal of Statistics*, 65–70.

Hossain, T. and Okui, R. (2013). "The binarized scoring rule". *Review of Economic Studies*, 80(3), 984–1001.

Imas, A., Kuhn, M. A., and Mironova, V. (2022). "Waiting to choose: The role of deliberation in intertemporal choice". *American Economic Journal: Microeconomics*, 14(3), 414–440.

Joyce, T. J., Henshaw, S. K., Dennis, A., Finer, L. B., and Blanchard, K. (2009). "The impact of state mandatory counseling and waiting period laws on abortion: A literature review". Guttmacher Institute.

Kagel, J. H. and Roth, A. E. (1995). *The Handbook of Experimental Economics*. Vol. 1. Princeton University Press. ISBN: 9780691058979.

Kalmanson, M. (2016). "The Second Amendment Burden: Arming Courts with a Workable Standard for Reviewing Gun Safety Legislation". *Fla. St. UL Rev.*, 44, 347.

Kartik, N. (2009). "Strategic communication with lying costs". *The Review of Economic Studies*, 76(4), 1359–1395.

Kearns, M. and Roth, A. (2019). *The ethical algorithm: The science of socially aware algorithm design*. Oxford University Press.

Khalmetski, K. and Ockenfels, A. (2024). "Why people ban others' actions: A norm conformism hypothesis". Mimeo.

Kiessling, L., Chowdhury, S. K., Schildberg-Hörisch, H., and Sutter, M. (2021). "Parental paternalism and patience". DICE Discussion Paper № 358. URL: `https://hdl.handle.net/10419/228523`.

Kirchgässner, G. (2017). “Soft paternalism, merit goods, and normative individualism”. *European Journal of Law and Economics*, 43, 125–152.

Kocher, M. G. and Sutter, M. (2006). “Time is money—Time pressure, incentives, and the quality of decision-making”. *Journal of Economic Behavior & Organization*, 61(3), 375–392.

Kokolakis, S. (2017). “Privacy attitudes and privacy behaviour: A review of current research on the privacy paradox phenomenon”. *Computers & Security*, 64, 122–134.

Kolm, S.-C. (1993). “The Impossibility of Utilitarianism”. In: *The Good and the Economical: Ethical Choices in Economics and Management*. Springer, 30–69.

Konrad, K. A. (2024). “The political economy of paternalism”. *Public Choice*, 201, 61–81.

Lee, J. (2013). “The impact of a mandatory cooling-off period on divorce”. *The Journal of Law and Economics*, 56(1), 227–243.

Lindo, J. M. and Pineda-Torres, M. (2021). “New evidence on the effects of mandatory waiting periods for abortion”. *Journal of Health Economics*, 80, 102533.

Lipman, S. A. (2024). “One size fits all? Designing financial incentives tailored to individual economic preferences”. *Behavioural Public Policy*, 8(2), 264–278.

Lipman, S. A., Sukurica, L., Alvarez-Colic, N., and Tholen, L. (2024). *Tailoring through choice: Comparing the effect of randomly assigned vs. self-chosen behavioural interventions in promoting healthier snack choice*. EsCHER Working Paper 2024006. Rotterdam, The Netherlands: Erasmus Centre for Health Economics Rotterdam.

Loewenstein, G. and Haisley, E. C. (2007). “The economist as therapist: Methodological ramifications of ’light’ paternalism”. *Available at SSRN 962472*.

Loewenstein, G., O’Donoghue, T., and Bhatia, S. (2015). “Modeling the interplay between affect and deliberation”. *Decision*, 2(2), 55.

Luca, M., Malhotra, D., and Poliquin, C. (2017). “Handgun waiting periods reduce gun deaths”. *Proceedings of the National Academy of Sciences*, 114(46), 12162–12165.

Mabsout, R. (2022). “John Stuart Mill, soft paternalist”. *Social Choice and Welfare*, 58(1), 161–186.

Manning, W. G., Blumberg, L., and Moulton, L. H. (1995). “The demand for alcohol: the differential response to price”. *Journal of Health Economics*, 14(2), 123–148.

Manski, C. F. and Sheshinski, E. (2023). *Optimal Paternalism in a Population with Bounded Rationality*. Working Paper 31349. National Bureau of Economic Research. URL: `https://www.nber.org/papers/w31349`.

Metcalfe, J. and Mischel, W. (1999). “A hot/cool-system analysis of delay of gratification: Dynamics of willpower.” *Psychological Review*, 106(1), 3.

Milgrom, P. R. (1981). “Good news and bad news: Representation theorems and applications”. *The Bell Journal of Economics*, 380–391.

Mill, J. S. (1869). *On Liberty*. 4th ed. London: Longmans, Green, Reader and Dyer. URL: `https://en.wikisource.org/wiki/On_Liberty`.

Musgrave, R. A. (1956). “A Multiple Theory of Budget Determination”. *Finanz-Archiv/Public Finance Analysis*, 17 (3), 333–343.

— (1959). *The Theory of Public Finance—A Study in Public Economy*. McGraw-Hill.

Mutz, D. C. (2011). *Population-Based Survey Experiments*. Princeton University Press.

Nielsen, K. and Rehbeck, J. (2022). “When Choices Are Mistakes”. *American Economic Review*, 112(7), 2237–2268.

Oechssler, J., Roider, A., and Schmitz, P. W. (2015). “Cooling off in negotiations: Does it work?” *Journal of Institutional and Theoretical Economics (JITE)/Zeitschrift für die gesamte Staatswissenschaft*, 565–588.

OpenAI et al. (2024). *GPT-4 Technical Report*. URL: `https://arxiv.org/abs/2303.08774`.

Opitz, S., Sliwka, D., Vogelsang, T., and Zimmermann, T. (2024). “The Algorithmic Assignment of Incentive Schemes”. *Management Science*.

Page, E. B. (1963). “Ordered Hypotheses for Multiple Treatments: A Significance Test for Linear Ranks”. *Journal of the American Statistical Association*, 58(301), 216–230. DOI: 10.1080/01621459.1963.10500843.

Pareto, V. (2014). *Manual of Political Economy: A Critical and Variorum Edition*. Oxford University Press.

Polman, E. and Wu, K. (2020). “Decision making for others involving risk: A review and meta-analysis”. *Journal of Economic Psychology*, 77, 102184.

Price, P. C. (1998). “Effects of a relative-frequency elicitation question on likelihood judgment accuracy: The case of external correspondence”. *Organizational Behavior and Human Decision Processes*, 76(3), 277–297.

Rekaiti, P. and Van den Bergh, R. (2000). “Cooling-off periods in the consumer laws of the EC Member States. A comparative law and economics approach”. *Journal of Consumer Policy*, 23(4), 371–408.

Rizzo, M. J. and Whitman, D. G. (2009). “Little brother is watching you: New paternalism on the slippery slopes”. *Arizona Law Review*, 51, 685.

Rockenbach, B., Sadrieh, A., and Schielke, A. (2021). “Paying with your personal data: the insensitivity of private information provision to asymmetric benefits”. *Journal of the Economic Science Association*, 7(1), 64–73.

Ross, L., Greene, D., and House, P. (1977). “The ”false consensus effect”: An egocentric bias in social perception and attribution processes”. *Journal of Experimental Social Psychology*, 13(3), 279–301.

Roth, A. E. (1986). “Laboratory experimentation in economics”. *Economics & Philosophy*, 2(2), 245–273.

— (2007). “Repugnance as a Constraint on Markets”. *Journal of Economic Perspectives*, 21(3), 37–58.

Rowlands, S. and Thomas, K. (2020). “Mandatory waiting periods before abortion and sterilization: theory and practice”. *International Journal of Women’s Health*, 577–586.

Samuelson, W. and Zeckhauser, R. (1988). “Status quo bias in decision making”. *Journal of Risk and Uncertainty*, 1, 7–59.

Schlag, K. H., Tremewan, J., and Van der Weele, J. J. (2015). “A penny for your thoughts: A survey of methods for eliciting beliefs”. *Experimental Economics*, 18(3), 457–490.

Schnellenbach, J. (2012). “Nudges and norms: On the political economy of soft paternalism”. *European Journal of Political Economy*, 28(2), 266–277.

Schwerter, F. (2024). “Social Reference Points and Risk Taking”. *Management Science*, 70(1), 616–632.

Scoccia, D. (2018). “The concept of paternalism”. In: *The Routledge Handbook of the Philosophy of Paternalism*. Routledge, 11–23.

Sher, B. D. (1967). “The cooling-off period in door-to-door sales”. *UCLA L. Rev.*, 15, 717.

Smith, H. J., Milberg, S. J., and Burke, S. J. (1996). “Information privacy: Measuring individuals’ concerns about organizational practices”. *MIS Quarterly*, 167–196.

Snowdon, C. (2017). *Killjoys. A Critique of Paternalism*. The Institute of Economic Affairs. ISBN: 9780255367493.

Solove, D. J. (2021). “The Myth of the Privacy Paradox”. *Geo. Wash. L. Rev.*, 89, 1.

Sovern, J. (2013). “Written Notice of Cooling-off Periods: A Forty-Year Natural Experiment in Illusory Consumer Protection and the Relative Effectiveness or Oral and Written Disclosures”. *U. Pitt. L. Rev.*, 75, 333.

Špecián, P. (2019). “The Precarious Case of the True Preferences”. *Society*, 56, 267–272.

Sugden, R. (2022). “Debiasing or regularisation? Two interpretations of the concept of ’true preference’ in behavioural economics”. *Theory and Decision*, 92(3-4), 765–784.

Sunstein, C. R. (2014). *Why Nudge?: The Politics of Libertarian Paternalism*. Yale University Press.

Sunstein, C. R. and Thaler, R. H. (2003). “Libertarian paternalism is not an oxymoron”. *The University of Chicago Law Review*, 1159–1202.

Supreme Court of The United States (1973). *Roe v. Wade*. 410 U.S. 113. Available at: `https://www.oyez.org/cases/1971/70-18`.

Supreme Court of The United States (1992). *Planned Parenthood v. Casey*. 505 U.S. 833. Available at: https://www.oyez.org/cases/1991/91-744.

Sweeney, L. (2000). "Simple demographics often identify people uniquely". *Health (San Francisco)*, 671(2000), 1–34.

Taddicken, M. (2014). "The 'Privacy Paradox' in the Social Web: The Impact of Privacy Concerns, Individual Characteristics, and the Perceived Social Relevance on Different Forms of Self-Disclosure". *Journal of Computer-Mediated Communication*, 19(2), 248–273. DOI: 10.1111/jcc4.12052.

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

— (2008). *Nudge: Improving decisions about health, wealth, and happiness*. Penguin.

Tor, A. (2022). "The law and economics of behavioral regulation". *Review of Law & Economics*, 18(2), 223–281.

Traxler, C. and Winter, J. (2012). "Survey evidence on conditional norm enforcement". *European Journal of Political Economy*, 28(3), 390–398.

Tversky, A. and Kahneman, D. (1971). "Belief in the law of small numbers". *Psychological Bulletin*, 76(2), 105.

Viscusi, W. K. (1990). "Do Smokers Underestimate Risks?" *Journal of Political Economy*, 98(6), 1253–1269.

Wooldridge, J. M. (2010). *Econometric analysis of cross section and panel data*. MIT Press.

Note: All URLs were last accessed January 2, 2025.

Appendix A

Appendix to Chapter 1

A.1 Additional analyses

	Model 1
Intercept	-1.310^{***}
	(0.277)
$[\text{WTA} \leq 4] \cdot \text{WTA}$	0.458^{*}
	(0.178)
$\text{WTA} > 4$	0.722
	(0.623)
Model	Logistic regression
Outcome	Set minimum price
AIC	157.831
BIC	166.433
Log Likelihood	−75915
Deviance	151.831
Num. obs.	130

$^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$

Table A.1: Extensive margin and CAs' WTA

	Model 1
Intercept	0.559^{***}
	(0.164)
$\widehat{\log \sigma}$	-0.388^{***}
	(0.114)
$[\text{WTA} \leq 4] \cdot \text{WTA}$	0.255^{**}
	(0.091)
WTA > 4	2.180^{***}
	(0.346)
Model	Tobit$(-\infty, 4.01)$
Subset	Set minimum price
Outcome	Minimum price
Log Likelihood	−40.563
Num. obs.	39

$^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$

Table A.2: Explaining minimum prices with WTA

	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
Intercept	0.319^{***}	16.847^{***}	5.369^{***}	5.029^{***}	5.351^{***}	5.617^{***}
	(0.055)	(1.949)	(0.096)	(0.107)	(0.094)	(0.100)
$[\text{WTA} \leq 4] \cdot \text{WTA}$	−0.005	0.083	0.153^{*}	0.151^{*}	0.193^{**}	0.169^{**}
	(0.037)	(1.438)	(0.059)	(0.075)	(0.059)	(0.059)
WTA > 4	0.109	6.285	0.470	0.507	0.417	0.645^{**}
	(0.153)	(9.137)	(0.297)	(0.309)	(0.287)	(0.211)
Outcome	Ad-blocker use	Time on policy	CFIP–Collection	CFIP–Errors	CFIP–USU	CFIP–IA
Standard errors	HC3	HC3	HC3	HC3	HC3	HC3
R^2	0.006	0.007	0.053	0.045	0.072	0.081
Adj. R^2	−0.010	−0.009	0.039	0.030	0.057	0.066
Num. obs.	125	130	130	130	130	130

$^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$

Table A.3: CAs' WTA and other measures of privacy preferences

	Model 1	Model 2
Intercept	0.829*	−5.840
	(0.320)	(17.231)
CFIP–Collection	−0.121	−3.976
	(0.082)	(7.407)
CFIP–Errors	−0.019	0.263
	(0.068)	(2.503)
CFIP–USU	0.160	1.845
	(0.105)	(4.892)
CFIP–IA	−0.106	5.822
	(0.075)	(4.596)
Outcome	Ad-blocker use	Time on policy
Standard errors	HC3	HC3
F-test	$p = 0.146$	$p = 0.358$
R^2	0.055	0.034
Adj. R^2	0.023	0.003
Num. obs.	125	130

$^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$

Table A.4: CFIP and other constructs

	Model 1	Model 2
Intercept	0.505^{***}	0.626^{***}
	(0.053)	(0.060)
Set minimum	0.059	0.125
	(0.097)	(0.089)
$[\text{WTA} \leq 4] \cdot \text{WTA}$		-0.091^{*}
		(0.043)
WTA > 4		-0.670^{***}
		(0.060)
Standard errors	HC3	HC3
R^2	0.003	0.172
Adj. R^2	−0.005	0.153
Num. obs.	130	130

$^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$
The outcome in all regressions is a binary indicator of whether the CA sent the message "I recommend sharing" to the Chooser.

Table A.5: Recommendation behavior

A.2 Beliefs about WTA when Chooser advised against sharing data

Table A.6 contains a regression model similar to the one implied by Equation 1.2 in the main text:

$$\text{Belief} = \sum_{j=1}^{6} \beta_j D_j + \sum_{j=1}^{6} \gamma_j \Delta D_j + \epsilon, \tag{A.1}$$

where the indicator ΔD_j turns on if the belief is concerning a Chooser who was advised against sharing their data. We use cluster-robust standard errors on the CA level.

The first six coefficients are identical to Model 1 in Table 1.2 in the main text; as in that Table, γ_j can be conveniently interpreted as differences. Here, differences are between two kinds of belief, not between beliefs and actual WTAs. As is evident from Table A.6, advising a Chooser against sharing their data prompts CAs to shift their beliefs about WTA upward. This effect is most

	Model 1
WTA = 0	0.320^{***}
	(0.025)
0 < WTA ≤ 1	0.120^{***}
	(0.009)
1 < WTA ≤ 2	0.129^{***}
	(0.008)
2 < WTA ≤ 3	0.143^{***}
	(0.009)
3 < WTA ≤ 4	0.139^{***}
	(0.011)
WTA > 4	0.150^{***}
	(0.015)
Δ [WTA = 0]	-0.156^{***}
	(0.019)
Δ [0 < WTA ≤ 1]	-0.035^{***}
	(0.008)
Δ [1 < WTA ≤ 2]	-0.035^{***}
	(0.007)
Δ [2 < WTA ≤ 3]	-0.018^{*}
	(0.009)
Δ [3 < WTA ≤ 4]	0.031^{*}
	(0.013)
Δ [WTA > 4]	0.212^{***}
	(0.022)
Coefficients represent	Beliefs
Standard errors	Clustered HC3
R^2	0.578
Adj. R^2	0.575
Num. obs.	1560

$^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$

Table A.6: Beliefs about the effect of advising against sharing striking for the proportion of Choosers thought to have a WTA of above €4: in the baseline, the mean proportion allocated to that category is 0.15. When Choosers are advised against sharing data, this category is believed to comprise 36.2% of subjects—an increase of $\gamma_6 = 0.212$.

A.3 Highly variegated questions (HVQs)

1. Which gender do you identify with?
2. Are you an only child?
3. What is your field of study?
4. Which party would you vote for if the federal election were next Sunday?
5. Which religion do you identify with?
6. Which football club do you like the most?

Subjects were asked to indicate their agreement (yes/no) with the following statements:

1. The right to travel within the federal territory should be dependent on the installation of the Corona-Warn-App.
2. A general speed limit should be introduced on German highways.
3. Animal-based food products should be additionally taxed.
4. Half of the seats in the Bundestag should be reserved for women.

Please see also screen 5 in Section A.5 in this Appendix.

A.4 Beliefs about conditional means

The mean WTA cannot be calculated and its expectation may be infinite, as some CAs have WTAs in excess of €4, with the precise value unknown. Similarly, we cannot calculate the expectation of the distribution implied by beliefs.

However, recall that we can compute the mean WTA given that the WTA is at most €4 (the *conditional* WTA). The population equivalent of this statistic is a conditional expectation. Secondly, note that we elicit beliefs about the relative frequency of WTA intervals, not just a single summary statistic. We can use these beliefs to compute lower and upper bounds for the conditional expectation of each CA. Denote the belief about the proportion of WTAs falling into category i by $\tilde{x}_i$. It follows that for CA j, their conditional expectation is bounded by

$$E_j\left[r \mid r \leq 4\right] \in \left[E_j^{\text{lo}}, E_j^{\text{hi}}\right], \tag{A.2}$$

where

$$E_j^{\text{lo}} = \frac{0 \cdot \tilde{x}_1 + 0.01 \cdot \tilde{x}_2 + 1.01 \cdot \tilde{x}_3 + 2.01 \cdot \tilde{x}_4 + 3.01 \cdot \tilde{x}_5}{\sum_{j=1}^{5} \tilde{x}_j}, \tag{A.3}$$

$$E_j^{\text{hi}} = \frac{0 \cdot \tilde{x}_1 + 1 \cdot \tilde{x}_2 + 2 \cdot \tilde{x}_3 + 3 \cdot \tilde{x}_4 + 4 \cdot \tilde{x}_5}{\sum_{j=1}^{5} \tilde{x}_j}. \tag{A.4}$$

Beyond the t-test on lower bounds reported in the main text, we can test whether the lower bounds, E_j^{lo}, are significantly different from the conditional WTAs elicited in our experiment. The null hypothesis is rejected (two-sample Mann–Whitney U tests, $W = 5344$, $p < 0.001$). The same is true for upper bounds ($W = 3972$, $p < 0.001$).

A.5 Instructions

Session	oTree app sequence	Screens in this Appendix
1	PiDS_Consent, PiDS2_PRE	1–6, 18, 14, 16, 19
2	PiDS_Consent, PiDS2	1–17
3	PiDS_Consent, PiDS2_IMPL	1–6, 20, 14, 16, 21

Note: The following screenshots represent English translations of particular runs through the experiment. The experiment was originally conducted in German.

Screen 1

Screen 1 showed IRB-related information and the first consent screen.

Screen 2

Screen 2 showed the experiment's privacy policy and the final consent screen.

Screen 3

Participants waited on screen 3. Here, their ad-blocker usage was automatically ascertained.

Screen 4

General Information

Time left to complete this page: **3:54**

Welcome to this experiment.

In this experiment, you can earn money. The duration of the experiment is approximately 30 minutes.

Your decisions and your payout will be treated with **strict confidentiality**. Neither other participants nor the experimenters will know your identity.

Please give your full attention to this experiment during the session.

For better readability, we use only the masculine form of personal nouns in these instructions. Of course, we are referring to members of all genders.

What the experiment is about

Today's experiment consists of several parts. In all parts, you will make decisions or answer questions that can earn you money. Every part is relevant for the payout. Additionally, you will receive €2.50 for arriving on time.

Next

Screen 5

Part A

Time left to complete this page: **5:17**

Please answer the following questions.

You will receive a credit of €4 for answering the questions.

Which gender do you identify with?

- diverse
- female
- male

Are you an only child?

- No
- Yes

What is your field of study?

- Education
- Medicine
- Psychology
- Law
- Economics
- Social Sciences
- Business Administration
- other
- none

Which party would you vote for if the federal election were next Sunday?

- SPD
- FDP
- Die Linke
- AfD
- CDU/CSU
- Bündnis 90/Die Grünen
- other
- none

Which religion do you identify with?

- Islam
- Christianity (other)
- Buddhism
- Christianity (Protestant)
- Hinduism
- Christianity (Catholic)
- Judaism
- other
- non-religious

Which football club do you like the most?

- Real Madrid
- Borussia Mönchengladbach
- FC Bayern München
- FC Schalke 04
- Borussia Dortmund
- 1. FC Köln
- RB Leipzig
- other/none

The right to travel within the federal territory should be dependent on the installation of the Corona-Warn-App.

- I agree with this statement.
- I disagree with this statement.

A general speed limit should be introduced on German highways.

- I agree with this statement.
- I disagree with this statement.

Animal-based food products should be additionally taxed.

- I agree with this statement.
- I disagree with this statement.

Half of the seats in the Bundestag should be reserved for women.

- I agree with this statement.
- I disagree with this statement.

Next

Screen 6

Part A

Time left to complete this page: **3:57**

Please make the following decision to conclude Part A of the experiment:

You can now choose whether to donate €3 to UNICEF or not. The donated amount will be deducted from your payout. The money will then be paid by us to UNICEF to make the donation. The receipts for this will be published within the next few weeks along with the **number** of donors on our website (ockenfels.uni-koeln.de) in the "News" section.

UNICEF is the United Nations Children's Fund. The Federal Agency for Civic Education describes UNICEF as follows: "[UNICEF] is tasked with improving the protection and health of children, supporting mothers, advising on family planning, assisting with education and training issues, and providing aid in emergencies and disasters (especially to children). These tasks are primarily carried out in developing countries and in cooperation with other United Nations organizations. In 1965, UNICEF was awarded the Nobel Peace Prize."

- (●) Yes, I would like to donate €3 to UNICEF.
- ( ) No, I do not want to donate.

Next

Screen 7

Part B

Time left to complete this page: **39:51**

In the context of a previous experiment recently conducted at the Cologne Laboratory for Economic Research, participants answered the same questions as you. Like you, each participant had the opportunity to donate €3 to UNICEF.

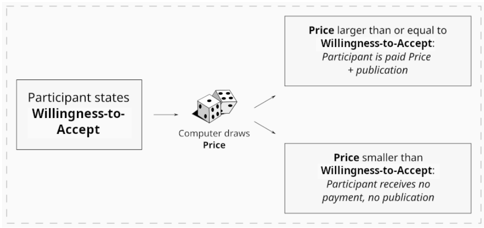
In the previous experiment, each participant then had the opportunity to earn additional money by publishing their data. By publication, we mean that the responses to the questions and the individual participant's donation decision were shared with the other participants of the previous experiment. Furthermore, these data were published on our website at ockenfels.uni-koeln.de under the "News" section after the experiment.

You can review exactly which data could be published per participant here if you wish:

Which data could be published? ▾

In the previous experiment, each participant was asked to indicate their Willingness-to-Accept in euros for the publication of their own data. This is the minimum amount the participant wanted to receive to agree to the publication of the data. Here, €0 means that the participant accepts any amount for publication. €4 means that the participant only accepts €4 or more. For each participant, the computer subsequently drew a price randomly between €0 and €4 inclusive. If this randomly drawn price exceeded or matched the participant's Willingness-to-Accept, the participant was paid the price in euros and the publication was carried out. However, if the price was less than the Willingness-to-Accept, the participant did not receive any payment for this part of the experiment and no publication took place. Please note that each participant could object to the publication by stating that their Willingness-to-Accept was higher than €4. Since the price was only drawn between €0 and €4, publication would then never have occurred. At the same time, it is valid that with a Willingness-to-Accept of €0, any randomly determined euro amount is accepted and leads to publication.

Here you see an illustration of the procedure:



Control Questions

Before you can proceed with the experiment, you must correctly answer the following control questions. If you answer all control questions correctly the first time, you will receive a bonus of €0.50.

1. Assume the participant has indicated a Willingness-to-Accept of €3. The computer randomly drew the price €3.49. How much will the participant be paid, and will a publication take place?

- ○ The participant will be paid €0.
- ○ The participant will be paid €3.
- ● The participant will be paid €3.49.
- ○ The participant will be paid €3.50.

- ○ No publication will take place.
- ● A publication will take place.
- ○ Whether a publication takes place will be randomly determined by the computer.

2. Assume the participant has indicated a Willingness-to-Accept of €0. The computer randomly drew the price €3.49. How much will the participant be paid, and will a publication take place?

- ○ The participant will be paid €0.
- ○ The participant will be paid €3.
- ● The participant will be paid €3.49.
- ○ The participant will be paid €3.50.

- ○ No publication will take place.
- ● A publication will take place.
- ○ Whether a publication takes place will be randomly determined by the computer.

3. Assume the participant has indicated a Willingness-to-Accept of €3. The computer randomly drew the price €2.49. How much will the participant be paid, and will a publication take place?

- ● The participant will be paid €0.
- ○ The participant will be paid €2.49.
- ○ The participant will be paid €3.49.
- ○ The participant will be paid €3.50.

- ● No publication will take place.
- ○ A publication will take place.
- ○ Whether a publication takes place will be randomly determined by the computer.

4. Assume the participant has stated that their Willingness-to-Accept is higher than €4.

- ○ A publication is possible.
- ● A publication is categorically excluded, regardless of the price the computer draws.

5. Is the following statement correct? **"This procedure ensures that each participant should state their 'true' Willingness-to-Accept.** Since the amount of the payment is determined by chance, the participant cannot influence it, but only whether they accept the payment or not. They will want to accept the payment exactly when it is at least as large as the true Willingness-to-Accept. The participant can guarantee this by stating the true Willingness-to-Accept for the publication of the data."

- ● This statement is correct.
- ○ This statement is incorrect.

6. The Willingness-to-Accept for publication was asked **after** answering the questions and making the donation decision; not before or simultaneously.

- ● This statement is correct.
- ○ This statement is incorrect.

Next

Screen 8

Part B

Time left to complete this page: **9:51**

In this part, you can earn up to €2.

Below are six groups listed. Each participant from the previous experiment belonged to one of these groups depending on their Willingness-to-Accept. Please estimate: What was the percentage of participants in the different groups?

Willingness-to-Accept for publication	**Your estimate of how many participants (in percent) in the previous experiment had this Willingness-to-Accept**
exactly €0	20 %
from €0.01 to €1	10 %
from €1.01 to €2	10 %
from €2.01 to €3	10 %
from €3.01 to €4	20 %
more than €4 (no publication)	30 %
Total	**100%**

Before you can submit your entries, the sum of the percentages must equal 100%.

The closer your estimate is to the actual distribution of the data, the higher the probability that you will receive a bonus of €2 in this part.

More specifically: One of the six groups mentioned above will be randomly selected. Your estimate will then be compared to the actual percentage of the group: In this part, you will receive €2 with a k% probability, where $k = 100 \cdot (1 - 0.05 \cdot (\text{Estimate} - \text{TrueValue})^2)$. The probability of winning €2 is thus higher the closer you are to the true values. It is therefore in your interest to give your best possible estimate.

Next

Screen 9

Part C

Time left to complete this page: **39:57**

For your attentive completion of this part, you will be credited a flat €3.

With a certain probability, we will randomly assign you **exactly one participant** of a **future experiment**. This participant will answer the same questions as you. Like you, this participant will also have the opportunity to donate €3 to UNICEF. Thus, this future participant will undergo the same experiment as you and the participants of the previous experiment.

In the context of this future experiment, the participant will also have the opportunity to earn additional money by publishing the data. By publication, we mean that the responses to the questions and the donation decision of the individual participant will be shared with the other participants of the future experiment. Additionally, these data will be published after the experiment on our website (ockenfels.uni-koeln.de) in the "News" section. **The decision situation is identical to that in the previous experiment (from Part B). To recall:**

> **In the context of the future experiment, each participant will be asked to state their Willingness-to-Accept in euros for the publication of their own data.** This is the amount the participant wants to receive *at least* to agree to the publication of the data. Here, €0 means that the participant accepts any amount for publication. €4 means that the participant only accepts €4 or more. For each participant, the computer will subsequently draw a price between €0 and €4 inclusively. If this randomly drawn price exceeds or matches the participant's Willingness-to-Accept, the price in euros will be paid out to the participant and the publication will be made. However, if the price is less than the Willingness-to-Accept, the participant will not receive a payout for this part of the experiment and no publication will take place. Please note that each participant can object to the publication by stating that their Willingness-to-Accept is higher than €4. Since the price is only drawn between €0 and €4, there would then never be a publication. At the same time, it is valid that with a Willingness-to-Accept of €0, any randomly determined euro amount is accepted and leads to publication.

As you can see, this is essentially **the same decision** as that of the participant in the previous experiment (Part B).

Which data per participant can exactly be published, you can review here again if you wish:

Which data can be published? ▾

You now have the opportunity to set a **minimum price** and/or a **maximum price** for the publication for **a future participant**. A minimum price ensures that the offered price for publication does not fall below the minimum price, and the maximum price ensures that the offered price does not exceed the maximum price for publication.

In detail, the **minimum price** works as follows: The publication will only be made if the price drawn by the computer **exceeds both your minimum price and the future participant's Willingness-to-Accept or is identical to it.** In this case, the publication is made and the future participant receives the price as a payout.
The **maximum price** works as follows: The publication will only be made if the price drawn by the computer **exceeds the Willingness-to-Accept or is identical to it and is less than the maximum price**. In this case, the publication is made and the future participant receives the price as a payout.

You can decide in the next step whether you want to set a minimum price, a maximum price, neither, or both.

Additionally, you can recommend the future participant to publish, advise against publication, or give no recommendation. The recommendation can be made alongside the minimum and maximum price. Any recommendation will be shown to the future participant during the decision on Willingness-to-Accept.

In the future experiment, the rules you set will be implemented. The participant will learn that they have been assigned to a participant from today's experiment; any minimum or maximum price will only be shown to them after the Willingness-to-Accept has been given.

Control Questions

Before you can continue with the experiment, you must answer the following control questions correctly. If you answer all control questions correctly the first time, you will receive a bonus of €0.50.

1. Assume the minimum price is €0.20; the computer randomly drew €2.96 and the participant's Willingness-to-Accept is €1.20. The maximum price is €2.78. Which statement is correct?

- ○ The publication takes place; the participant receives €0.20.
- ○ The publication takes place; the participant receives €1.20.
- ○ The publication takes place; the participant receives €2.96.
- ◉ The publication does not take place; the participant receives €0.

2. Assume the participant's Willingness-to-Accept is €2.18; the minimum price is €2.51 and the computer randomly drew €2.43. The maximum price is €3.90. Which statement is correct?

- ○ The publication takes place; the participant receives €0.57.
- ○ The publication takes place; the participant receives €3.18.
- ○ The publication takes place; the participant receives €2.65.
- ◉ The publication does not take place; the participant receives €0.

3. Assume the computer randomly drew €2.84; the minimum price is €1.16 and the participant's Willingness-to-Accept is €0.09. No maximum price was set. Which statement is correct?

- ◉ The publication takes place; the participant receives €2.84.
- ○ The publication takes place; the participant receives €1.16.
- ○ The publication takes place; the participant receives €0.09.
- ○ The publication does not take place; the participant receives €0.

4. Assume the computer randomly drew €1.13; the maximum price is €2.20 and the participant's Willingness-to-Accept is €0.71. No minimum price was set. Which statement is correct?

- ○ The publication takes place; the participant receives €2.20.
- ○ The publication takes place; the participant receives €0.71.
- ◉ The publication takes place; the participant receives €1.13.
- ○ The publication does not take place; the participant receives €0.

5. Which statement is **not** correct?

- ○ It is still in the best interest of the future participant to state their true Willingness-to-Accept.
- ○ A minimum price exceeding €0, or a maximum price less than €4.01, generally reduces the likelihood of publication.
- ◉ Setting a minimum price generally increases the payout of the future participant.
- ○ It is optional to give a recommendation and I can freely combine this measure with minimum and maximum prices.
- ○ Setting minimum and/or maximum prices has no consequences for me; I receive a flat €3 in this part.
- ○ Publication is never mandatory.

Next

Screen 10a

Part C

Time left to complete this page: **7:55**

In this part, you will receive a flat rate of €3. There are no right or wrong answers in the following. Please make your decisions freely according to your preferences.

Please first indicate whether you would like to give a recommendation to the participant:

- ○ I do not give any recommendation.
- ◉ I send the following message: "Publication is not recommended."
- ○ I send the following message: "Publication is recommended."

Continue

Screen 10b

Part C

Time left to complete this page: **7:42**

In this part, you will receive a flat rate of €3. There are no right or wrong answers in the following. Please make your decisions freely according to your preferences.

Please choose whether you want to set a minimum price and/or a maximum price. Remember: You do not have to set either.

- [x] I want to set a **minimum price** for publication.
- [ ] I want to set a **maximum price** for publication.

Continue

Screen 10c

Part C

Time left to complete this page: **7:24**

In this part, you will receive a flat rate of €3. There are no right or wrong answers in the following. Please make your decisions freely according to your preferences.

You have the option here to set the **minimum price** for publication. Remember: The minimum price means that no publication will occur if the price randomly drawn by the computer is less than the minimum price.

€0.00 €4.01

Minimum price: **0,50**

Next

Screen 11

Part D

Time left to complete this page: **9:46**

In this part, you can earn up to €2.

During the previous experiment, we displayed the following message to some participants: "Publication is not recommended."

Below are six groups. Each participant from the previous experiment belonged to one of these groups depending on their Willingness-to-Accept. Please estimate: What was the percentage of participants in the various groups **when publication was recommended against**? In the table, you see your estimates of the distribution *without* the recommendation against publication on the right.

Willingness-to-Accept for Publication	Your estimate of how many participants (in percent) exhibited this Willingness-to-Accept in the previous experiment	Your estimate without the recommendation against publication
exactly €0	5 %	20.0%
from €0.01 to €1	5 %	10.0%
from €1.01 to €2	5 %	10.0%
from €2.01 to €3	5 %	10.0%
from €3.01 to €4	30 %	20.0%
more than €4 (no publication)	50 %	30.0%
Total	**100%**	**100%**

Before you can submit your entries, the sum of the percentages must equal 100%.

The more accurate your estimate is to the actual distribution of the data, the higher the probability that you will receive a bonus of €2 in this part.

Specifically: One of the six groups mentioned above will be randomly selected. Your estimate will then be compared to the actual percentage of the group: You will receive €2 in this part with a k% probability, where $k = 100 \cdot (1 - 0.05 \cdot (\text{Estimate} - \text{TrueValue})^2)$. The probability of winning €2 is thus higher the closer you are to the true values. It is therefore in your interest to provide your best possible estimate.

Next

Screen 12

Part E

Time left to complete this page: **9:20**

In this experiment, you also have the opportunity to earn additional money by allowing your data to be published. By publication, we mean that the responses to the questions and the donation decision of the individual participant will be shared with other participants of today's experiment. Additionally, these data will be published on our website ockenfels.uni-koeln.de under "News" after the experiment.

You can review exactly which data will be published for you here if you wish:

Which data can be published?

Below, you must indicate your Willingness-to-Accept in euros for the publication of your own data. This is the minimum amount you want to receive to agree to the publication of the data. Here, €0 means that you accept any amount for publication. €4 means that you only accept €4 or more. Later, the computer will randomly draw a price between €0 and €4 inclusive. If this randomly drawn price exceeds or matches your Willingness-to-Accept, you will be paid the price in euros and the publication will proceed. However, if the price is less than your Willingness-to-Accept, you will not receive any payment for this part of the experiment and no publication will occur. Please note that you can object to the publication by indicating that your Willingness-to-Accept is higher than €4. Since the price is only drawn between €0 and €4, there will then never be a publication. At the same time, it is valid that with a Willingness-to-Accept of €0, any randomly determined euro amount is accepted and leads to publication.

Important: This decision only concerns you — you should therefore decide independently of your decisions in previous parts of the experiment. It is in your own interest to state your true "value" for the publication of your data. If you are unsure, then ask yourself how much money you would need to receive at a minimum for your data to be published.

If you want to indicate a Willingness-to-Accept from €0 to €4, click on the blue bar to display the slider.

€0.00 €4.00

Your Willingness-to-Accept: **3,00**

☐ Your Willingness-to-Accept is higher than €4. (This precludes publication.)

Next

Screen 13

Part F

Time left to complete this page: **2:55**

In Part C, you were able to set the rules for a participant of a future experiment. Each participant of the future experiment will also be paired with another participant from this experiment. It will definitely be a different participant, that is, not the participant from Part C. You can now influence the payout of this participant if you are paired with them. Please choose:

◉ The future participant should receive an additional **€0.50**.
○ **€0.50** should be deducted from the future participant's payout amount.

Please note that your decision does not affect your own payout.

Next

Screen 14

Just a moment...

Time left to complete this page: **9:29**

Please select to what extent you agree with each of the following statements. Here, 1 means "strongly disagree" and 7 means "strongly agree".

1. It usually bothers me when companies ask me for personal information.

strongly disagree 1 2 3 4 5 6 7 *strongly agree*

2. When companies ask me for personal information, I sometimes think twice before providing it.

strongly disagree 1 2 3 4 5 6 7 *strongly agree*

3. It bothers me to give personal information to so many companies.

strongly disagree 1 2 3 4 5 6 7 *strongly agree*

4. I'm concerned that companies are collecting too much personal information about me.

strongly disagree 1 2 3 4 5 6 7 *strongly agree*

5. All the personal information in computer databases should be double-checked for accuracy---no matter how much this costs.

strongly disagree 1 2 3 4 5 6 7 *strongly agree*

6. Companies should take more steps to make sure that the personal information in their files is accurate.

strongly disagree 1 2 3 4 5 6 7 *strongly agree*

7. Companies should have better procedures to correct errors in personal information.

strongly disagree 1 2 3 4 5 6 7 *strongly agree*

8. Companies should devote more time and effort to verifying the accuracy of the personal information in their databases.

strongly disagree 1 2 3 4 5 6 7 *strongly agree*

9. Companies should not use personal information for any purpose unless it has been authorized by the individuals who provided the information.

strongly disagree 1 2 3 4 5 6 7 *strongly agree*

10. When people give personal information to a company for some reason, the company should never use the information for any other reason.

strongly disagree 1 2 3 4 5 6 7 *strongly agree*

11. Companies should never sell the personal information in their computer databases to other companies.

strongly disagree 1 2 3 4 5 6 7 *strongly agree*

12. Companies should never share personal information with other companies unless it has been authorized by the individuals who provided the information.

strongly disagree 1 2 3 4 5 6 7 *strongly agree*

13. Companies should devote more time and effort to preventing unauthorized access to personal information.

strongly disagree 1 2 3 4 5 6 7 *strongly agree*

14. Computer databases that contain personal information should be protected from unauthorized access---no matter how much it costs.

strongly disagree 1 2 3 4 5 6 7 *strongly agree*

15. Companies should take more steps to make sure that unauthorized people cannot access personal information in their computers.

strongly disagree 1 2 3 4 5 6 7 *strongly agree*

Next

Screen 15

Just a moment...

Time left to complete this page: **59:11**

In the next step, you will receive an overview of the data shared by you and the other participants. Can you describe why you made your decisions in this experiment? Why did you set a minimum price of €0,50 €? Why did you not set a maximum price? Why is your Willingness-to-Accept (€3,00 €) higher than your minimum price? Can you also briefly describe what research question, in your opinion, this experiment aims to answer?

I just heckin' love being paternalistic

Next

Screen 16

Data

Time left to complete this page: **119:57**

As indicated on the previous page, we are now showing you and all other participants in the current experiment the data of those participants whose data was published. This data will also be published on our website (ockenfels.uni-koeln.de) in the "News" section within the next few weeks. The name of this experiment is "Experiment5".

The data of other participants will appear here shortly. This page will automatically refresh.

Person	Donate to UNICEF?	Gender	Only Child	Field of Study	Party at Fed	Religion	Football Club	Travel in the country only with CWA?	General Speed Limit on Highways?	Tax Meat Additionally?	50% of Seats in Parliament for Women?
You	*No publication*										

On the next page, you will find the payout information.

Next

Screen 17

End

Time left to complete this page: **4:34**

The experiment is now concluded.

Your total payout is 9,50 €. It is composed as follows:

- For your **punctual arrival**, you receive €2.50.
- **Part A**: For answering the questions, you receive a flat rate of 4,00 €. You have decided to donate to UNICEF. This changes your payout by -3,00 €.
- **Part B**: Since you answer the control questions correctly the first time, you receive 0,50 €. Additionally, we have compared your estimate with the actual data from the previous experiment. Based on your estimate for the 4th group, you receive 2,00 €.
- **Part C**: Since you answer the control questions correctly the first time, you receive 0,50 €. Additionally, you receive a flat rate of 3,00 € in this part.
- **Part D**: We have compared your estimate with the actual data from the previous experiment. Based on your estimate for the 2th group, you receive 0,00 €.
- **Part E**: You have indicated a Willingness-to-Accept of 3,00 € . The computer randomly drew the price 0,00 €. Since the price is less than your Willingness-to-Accept, no publication will occur. You receive 0,00 €.

Thank you for participating in this experiment.

Payout

To proceed with the payout, we need your PayPal email address.

Please click "Next" now to access the payout form.

Next

Screen 18 (session 1 only)

Part B

Time left to complete this page: **8:01**

In this experiment, we offer you the opportunity to earn additional money by publishing the data. By publication, we mean that the responses to the questions and the donation decision of the individual participant will be shared with the other participants of today's experiment. Furthermore, these data will be published on our website ockenfels.uni-koeln.de under "News" after the experiment.

You can review exactly which data will be published for you here if you wish:

Which data can be published?

Below, you must indicate your Willingness-to-Accept in euros for the publication of your own data. This is the minimum amount you want to receive to agree to the publication of the data. Here, €0 means that you accept any amount for publication. €4 means that you only accept €4 or more. Later, the computer will randomly draw a price between €0 and €4 inclusive. If this randomly drawn price exceeds or matches your Willingness-to-Accept, you will be paid the price in euros and the publication will proceed. However, if the price is less than your Willingness-to-Accept, you will not receive any payment for this part of the experiment and no publication will occur. Please note that you can object to the publication by indicating that your Willingness-to-Accept is higher than €4. Since the price is only drawn between €0 and €4, there will never be a publication. At the same time, a Willingness-to-Accept of €0 means that any randomly determined euro amount is accepted and leads to publication.

Important: It is in your own interest to state your true "value" for the publication of your data. If you are unsure, then ask yourself how much money you would need to receive at a minimum for your data to be published.

Publication is advised against.

If you want to indicate a Willingness-to-Accept from €0 to €4, click on the blue bar to display the slider.

☑ Your Willingness-to-Accept is higher than €4. (This precludes publication.)

Next

Screen 19 (session 1 only)

End

Time left to complete this page: **4:59**

The experiment is now over.

Your total payout is 3,50 €. It is composed as follows:

- For your **punctual arrival**, you receive €2.50.
- **Part A**: For answering the questions, you receive a flat rate of 4,00 €. You have decided to donate to UNICEF. This changes your payout by -3,00 €.
- **Part B**: You have indicated a Willingness-to-Accept of more than €4 . The computer randomly drew the price 0,00 €. Since the price is less than your Willingness-to-Accept, no publication will take place. You receive 0,00 €.

Thank you for participating in this experiment.

Payout

To proceed with the payout, we need your PayPal email address.

Please click "Continue" now to access the payout form.

Next

Screen 20 (session 3 only)

Part B

Time left to complete this page: **9:15**

As part of this experiment, you have the opportunity to earn additional money by agreeing to the publication of your data. By publication, we mean that the responses to the questions and the donation decision of the individual participant will be shared with other participants of today's experiment. Additionally, these data will be published on our website ockenfels.uni-koeln.de under the "News" section after the experiment.

You can review exactly which data will be published here if you wish:

Which data can be published?

Below, you must indicate your Willingness-to-Accept in euros for the publication of your own data. This is the minimum amount you want to receive to agree to the publication of the data. Here, €0 means that you accept any amount for publication. €4 means that you only accept €4 or more. Later, the computer will randomly draw a price between €0 and €4 inclusive. If this randomly drawn price exceeds or matches your Willingness-to-Accept, the price in euros will be paid to you and the publication will proceed. However, if the price is less than your Willingness-to-Accept, you will not receive any payment for this part of the experiment and no publication will occur. Please note that you can object to the publication by indicating that your Willingness-to-Accept is higher than €4. Since the price is only drawn between €0 and €4, there will never be a publication. Similarly, a Willingness-to-Accept of €0 means that any randomly determined euro amount is accepted and leads to publication.

Important: It is in your own interest to state your true "value" for the publication of your data. If you are unsure, then consider how much money you would need to receive at a minimum for your data to be published.

Note: Today's experiment also involves implementing a decision made by a person from a previous experiment. In this previous experiment, the person could set a minimum and also a maximum price for you (either none, one, or both). At the end of this experiment, you will not only find out if the computer-drawn price was above your Willingness-to-Accept, but also if the price was less than the minimum price or at least as high as the maximum price set by the person from the previous experiment. If one of these conditions is met, your data will not be published, even if the drawn price exceeds your Willingness-to-Accept. Despite these circumstances, it remains in your own interest to state your true "value" for the publication of your data.

The person from the previous experiment sends you the following message: **Publication is advised against.**

If you want to indicate a Willingness-to-Accept from €0 to €4, click on the blue bar to display the slider.

€0.00 €4.00

Your Willingness-to-Accept: **0,01**

☐ Your Willingness-to-Accept is higher than €4. (This precludes publication.)

Next

Screen 21 (session 3 only)

End

Time left to complete this page: **4:55**

The experiment is now over.

Your total payout is 4,50 €. It is composed as follows:

- For your **punctual arrival**, you receive €2.50.
- **Part A**: For answering the questions, you receive a flat rate of 4,00 €. You have decided to donate to UNICEF. This changes your payout by -3,00 €.
- **Part B**: You have indicated a Willingness-to-Accept of 0,01 € . The computer randomly drew the price 2,53 €. The person assigned to you from the previous experiment has set a minimum price of €3.0. The person assigned to you from the previous experiment has set a maximum price of €4.0. Since the price drawn by the computer does not lie between the set minimum and maximum prices, no publication will occur, although the drawn price is at least as high as your Willingness-to-Accept. You receive 0,00 €.
- You were also assigned to another person from the previous experiment. This person has increased your payout by €1.

Thank you for participating in this experiment.

Payout

To proceed with the payout, we need your PayPal email address.

Please click "Continue" now to access the payout form.

Next

Appendix B

Appendix to Chapter 2

B.1 Data availability

Software, materials, data and analysis code used in this article are freely available at `https://gitlab.com/gr0ssmann/knf`. We welcome attempts at replication and reproduction.

	Model 1	Model 2	Model 3	Model 4	Model 5
Intercept	0.210***	0.042*	0.599***	0.732***	0.829***
	(0.017)	(0.019)	(0.018)	(0.040)	(0.018)
CA prefers 1	0.094***	0.113***	0.157***	0.220***	−0.012
	(0.023)	(0.033)	(0.019)	(0.041)	(0.020)
Experiment	1	1	2	2	2
Outcome	π	$\pi > 0.5$	π	$\pi > 0.5$	π'
Belief about choice under	Risk	Risk	Risk	Risk	Ambiguity
Standard errors	HC3	HC3	HC3	HC3	HC3
R^2	0.048	0.031	0.142	0.093	0.001
Adj. R^2	0.045	0.028	0.141	0.092	−0.001
Num. obs.	300	300	600	600	600

$^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$

Table B.2: False consensus bias across experiments

The binary outcome $\pi > 0.5$ indicates whether Option 1 is viewed to be the majority choice.

B.2 Additional tables

k	Mean	Median	Mode	Var.	MAE	RMSE	D_{KL}	W_1
0	0.500	0.500	—	0.083	0.340	0.416	0.000	0.000
1	0.400	0.362	0.000	0.073	0.261	0.337	0.062	0.100
2	0.350	0.307	0.000	0.062	0.222	0.290	0.148	0.150
5	0.286	0.252	0.083	0.040	0.167	0.217	0.338	0.214
10	0.250	0.228	0.163	0.025	0.128	0.165	0.526	0.250
25	0.222	0.212	0.187	0.011	0.086	0.109	0.852	0.280
50	0.212	0.206	0.194	0.006	0.062	0.079	1.152	0.297
1000	0.201	0.200	0.200	0.000	0.014	0.018	2.607	0.330
∞	0.200	0.200	0.200	0.000	0.000	0.000	∞	0.340

Table B.1: Properties of the marginal posterior distribution of p

MAE and RMSE are calculated as departures from $p = 0.2$.
The Kullback–Leibler divergence and the Wasserstein 1-distance are calculated as differences between the marginal posterior and $U(0, 1)$.
$U(0, 1)$ is the distribution used for $k = 0$.

B.3 Experiment 1: Analyses

B.3.1 Transformation of independent variable

k	0	1	2	5	10	25	50	1000	∞
k_{rank}	0	1	2	3	4	5	6	7	8

B.3.2 Main analysis

	Model 1†	Model 2	Model 3
k_{rank}	-0.188^{***}	-0.192^{***}	-0.139^{***}
	(0.019)	(0.019)	(0.040)
Round		0.043^{***}	0.057^{*}
		(0.014)	(0.032)
Subset	—	—	Inconsistent CAs‡
Controls	No	Yes	Yes
AIC	3518.186	3478.862	756.761
BIC	3529.988	3520.169	787.032
Log Likelihood	−1757.093	−1732.431	−371.381
Deviance	3514.186	3464.862	742.761
Num. obs.	2700	2700	558

$^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.1$

Table B.3: Logistic regressions on the extensive margin

All standard errors are clustered by subject. †Original preregistered specification. ‡Consistency is defined in the preregistration: "a [CA] is consistent if and only if they never intervened or if they imposed Option 1 OR Option 2, but never switched between the Options."

Define by intervene_{ij} a binary variable that equals '1' if subject i intervened in round j, and '0' otherwise. We can run a logistic regression of intervene_{ij} on k_{rank}, with standard errors clustered by subject. As Table B.3 indicates,[1] there is a statistically significant reduction in intervene_{ij} along with k_{rank}, and thus k.

[1] Regression outputs courtesy of `texreg`, see Leifeld (2013).

This holds both for the “naked”, preregistered, regression (Model 1) as well as when including the round and control variables.[2]

B.3.3 Beliefs and the intensive margin

	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	Model 7	Model 8	Model 9
Intercept	0.512^{***}	0.484^{***}	0.503^{***}	0.234^{***}	0.223^{***}	0.222^{***}	0.223	0.219	0.213
	(0.038)	(0.041)	(0.045)	(0.050)	(0.050)	(0.057)	(0.145)	(0.145)	(0.146)
π (standardized)	0.087^{*}		0.067	0.025		−0.001	0.018		−0.017
	(0.037)		(0.067)	(0.031)		(0.063)	(0.031)		(0.067)
$\pi > 0.5$		0.253^{*}	0.079		0.099	0.101		0.089	0.131
		(0.114)	(0.210)		(0.092)	(0.192)		(0.092)	(0.200)
CA prefers 1				0.492^{***}	0.493^{***}	0.493^{***}	0.486^{***}	0.484^{***}	0.488^{***}
				(0.070)	(0.069)	(0.071)	(0.073)	(0.072)	(0.073)
Outcome	Imposed 1	Imposed 1	Imposed 1	Imposed 1	Imposed 1	Imposed 1	Imposed 1	Imposed 1	Imposed 1
Subset	c.n.l.	c.n.l.	c.n.l.	c.n.l.	c.n.l.	c.n.l.	c.n.l.	c.n.l.	c.n.l.
Controls	No	No	No	No	No	No	Yes	Yes	Yes
R^2	0.030	0.025	0.031	0.254	0.255	0.255	0.272	0.273	0.274
Adj. R^2	0.025	0.019	0.020	0.245	0.246	0.242	0.245	0.247	0.243
Num. obs.	172	172	172	172	172	172	172	172	172

$^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$
c.n.l. refers to “consistent non-libertarians.”

Table B.4: Regressions of intensive margin on CA beliefs and preferences

B.4 Experiment 2

B.4.1 Relative importance of beliefs and preferences

Let π_1 be the belief of an Option 1-preferring CA, and π_2 the belief of an Option 2-preferring CA.

We can use the coefficients of Table B.5 to find the π_2 that solves $0.115 + 0.544\pi_2 = 0.115 + 0.544\pi_1 + 0.263$. This would be the belief of the Option 2-preferring CA required to attain the same probability of imposing Option 1 as when the CA prefers Option 1. We find $\pi_2 \approx 0.484 + \pi_1$. The average π_1 in our sample (Table B.2) is 0.756. Obviously, such a large π_2 would not be achievable (since it exceeds 1). The counterfactual Option 1-preferring CA would have to have much lower beliefs. Thus, it is difficult to achieve parity through beliefs alone. Beliefs are weak predictors of interventions when compared to CAs’ preferences.

[2]Control variables are the university entrance exam grade or a dummy variable if it was not provided to us, a dummy indicating whether the subject ever took a class on introductory microeconomics and a dummy indicating whether the subject identified as male.

	Model
Intercept	0.115
	(0.150)
π	0.544^{**}
	(0.224)
CA prefers 1	0.263^{***}
	(0.096)
Outcome	Imposed 1
Subset	Intervened
Standard errors	HC3
R^2	0.164
Adj. R^2	0.151
Num. obs.	132

$^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.1$

Table B.5: Regression explaining the intensive margin with CA preference and raw beliefs

B.4.2 Performance of binary-choice model

Predictions (i.e., fitted values) for the probability of the outcome "Imposed (Option) 1" can be obtained from model 5 in Table 2.3 and model 2 in Table 2.4. Call these predictions $\hat{v}_{1,i}$, with i indicating the CA's index. Similarly, we can calculate

$$\hat{v}_{2,i} = \Phi\left(\frac{W_{\theta_i}(1) - W_{\theta_i}(2)}{\sigma}\right) \tag{B.1}$$

for each CA after maximization of Equation 2.9.

The threshold 1/2 is then used to form categorical predictions. If $\hat{v}_{m,i} \geq 1/2$ for any m, i, we predict "Imposed 1." Otherwise, we predict "Imposed 2." We can classify whether the predictions are correct by considering the actual outcome. A comparison of these classifiers reveals the following result for Chooser 1 (model 1 is model 5 of Table 2.3):

	Model 2 wrong	Model 2 correct	**Sum**
Model 1 wrong	32	4	36
Model 1 correct	2	94	96
Sum	34	98	132

For Chooser 5 (model 1 is model 2 of Table 2.4):

	Model 2 wrong	Model 2 correct	**Sum**
Model 1 wrong	50	23	73
Model 1 correct	24	91	115
Sum	74	114	188

Clearly, both our binary-choice models and the linear probability models perform about equally well. In both cases, correlations between the raw values of $\hat{v}_{m,i}$ exceed 0.9. Similar results can be obtained by running logistic regressions in place of the linear probability models. Our data package provides code and data for these analyses.

B.4.3 Complementarities

	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	Model 7	Model 8	Model 9
Intercept	0.635^{***}	0.819^{***}	0.522^{***}	0.544^{***}	0.699^{***}	0.914^{***}	0.598^{***}	0.882^{***}	0.653^{***}
	(0.034)	(0.037)	(0.033)	(0.036)	(0.057)	(0.059)	(0.055)	(0.057)	(0.064)
Int. for Chooser 1	1.312^{***}				1.442^{***}				
	(0.078)				(0.128)				
Int. for Chooser 2		1.354^{***}				1.670^{***}			
		(0.124)				(0.212)			
Int. for Chooser 3			1.281^{***}				1.446^{***}		
			(0.066)				(0.113)		
Int. for Chooser 4								1.761^{***}	
								(0.212)	
Int. for Chooser 5				1.212^{***}					1.270^{***}
				(0.066)					(0.111)
Outcome	IC_{-4}	IC_{-4}	IC_{-4}	IC_{-4}	IC_{+4}	IC_{+4}	IC_{+4}	IC_{+4}	IC_{+4}
Subset	—	—	—	—	Plus	Plus	Plus	Plus	Plus
Standard errors	HC3	HC3	HC3	HC3	HC3	HC3	HC3	HC3	HC3
R^2	0.345	0.151	0.412	0.368	0.340	0.187	0.398	0.239	0.307
Adj. R^2	0.344	0.150	0.411	0.367	0.337	0.184	0.396	0.236	0.304
Num. obs.	600	600	600	600	290	290	290	290	290

$^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.1$

Table B.6: Regression of intervention counts on intervention behavior

We can define the following variable:

$$IC_{\mathbb{S}} \equiv \sum_{i \in \mathbb{S}} [\text{Intervened for Chooser } i] . \tag{B.2}$$

This sum of indicators is essentially a counter. Table B.6 presents regressions of Equation B.2 on the individual indicators, with $\mathbb{S} = +4 = \{1, 2, 3, 4, 5\}$, $\mathbb{S} = -4 = \{1, 2, 3, 5\}$. Obviously, each coefficient must be at least 1 if the Chooser on the right-hand side of the regression equation is included in $\mathbb{S}$. To the extent that the coefficient exceeds 1, this hints at a general pattern where those who intervene for a Chooser are more likely to intervene for *other* Choosers, too.

B.4.4 Preregistered analyses

B.4.4.1 Correction to preregistration

In the preregistration, analysis A4 was specified as follows:

```
glm(I(intervene_known == 1) ~ I(own_preference == 1)
+ I(other == 1), family = binomial, data = data)
```

However, as is clear from the context ("This is to test for the relative importance of knowing the other's counterfactual perfect-knowledge choice.") this analysis relates to the intensive margin. Hence, the data used for analysis must be restricted to interveners. The corrected code used below is

```
... data = data[data$intervene_known > 0, ])
```

B.4.4.2 A1

Out of 600 CAs (Section 2.3.5.2), 290 are in treatment "Plus." 310 are in the baseline.

Out of the 290 CAs in "Plus," 236 (81.3%) provided information.

Out of the 310 CAs in the baseline, 249 (80.3%) provided information.

Using a one-sided test of equal proportions (the alternative being that fewer CAs provide information in "Plus"), we find $\chi^2 = 0.0506$, $p = 0.59$ (95% confidence interval for the difference: $[-1, 0.067]$).

B.4.4.3 A2

	Model 1
(Intercept)	-1.266^{***}
	(0.099)
Full Knowledge	-1.223^{***}
	(0.180)
AIC	960.952
BIC	971.132
Log Likelihood	−478476
Deviance	956.952
Num. obs.	1200

$^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.1$

Here, $z = -1.223/0.180 = -6.79$, $p < 0.001$.

B.4.4.4 A3

	Model 1
(Intercept)	-1.035^{***}
	(0.108)
Int. for Chooser 1	0.706^{***}
	(0.205)
Int. for Chooser 2	0.942^{***}
	(0.313)
AIC	731.148
BIC	744.339
Log Likelihood	−362574
Deviance	725.148
Num. obs.	600

$^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.1$

A Wald test on equality of the two non-Intercept independent variables reveals $\chi^2 = 0.4$. With 1 degree of freedom, $p = 0.53$.

B.4.4.5 A4

Note the correction in Section B.4.4.1.

	Model 1
(Intercept)	-0.571^{*}
	(0.343)
CA prefers 1	0.557
	(0.354)
Chooser prefers 1	1.094^{***}
	(0.321)
AIC	249.466
BIC	259.175
Log Likelihood	−121.733
Deviance	243.466
Num. obs.	188

$^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.1$

A Wald test on equality of the two non-Intercept independent variables reveals $\chi^2 = 1.5$. With 1 degree of freedom, $p = 0.21$.

B.5 Instructions

Note: The following screenshots represent English translations of experiments that were wholly (experiment 1) or partially (experiment 2) conducted in German.

B.5.1 Experiment 1

Screen 1

(Consent form.)

Screen 2

Important Information

*Time left to complete this page: **3:59***

In today's experiment, you will make decisions for yourself and others.

Please pay full attention to the contents of the experiment. In order to receive a payment, it is necessary that you answer later comprehension questions correctly.

The experiment consists of two parts (part 1 and part 2). You will only be paid for one of the two parts. With 95% probability, part 1 is relevant for the payment, and with the remaining probability, part 2 is relevant for the payment. In addition, you will receive €1 for punctual attendance.

Next

Screen 3

Part 1: Information about today's experiment

*Time left to complete this page: **9:46***

Imagine a choice between two options (**Option 1** and **Option 2**). **Option 1** always pays out €15. On the other hand, **Option 2** pays out either €0 or €20. Please take a look at the next table.

Option 1	Option 2
With 100% probability €15 will be received	With *42%* probability €0 will be received
	With a *58%* probability €20 will be received

The probability of the low outcome for **Option 2** is *p*%.

If *p* = 0%, then **Option 2** would be equivalent to a sure payment of €20. If *p* = 100%, then **Option 2** would be equivalent to a sure payment of €0.

By moving the following slider, you can see how **Option 2** changes depending on how high or low *p* is. This is for your information only. Please move this slider and observe the change in the table above:

0% 100%

Next

Screen 4

Part 1: Information about today's experiment

Time left to complete this page: ***9:26***

Option 1	Option 2
With 100% probability €15 will be received	With *p*% probability €0 will be received Otherwise €20 will be received

Imagine that we present the choice between **Option 1** and **Option 2** to other participants in an online experiment of the Cologne Laboratory for Economic Research. In this case, *p* would take on different values.

For the following values of *p*, what do you believe how many of 1,000 other participants prefer **Option 2**? You can enter a value by clicking on the blue bar and then moving the slider. By clicking on the different values of *p*, you can display them in the table above.

Please give us your best estimate.

p	How many of 1,000 other participants prefer Option 2?
0%	0 – 1000 Your estimate: **1000**
10%	0 – 1000 Your estimate: **900**
20%	0 – 1000 Your estimate: **636**
30%	0 – 1000 Your estimate: **373**
40%	0 – 1000 Your estimate: **257**
50%	0 – 1000 Your estimate: **199**
60%	0 – 1000 Your estimate: **119**
70%	0 – 1000 Your estimate: **64**
80%	0 – 1000 Your estimate: **38**
90%	0 – 1000 Your estimate: **0**
100%	0 – 1000 Your estimate: **0**

Next

Screen 5

Part 1: Information about today's experiment

Time left to complete this page: ***11:58***

1	2	3	4	5	6

If this part is relevant for the payment, you will receive at least €4.

In this part of today's experiment, you will make decisions that will affect another participant in a future online experiment at the Cologne Laboratory for Economic Research. We refer to this future person as the **other person**.

Next

Screen 6

Part 1: Information about today's experiment

Time left to complete this page: ***11:49***

1 2 3 4 5 6

The other person will make a choice between **Option 1** and **Option 2**. As a reminder: **Option 1** always pays out €15. On the other hand, **Option 2** pays out either €0 or €20. The other person will see the following table:

Option 1	Option 2
With 100% probability €15 will be received	With *p*% probability €0 will be received Otherwise €20 will be received

Back Next

Screen 7

Part 1: Information about today's experiment

Time left to complete this page: ***11:41***

1 2 3 4 5 6

The other person does not know the exact value of *p*.

The other person does know, however, that the value of *p* will be determined randomly. With equal probability, *p* will take on any whole percentage from 0% to and including 100%.

Today you will make a decision for the other person if *p* randomly takes on the value 20%.

Since there are several decision makers like you who decide about another person whose *p* takes on the value 20, your decision will only be implemented if you were randomly selected among all matching decision makers.

You should behave in each decision as if your decision would be implemented, because maybe it will.

Back Next

Screen 8

Part 1: Information about today's experiment

Time left to complete this page: ***11:36***

1 2 3 4 5 6

You can decide below,

1. whether the other person's own choice between **Option 1** and **Option 2** will be implemented or
2. whether you will implement **Option 1** or **Option 2** and override the other person's choice.

In any case, the other person will choose between **Option 1** and **Option 2**. You can only decide whether this own choice of the other person will be implemented.

If you decide that the other person's own choice will be implemented, then the other person will receive **Option 1** or **Option 2**, depending on how they decide for themselves.

If you decide to implement **Option 1** or **Option 2** and override the other person's choice, then the other person will still make a choice, but instead they will receive **Option 1** or **Option 2**, depending on how you decide today. In this case, the other person's own choice would not be taken into account.

Back Next

Screen 9

Part 1: Information about today's experiment

Time left to complete this page: ***11:20***

1 2 3 4 5 6

Before choosing between **Option 1** and **Option 2**, the other person may see the result of draws from **Option 2**.

How many draws will be made for the other person is determined randomly. You will make a decision today for different numbers of draws.

The draws are based on p = 20%. Through the result of the draws, the other person can learn more about **Option 2**, but they will not find out the exact value of p.

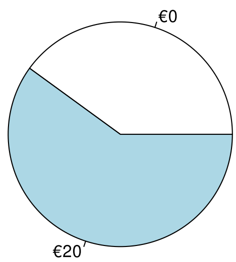
For example, the other person could see the result of 5 draws:

For the other person, the question mark will be replaced by the actual result of the draws. For illustration, you can also click here to perform example draws every few seconds: Start animation ▸

Result of draws

Option 2 was drawn 5 times.

Number of draws:	**5**
of which €0:	**2**



Back Next

Screen 10

Part 1: Information about today's experiment

Time left to complete this page: ***10:57***

1	2	3	4	5	6

Before you can decide, please answer the following comprehension questions.

If you answer all comprehension questions correctly on the first try and this part is selected for the payment, you will receive a bonus of €2.

Question 1: *p* is the probability for ...

- (selected) ... €0 in **Option 2**.
- ... €20 in **Option 2**.

Question 2: What is the probability *p* if your decisions can be implemented?

- 10%
- (selected) 20%
- 40%
- This cannot be said yet, since *p* is determined randomly.

Question 3: What does the other person know about *p*?

- Nothing.
- The exact value of *p*.
- (selected) That *p* is determined randomly, a whole percentage between 0% and 100% is possible and every possible value of *p* is equally likely.

Question 4: You decide today whether for the other person either **Option 1**, **Option 2** or the other person's own choice should be implemented.

- This is not correct.
- (selected) This is correct.

Question 5: How often is **Option 2** drawn for the other person before they can decide?

- (selected) This is determined randomly.
- Not at all.
- As often as they want.

Question 6: If you decide that the other person's choice is overridden, would the person still make a choice between **Option 1** and **Option 2**?

- The other person would not make any own choice anymore.
- (selected) The other person would still make an own choice between both options, but instead of their choice your choice today will be implemented.

Question 7: How often does the other person choose?

- (selected) Exactly once.
- This is determined randomly.

Back Next

Screen 11

Your decisions in today's experiment

Time left to complete this page: ***9:57***

Your participation in this experiment is an aspect of a larger research project that is of great importance to the researchers and the broader scientific community. It is therefore very important that you pay full attention to the contents of the experiment, as the time and effort that the researchers have put into the project could otherwise be wasted.

Please honestly tell us if you have been fully devoted to the experiment so far and if we should use your data in your opinion. **You will receive your payout regardless of what you answer here.**

- ○ Yes, I paid full attention to the experiment. My data should be used.
- ○ No, I did not pay full attention to the experiment. My data should not be used.

Next

Screen 12

Part 1: Information about today's experiment

Time left to complete this page: ***9:59***

Reminder: The other person only knows

- the certain payment of €15 from **Option 1**, and that it is certain,
- the possible payment amounts from **Option 2**, namely €0 and €20,
- that the probability *p* for €0 in **Option 2** randomly takes on a whole percentage from 0% to and including 100%, but not the exact value of *p*,
- possibly a result from draws from the real **Option 2**.

You and only you know in addition that *p* for this person actually takes on the value 20%.

The actual number of draws is determined randomly; thus at most one of your decisions is implemented. You should behave in each decision as if your decision were implemented, because maybe it will be.

Next

Screen 13

Part 1: Your Decision

*Time left to complete this page: **9:51***

If this decision is randomly selected and thus implemented, then the other person will, before making the choice between **Option 1** and **Option 2**, see the result of 50 draws:

For the other person, the question mark will be replaced by the actual result of the draws. For illustration, you can also click here to perform example draws every few seconds: Start animation ▸

Option 2 was drawn 50 times.

Number of draws:	**50**
of which €0:	**?**

You can also take a look at the table that the other person sees before the decision:

Option 1	Option 2
With 100% probability €15 will be received	With *p*% probability €0 will be received Otherwise €20 will be received

You and only you know in addition that *p* actually takes the value 20% for this person.

Your Decision

Should the **other person's own choice** between **Option 1** and **Option 2** be implemented for the other person?

Yes No

☑ I would like to make my decision as indicated.

Continue

Screen 14

Part 1: Your Decision

Time left to complete this page: ***9:47***

If this decision is randomly selected and thus implemented, then the other person will see a choice between **Option 1** and **Option 2** the result of 0 (i.e. no) draws.

You can also take a look at the table that the other person sees before the decision:

Option 1	Option 2
With 100% probability €15 will be received	With *p*% probability €0 will be received
	Otherwise €20 will be received

You and only you know in addition that *p* actually takes the value 20% for this person.

Your Decision

Should the **other person's own choice** between **Option 1** and **Option 2** be implemented for the other person?

Yes | No

Which option would you like to implement for the other person and thus overwrite the other person's own choice?

Option 1 | Option 2

☑ I would like to make my decision as indicated.

Continue

Screen 15

Part 1: Your Decision

*Time left to complete this page: **9:36***

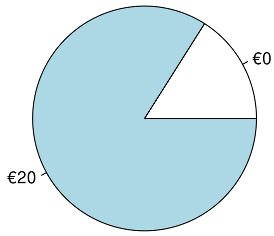
If this decision is randomly selected and thus implemented, then the other person will, before making the choice between **Option 1** and **Option 2**, see the result of 25 draws:

For the other person, the question mark will be replaced by the actual result of the draws. For illustration, you can also click here to perform example draws every few seconds: Start animation ▸

Result of draws

Option 2 was drawn 25 times.

Number of draws:	**25**
of which €0:	**4**



You can also take a look at the table that the other person sees before the decision:

Option 1	Option 2
With 100% probability €15 will be received	With *p*% probability €0 will be received Otherwise €20 will be received

You and only you know in addition that *p* actually takes the value 20% for this person.

Your Decision

Should the **other person's own choice** between **Option 1** and **Option 2** be implemented for the other person?

Yes No

☑ I would like to make my decision as indicated.

Continue

Screen 16

Part 1: Your Decision

Time left to complete this page: ***9:52***

We are also conducting an experiment with another person who knows the exact value of *p*. You can also make a decision for this person, which may be randomly selected and implemented. The experiment for this person is otherwise set up in the same way as for the other person who does not know the exact value of *p*.

If this decision is randomly selected and thus implemented, then the other person will see a choice between **Option 1** and **Option 2** the exact value of *p*, i.e. 20%.

You can take a look at the table that the other person sees before the decision:

Option 1	Option 2
With 100% probability €15 will be received	With *20%* probability €0 will be received With *80%* probability €20 will be received

Your Decision

Should the **other person's own choice** between **Option 1** and **Option 2** be implemented for the other person?

Yes No

☑ I would like to make my decision as indicated.

Continue

Screen 17

Part 2: Your Estimate

Time left to complete this page: ***9:52***

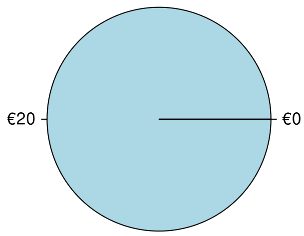
Reminder: *p* is the probability of €0 in **Option 2**. *p* takes a random whole percentage from 0% to 100% in the future experiment. The exact value of *p* is not known to the other person.

Imagine the other person sees the following result of the draws:

Result of draws

Option 2 was drawn 2 times.

Number of draws:	**2**
of which €0:	**0**



What do you think: What value of *p* does the other person assume?

Please give us your best estimate.

0% 100%

Your estimate: **10%**

Next

Screen 18

Part 2: Your own decision

*Time left to complete this page: **9:54***

Reminder: If *p* = 20%, the options look like this:

Option 1	Option 2
With 100% probability €15 will be received	With *20%* probability €0 will be received
	With *80%* probability €20 will be received

If this part is relevant for the payout, then you will be paid out the option you personally prefer.

This decision only affects you.

Which option do you prefer?

- (●) **Option 1**
- ( ) **Option 2**

Next

Screen 19

Demographic Information

*Time left to complete this page: **9:44***

Which gender do you identify with?

- (●) Male
- ( ) Female
- ( ) Diverse

Have you already taken a basic course in microeconomics as part of your studies?

- (●) Yes
- ( ) No

What overall grade did you receive in the general university entrance qualification ("Abitur")? If you do not wish to disclose this, or if you obtained your university entrance qualification outside of the German grading system, please enter 0.

1.2

Next

Screen 20

Your decisions in today's experiment

Time left to complete this page: ***119:57***

We are interested in why you made the decisions you did in today's experiment. Could you tell us your motives? Thank you. **Please explain why you decided (not) to implement the other person's choice at any given time.**

Please do not enter any personal data in the following fields.

What do you think: What research question is addressed with this experiment?

Next

Screen 21

End of Experiment

Time left to complete this page: ***9:59***

The experiment has concluded. It was randomly determined that part 1 is relevant for the payment.

Your total payment is **€7.00**. This consists of the following:

1. Punctual appearance: **€1.00**
2. Payment from part 1: **€6.00**

The payment for part 1 includes a bonus of €2, as you answered the comprehension questions correctly on the first attempt.

Thank you for your participation.

On the next page you will find the payment form.

Next

B.5.2 Experiment 2

Screen 1

(Consent form.)

Screen 2

Schedule of Today's Experiment

Welcome. Today's experiment consists of three parts and will take about 10 minutes.

Your final payout is based on the payout of a randomly selected part. The following table shows the probability of each part being selected:

Part 1	1%
Part 2	9%
Part 3	90%

In addition, you will receive **€1** for being on time. Behave in each part as if it were relevant for the payout.

All information in this experiment is truthful.

Continue

Screen 3

Schedule of Today's Experiment

In today's experiment, the choice between two **Options** plays an important role. Please carefully consider **Option 1** and **Option 2**:

Option 1	**Option 2**
With 100% probability receive €15	With 20% probability receive €0
	Otherwise receive €20

Option 1 and **Option 2** remain the same throughout the entire experiment. In a choice between **Option 1** and **Option 2**, only one of the two **Options** can be selected.

Click "Continue" only after you have understood **Option 1** and **Option 2**.

Back

Continue

Screen 4

Your Choice

This is **Part 1 of 3** of today's experiment.

Option 1	Option 2
With 100% probability receive €15	With 20% probability receive €0
	Otherwise receive €20

Which **Option** would you like to choose for yourself in this part?

Option 1 Option 2

Continue

Screen 5

Your Estimate

This is **Part 2 of 3** of today's experiment. On this page, you have the opportunity to win **€10** by making an accurate estimate.

Please review the two **Options** again:

Option 1	Option 2
With 100% probability receive €15	With 20% probability receive €0
	Otherwise receive €20

In March 2023, we presented this choice to 300 participants at the Cologne Laboratory for Economic Research.

Please estimate now: How many of the 300 participants chose **Option 1**?

The more accurate your estimate, the higher your chance of winning. It is therefore in your own interest to estimate as accurately as possible.

How does this work?

If your estimate is x and the true (known to us) value is y, then $p = \max(0, \min(1, 1 - 0.15 \cdot |x - y|))$ is the probability of winning. **Important**: You do not need to understand this calculation. However, it is in your own interest to estimate as accurately as possible, as this increases your chances of winning.

Your Estimate

Enter a number between 0 and 300.

242 That's **81%** of 300.

Continue

Screen 6

Your Estimate

This is **Part 2 of 3** of today's experiment. For your full attention on this page, you will receive an additional **€5** in this part.

Consider these two **Options**:

Option 1	**Option 2**
With 100% probability receive €15	With ▒% probability receive €0
	Otherwise receive €20

This means: The probability that the payout for **Option 2** is €0 is unknown to the participant here.

In March 2023, we could have presented this choice to 300 other participants of the Cologne Laboratory for Economic Research.

Please estimate now: How many of the 300 participants would have chosen **Option 1**?

Your Estimate

Enter a number between 0 and 300.

260 That's **87%** of 300.

Continue

Screen 7

Instructions

This is **Part 3 of 3** of today's experiment. In this part, you will receive at least **€2.**

The choice between **Option 1** and **Option 2** will also be made by five other people. **Today, you have the opportunity to set the rules according to which the other people make *their* choice.** Here are illustrations of these five people:

Each of these five people will be in a specific scenario, which will be described on the following pages. They are five different individuals.

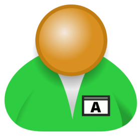
For each of these five people, a participant from the current series of experiments will be selected to determine the rules for the respective scenario. Therefore, your decisions could be implemented. Your decisions today can have real effects on real people. Please take your decisions seriously.

In any case, all five people will make their own choice. However, each of the five people will be informed that their own choice may not be implemented. This is among the things you will determine today. How the five people choose has no significance for your payout in today's experiment.

Continue

Screen 8

Your Decision Regarding Person A



Person A will see the following before choosing between **Option 1** and **Option 2**:

Option 1	Option 2
With 100% probability receive €15	With 20% probability receive €0 Otherwise receive €20

This means: Person A knows the possible outcomes, the structure of the **Options**, and also the probability that the payout in **Option 2** will be €0.

Afterwards, Person A will choose between the **Options**.

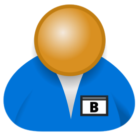
Should the **personal choice** of Person A between **Option 1** and **Option 2** be implemented for Person A?

Yes | No

Continue

Screen 9

Your Decision Regarding Person B



Person B will see the following before choosing between **Option 1** and **Option 2**:

Option 1	**Option 2**
With 100% probability receive €15	With ___% probability receive €0 Otherwise receive €20

This means: Person B knows the possible outcomes and the structure of the **Options**, but not the probability that the payout in **Option 2** will be €0.

You and only you know that the probability that the payout in **Option 2** will be €0 is actually 20%.

Afterwards, Person B will choose between the **Options**.

Should the **personal choice** of Person B between **Option 1** and **Option 2** be implemented for Person B?

Yes | No

Which **Option** would you like to implement for the other person, thereby overriding the personal choice of the other person?

Option 1 | Option 2

Continue

Screen 10

Your Decision Regarding Person C

Person C is faced with a choice between **Option 1** and **Option 2**. Today, you can determine how well Person C is informed about **Option 2**. The choice made by Person C between the **Options** will be implemented. You do not have the option to override this choice.

Person C will not know that the information they receive has been set by you.

Should Person C know the probability of €0 in **Option 2**?

Click on the buttons to preview for Person C.

No Yes

Option 1	**Option 2**
With 100% probability receive €15	With 20% probability receive €0
	Otherwise receive €20

In this case, Person C would learn the probability of receiving €0 in **Option 2**.

Afterwards, Person C will choose between the **Options**.

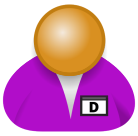
Regarding the Choice of Person C

Person C's **own choice** between **Option 1** and **Option 2** will be implemented. You cannot override this choice.

Continue

Screen 11

Your Decision Regarding Person D



Person D is also faced with the choice between **Option 1** and **Option 2**. However, it is left to chance how well Person D is informed about **Option 2**. There is a 50% chance that Person D sees the following:

Option 1	**Option 2**
With 100% probability receive €15	With ██% probability receive €0
	Otherwise receive €20

In this case, Person D would not be informed about the probability of receiving €0 in **Option 2**.

Otherwise, Person D receives this information:

Option 1	**Option 2**
With 100% probability receive €15	With 20% probability receive €0
	Otherwise receive €20

In this case, Person D would be informed about the probability of receiving €0 in **Option 2**.

You know that the probability of the payout being €0 in **Option 2** is actually 20%. Whether Person D is also aware of this depends on chance. It's akin to a coin toss: either Person D is informed or not. Thus, it is still uncertain how well Person D is informed.

Afterwards, Person D will choose between the **Options**.

Should the **personal choice** of Person D between **Option 1** and **Option 2** be implemented for Person D?

Yes | No

Continue

Screen 12

Your Decision Regarding Person E

Person E will see the following before choosing between **Option 1** and **Option 2**:

Option 1	Option 2
With 100% probability receive €15	With % probability receive €0
	Otherwise receive €20

This means: Person E knows the possible outcomes and the structure of the **Options**, but not the probability that the payout for **Option 2** will be exactly €0.

We guarantee that your following decision can only be implemented if Person E actually prefers **Option 2** and would choose **Option 2** if fully informed. How do we know this? Person E will be asked how they would choose between **Option 1** and **Option 2** at various possible probabilities for €0 in **Option 2**. (Person E does not know which probability actually applies.) Your decision today can only be implemented if Person E prefers **Option 2** when faced with the possibility that the probability for €0 is exactly 20%.

Afterwards, Person E chooses between the **Options** without knowing the actual probability.

You and only you know that the probability that the payout for **Option 2** is exactly €0 is actually 20%. Your rule for Person E can only be implemented if Person E would have preferred **Option 2** under full knowledge of the circumstances.

Should the **personal choice** of Person E between **Option 1** and **Option 2** be implemented for Person E?

Yes | No

Which **Option** would you like to implement for the other person, thereby overriding the personal choice of the other person?

Option 1 | Option 2

Continue

Screen 13

Just a moment...

This is still **Part 3 of 3** of today's experiment.

If you answer the following question correctly, you will receive an additional €0.25 in this part. Please recall Person E. Person E did not know the probability of €0 in **Option 2**. However, you have learned that your decision for Person E can only be implemented if Person E would have preferred a certain **Option** had they been fully informed.

Assume your decision for Person E can be implemented.

Which **Option** would Person E have preferred knowing all the circumstances?

Option 1 Option 2

Continue

Screen 14

End of the Experiment

Time remaining: 1196 seconds

The experiment for today is now concluded.

Your total payout is **€3.25**. The breakdown of your payout is as follows:

Show-up fee	€1.00
Part 1	Not selected for payout
Part 2	Not selected for payout
Part 3	€2.25
Total Payout	€3.25

We sincerely thank you for your participation.

Please register in the following payout form within the next hour.

Continue

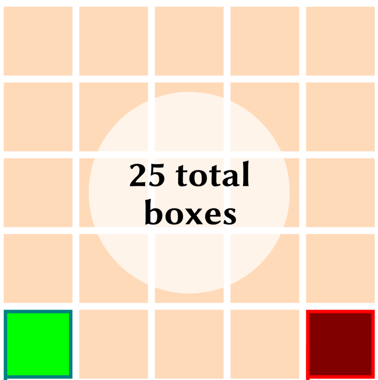
Appendix C

Appendix to Chapter 3

C.1 Survey items

C.1.1 Page 1

Imagine someone else who is about to play a simple game. We'll refer to this person as the 'player.' The player will participate in a two-day survey. On day 1 of the survey, the player will see 25 boxes:



① The player can open any number of boxes, from zero to 25. The player earns $20 for each opened box.
② However, one box contains a "curveball." The location of this curveball is completely random, and the player does not know which box it's in. The player knows there is a curveball somewhere.
③ If the player opens the box that contains the curveball, they will lose all their money.
④ Opening more boxes means a chance to win more money. However, it also increases the risk of opening the box containing the curveball.

Once	EndoDelay	Delay	Undo
On day 1, the player will learn the rules of the game and decide right away which boxes to open.	*(as in Once)*	On day 1, the player will learn the rules of the game and decide one day later (on day 2) which boxes to open.	On day 1, the player will learn the rules of the game and decide right away which boxes to open. One day later (on day 2), they are allowed to change their decision or stick with their initial decision. The boxes that are chosen will be opened on day 2.

In your personal opinion, how many boxes should the player open?

[Numeric input field, allowable values: 0, 1, …, 25]

Remember: The survey always takes two days. Assume that the player participates on both days. There is no way to shorten the survey.

What do you think: How many boxes will the average player ultimately open? Tell us your best guess.

[Numeric input field, allowable values: 0, 1, …, 25]

How happy do you think the average player will feel after making their choice? Rate the player's expected happiness on a scale:

[Numeric input field, Likert, radio buttons, allowable values: 1 (extremely unhappy) to 7 (extremely happy)]

Pagebreak

C.1.2 Page 2

Today, you are granted the power to either let the player have their choice or set a maximum number of boxes they can open.

Once	EndoDelay	Delay	Undo
Remember, on day 1, the player will learn the rules of the game and decide right away which boxes to open.	You will also decide when they are to make the choice. Remember, on day 1, the player will learn the rules of the game.	Remember, on day 1, the player will learn the rules of the game and decide one day later (on day 2) which boxes to open.	Remember, on day 1, the player will learn the rules of the game and decide right away which boxes to open. One day later (on day 2), they are allowed to change their decision or stick with their initial decision.

- If you want the player to be able to open any number of boxes, up to the maximum of 25, enter '25' into the field below.
- If you want to set a limit, enter the maximum number of boxes they should be able to open into the field below. Remember, the player can open anywhere from 0 up to the number you set.
- Remember that this game is just one part of a two-day survey that the player will take. They will learn the game rules on day 1. They will always be paid after day 2.

Your decision will be implemented for a real player if your response is randomly selected, so take it seriously. It could affect someone else.

Maximum number of boxes the player is able to open:

[Numeric input field, allowable values: 0, 1, ..., 25]

Once	EndoDelay	Delay	Undo
	Decide when they are to make the choice: [Radio button] The player is to make the choice right away (on day 1). [Radio button] The player is to make the choice only after one day (on day 2).		

C.2 Preregistered analyses

The preregistration may be perused at `https://osf.io/m7526`.

All tests are two-sided.

C.2.1 Analysis 1

Welch's two-sample t-test between Cap in Once *and* Cap in EndoDelay

	Cap in Once	Cap in EndoDelay
Sample size	639	690
Sample mean	17.358	17.691
Sample std.dev.	7.433	7.575
Difference in means	−0.333	
t	−0.808	
Degrees of freedom	1322.557	
p-value	0.419	
95% confidence interval	$[-1.141, 0.475]$	

C.2.2 Analysis 2

639 CAs are in Once. 690 are in EndoDelay.

In Once, 434 CAs (67.9%) set a cap of less than 25. In EndoDelay, 538 CAs (78.0%) set a cap of less than 25, imposed a waiting period or both.

Using a test of equal proportions, we find $\chi^2 = 16.6$, $p < 0.001$ (95% confidence interval for the difference: $[0.051, 0.150]$).

C.2.3 Analysis 3

Welch's two-sample t-test between Cap in Once *and* Cap in Delay

	Cap in Once	Cap in Delay
Sample size	639	717
Sample mean	17.358	17.688
Sample std.dev.	7.433	7.596
Difference in means	−0.329	
t	−0.806	
Degrees of freedom	1342.249	
p-value	0.421	
95% confidence interval	$[-1.131, 0.472]$	

C.2.4 Analysis 4

Welch's two-sample t-test between Cap in Delay *and* Cap in Undo

	Cap in Delay	Cap in Undo
Sample size	717	656
Sample mean	17.688	17.291
Sample std.dev.	7.596	7.784
Difference in means	0.396	
t	0.954	
Degrees of freedom	1353.618	
p-value	0.340	
95% confidence interval	$[-0.419, 1.212]$	

C.2.5 Analysis 5

Welch's two-sample t-test between Cap in Undo *and* Cap in Once

	Cap in Undo	Cap in Once
Sample size	656	639
Sample mean	17.291	17.358
Sample std.dev.	7.784	7.433
Difference in means	−0.067	
t	−0.159	
Degrees of freedom	1292.491	
p-value	0.874	
95% confidence interval	$[-0.897, 0.762]$	

C.2.6 Analysis 6

A Kolmogorov–Smirnov test between the norms in Once and Delay reveals $D = 0.041$, $p \approx 0.62$. This test meant as a robustness check, and not intended to reject (Grossmann, 2023, p. 14).

C.2.7 Analysis 7

Welch's two-sample t-test between Error in Once *and* Error in Delay

	Error in Once	Error in Delay
Sample size	628	696
Sample mean	5.005	4.398
Sample std.dev.	4.159	4.211
Difference in means	0.607	
t	2.635	
Degrees of freedom	1311.265	
p-value	0.009	
95% confidence interval	[0.155, 1.059]	

C.2.8 Analysis 8

Welch's two-sample t-test between Forecast happiness in Once *and* Forecast happiness in Delay

	Forecast happiness in Once	Forecast happiness in Delay
Sample size	639	717
Sample mean	4.354	4.413
Sample std.dev.	1.264	1.243
Difference in means	−0.059	
t	−0.867	
Degrees of freedom	1330.606	
p-value	0.386	
95% confidence interval	[−0.193, 0.075]	

C.2.9 Analyses 9

C.2.9.1 Analysis 9a

Welch's two-sample t-test between Cap in EndoDelay (Chooser in Delay) *and* Cap in Delay

	Cap in EndoDelay (Chooser in Delay)	Cap in Delay
Sample size	267	717
Sample mean	17.775	17.688
Sample std.dev.	7.734	7.596
Difference in means	0.088	
t	0.159	
Degrees of freedom	468.939	
p-value	0.874	
95% confidence interval	$[-0.997, 1.172]$	

C.2.9.2 Analysis 9b

Welch's two-sample t-test between Cap in EndoDelay (Chooser in Once) *and* Cap in Once

	Cap in EndoDelay (Chooser in Once)	Cap in Once
Sample size	423	639
Sample mean	17.638	17.358
Sample std.dev.	7.481	7.433
Difference in means	0.280	
t	0.599	
Degrees of freedom	899.650	
p-value	0.550	
95% confidence interval	$[-0.638, 1.198]$	

C.2.10 Analysis 10

639 CAs are in Once. 656 are in Undo. 717 are in Delay.

In Once, 205 CAs (32.1%) did not set a cap (i.e., they set a cap of 25). In Undo, 199 CAs (30.3%) did not set a cap. In Delay, 253 CAs (35.3%) did not set a cap.

A 3-sample test for equality of proportions returns $\chi^2 = 3.96$ (2 degrees of freedom), $p = 0.14$.

C.2.11 Adjusting for multiple hypothesis testing

p-value	Subset	A1	A2	A3	A4	A5	A7	A8	A9a	A9b	A10
Raw	—	0.419	0.000	0.421	0.340	0.874	0.009	0.386	0.874	0.550	0.138
Holm adj.	Main effects	1.000	0.000	1.000	1.000	1.000					
Holm adj.	Mechanisms						0.017	0.386			
Holm adj.	—	1.000	0.000	1.000	1.000	1.000	0.077	1.000	1.000	1.000	1.000

Table C.1: Adjusted *p*-values

Table C.1 adjusts *p*-values for multiple hypothesis testing using Holm's (1979) method. Analysis 6 (Section C.2.6 in this Appendix) is omitted because it was not meant to reject.

Analysis 2 (Section C.2.2 in this Appendix) relates to hypothesis 5. It survives even the most conservative adjustment, where all preregistered tests are jointly adjusted.

Analysis 7 (Section C.2.7 in this Appendix) relates to hypothesis 9. It does not survive the most conservative adjustment ($p = 0.077 > 0.05$). However, the t-test is revealed to be the most conservative among common tests. A Wilcoxon rank sum test on the same data reveals $p = 0.00028$ (conservatively adjusted: $p = 0.003$). A two-sample Kolmogorov–Smirnov test shows $p \approx 0.00031$ (conservatively adjusted: $p = 0.003$). Wilcoxon-type tests have been criticized in the literature (Divine et al., 2018). We can use Mood's median test to evaluate a null hypothesis of equal medians. When we do so, we find $p \approx 0.005$, conservatively adjusted to $p \approx 0.047$.

Moreover, we can group Undo and Delay, and Once and EndoDelay. Recall from Section 3.4.4 that Once and EndoDelay were identical on page 1, and that Undo and Delay both share deliberation. Note how model 4 in Table 3.1 shows how errors are slightly, though not significantly, lower in EndoDelay than in Once, making the overall grouped error closer to Once and Delay. Yet, after grouping, Welch's t-test returns $p = 0.0015$ (conservatively adjusted: $p = 0.013$). Wilcoxon and Kolmogorov–Smirnov tests show, after conservative adjustment, $p < 0.001$. A conservatively adjusted $p < 0.001$ also obtains in a Fisher's exact test on the null hypothesis that the proportion of errors equal to zero (model 1 in Table C.2 in this Appendix) is equal between the two groups of treatments.

The following table reports conservatively adjusted p-values from Welch's t-tests, Wilcoxon rank sum tests and Kolmogorov–Smirnov tests with all possible reasonable groupings of treatments:

Sample 1	Sample 2	Mean 1	Mean 2	Direction	Welch's t	Wilcoxon	K–S
Once	Delay	5.005	4.398	>	0.077	0.003	0.003
Once	Undo	5.005	4.262	>	0.018	0.000	0.000
Once	Delay & Undo	5.005	4.333	>	0.009	0.000	0.000
EndoDelay	Delay	4.715	4.398	>	1.000	0.423	1.000
EndoDelay	Undo	4.715	4.262	>	0.475	0.022	0.253
EndoDelay	Delay & Undo	4.715	4.333	>	0.465	0.035	0.421
Once & EndoDelay	Delay	4.854	4.398	>	0.183	0.011	0.035
Once & EndoDelay	Undo	4.854	4.262	>	0.039	0.000	0.001
Once & EndoDelay	Delay & Undo	4.854	4.333	>	0.013	0.000	0.000

From this evidence, we conclude that errors appear to be reduced through Chooser deliberation. However, it is a small effect.

C.3 Additional tables

	Model 1	Model 2
Intercept	0.168^{***}	0.072^{***}
	(0.010)	(0.020)
Delay or Undo	0.061^{***}	0.129^{***}
	(0.015)	(0.030)
Norm		0.010^{***}
		(0.002)
Delay or Undo × Norm		-0.007^{*}
		(0.003)
Outcome	Error is 0	Error is 0
R^2	0.006	0.018
Adj. R^2	0.006	0.016
Num. obs.	2649	2649

$^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$

Table C.2: Heterogeneous treatment effects on binary outcome

	Model 1	Model 2	Model 3	Model 4
Intercept	12.158^{***}	7.950^{***}	5.005^{***}	4.471^{***}
	(0.225)	(0.429)	(0.166)	(0.405)
Undo	-1.291^{***}	-1.802^{**}	-0.743^{**}	-1.326^{*}
	(0.325)	(0.602)	(0.240)	(0.566)
Delay	-1.024^{***}	-0.944	-0.607^{**}	-1.883^{***}
	(0.306)	(0.583)	(0.230)	(0.547)
EndoDelay	-0.531	-0.837	-0.290	-0.581
	(0.315)	(0.584)	(0.229)	(0.541)
Norm		0.439^{***}		0.055
		(0.045)		(0.043)
Undo × Norm		0.069		0.065
		(0.062)		(0.058)
Delay × Norm		-0.018		0.128^{*}
		(0.061)		(0.058)
EndoDelay × Norm		0.032		0.030
		(0.060)		(0.056)
Outcome	Belief	Belief	Error	Error
R^2	0.007	0.231	0.005	0.034
Adj. R^2	0.006	0.229	0.003	0.031
Num. obs.	2661	2649	2649	2649

$^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$

Table C.3: Heterogeneous treatment effects (without treatment grouping)

	Model 1	Model 2
Intercept	0.161***	0.062*
	(0.015)	(0.028)
Undo	0.084***	0.164***
	(0.022)	(0.044)
Delay	0.055*	0.112**
	(0.021)	(0.041)
EndoDelay	0.014	0.018
	(0.021)	(0.040)
Norm		0.010***
		(0.003)
Undo × Norm		−0.008
		(0.004)
Delay × Norm		−0.006
		(0.004)
EndoDelay × Norm		−0.000
		(0.004)
Outcome	Error is 0	Error is 0
R^2	0.007	0.019
Adj. R^2	0.006	0.016
Num. obs.	2649	2649

$^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$

Table C.4: Heterogeneous treatment effects on binary outcome (without treatment grouping)

Appendix D

Software and materials

Note: All URLs in this Appendix were last accessed January 2, 2025.

D.1 Software

The following specialist software was used in the preparation of this thesis:

- R 4.3.3 (R Core Team, 2024), *with packages*:
  - `aod` (Lesnoff et al., 2012)
  - `Cairo` (Urbanek and Horner, 2023)
  - `coin` (Hothorn et al., 2006, 2008), version 1.4-3
  - `dplyr` (Wickham, François, et al., 2023)
  - `kableExtra` (Zhu, 2024)
  - `lmtest` (Zeileis and Hothorn, 2002), version 0.9-40
  - `marginaleffects` (Arel-Bundock, Greifer, and Heiss, 2024), version 0.21.0
  - `sandwich` (Zeileis, 2004, 2006), version 3.1-0
  - `scales` (Wickham, Pedersen, and Seidel, 2023)
  - `texreg` (Leifeld, 2013)

  - `tidyr` (Wickham, Vaughan, and Girlich, 2024)
  - `VGAM` (Yee, 2015), version 1.1-11

  and their dependencies.

- Python 3.11.2,[1] *with packages*:
  - `numdifftools`,[2] version 0.9.41
  - `numpy` (Harris et al., 2020), version 1.26.4
  - `pandas` (McKinney, 2010; Pandas development team, 2020), version 2.2.0
  - `scipy` (Virtanen et al., 2020), version 1.12.0

  and their dependencies.

- Stata/MP 18.0, StataCorp LLC[3]
- SageMath 9.5[4] and Mathomatic 16.0.5[5] and their dependencies.
- oTree (Chen, Schonger, and Wickens, 2016) and uproot (Grossmann and Gerhardt, n.d.) and their dependencies.
- translate[6] and its dependencies (e.g., OpenAI et al., 2024, to translate instructions in Appendices A.5, B.5.1 and B.5.2).
- LuaLaTeX

D.2 Materials

The user icons in Figures 1 and 3.1 and Section B.5.2 derive from the (now defunct) Tango Desktop Project's icon library. These icons were released into the public domain and are available at Wikimedia Commons:

[1] `https://python.org`
[2] `https://github.com/pbrod/numdifftools`
[3] `https://www.stata.com`
[4] `https://www.sagemath.org`
[5] `https://orms.mfo.de/project@terms=&id=312.html`
[6] `https://gitlab.com/gr0ssmann/translate`

- https://commons.wikimedia.org/w/index.php?title=File:User_icon_2.svg&oldid=905903161
- https://commons.wikimedia.org/w/index.php?title=File:User_icon_2.svg&oldid=832009697

D.3 Software bibliography

Harris, C. R. et al. (2020). "Array programming with NumPy". *Nature*, 585(7825), 357–362. DOI: 10.1038/s41586-020-2649-2. URL: https://doi.org/10.1038/s41586-020-2649-2.

Hothorn, T., Hornik, K., van de Wiel, M. A., and Zeileis, A. (2006). "A Lego system for conditional inference". *The American Statistician*, 60(3), 257–263. DOI: 10.1198/000313006X118430.

— (2008). "Implementing a class of permutation tests: The coin package". *Journal of Statistical Software*, 28(8), 1–23. DOI: 10.18637/jss.v028.i08.

Leifeld, P. (2013). "texreg: Conversion of Statistical Model Output in R to LaTeX and HTML Tables". *Journal of Statistical Software*, 55(8), 1–24.

Lesnoff, M., Lancelot, and R. (2012). *aod: Analysis of Overdispersed Data*. R package version 1.3.3. URL: https://cran.r-project.org/package=aod.

McKinney, W. (2010). "Proceedings of the 9th Python in Science Conference". In: ed. by S. van der Walt and J. Millman, 56–61. DOI: 10.25080/Majora-92bf1922-00a.

Pandas development team (2020). *pandas-dev/pandas: Pandas*. DOI: 10.5281/zenodo.3509134. URL: https://doi.org/10.5281/zenodo.3509134.

R Core Team (2024). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. Vienna, Austria. URL: https://www.R-project.org/.

Urbanek, S. and Horner, J. (2023). *Cairo: R Graphics Device using Cairo Graphics Library for Creating High-Quality Bitmap (PNG, JPEG, TIFF), Vector (PDF,* *SVG, PostScript) and Display (X11 and Win32) Output*. R package version 1.6-2. URL: `https://cran.r-project.org/package=Cairo`.

Virtanen, P. et al. (2020). "SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python". *Nature Methods*, 17, 261–272. DOI: `10.1038/s41592-019-0686-2`.

Wickham, H., François, R., Henry, L., Müller, K., and Vaughan, D. (2023). *dplyr: A Grammar of Data Manipulation*. R package version 1.1.4. URL: `https://CRAN.R-project.org/package=dplyr`.

Wickham, H., Pedersen, T. L., and Seidel, D. (2023). *scales: Scale Functions for Visualization*. R package version 1.3.0. URL: `https://CRAN.R-project.org/package=scales`.

Wickham, H., Vaughan, D., and Girlich, M. (2024). *tidyr: Tidy Messy Data*. R package version 1.3.1. URL: `https://CRAN.R-project.org/package=tidyr`.

Yee, T. W. (2015). *Vector Generalized Linear and Additive Models: With an Implementation in R*. New York, USA: Springer.

Zeileis, A. (2004). "Econometric Computing with HC and HAC Covariance Matrix Estimators". *Journal of Statistical Software*, 11(10), 1–17. DOI: `10.18637/jss.v011.i10`.

— (2006). "Object-Oriented Computation of Sandwich Estimators". *Journal of Statistical Software*, 16(9), 1–16. DOI: `10.18637/jss.v016.i09`.

Zeileis, A. and Hothorn, T. (2002). "Diagnostic Checking in Regression Relationships". *R News*, 2(3), 7–10. URL: `https://CRAN.R-project.org/doc/Rnews/`.

Zhu, H. (2024). *kableExtra: Construct Complex Table with 'kable' and Pipe Syntax*. R package version 1.4.0. URL: `https://CRAN.R-project.org/package=kableExtra`.

Appendix E

Curriculum Vitae

Personal details

Address	Max R. P. Grossmann University of Cologne Albertus-Magnus-Platz 50923 Köln (Germany)
Website	*www.max.pm*
E-mail	*m@max.pm*
Telephone	+49 221 470-8667
Citizenship	Germany

Education

2020–2025 (expected)	Doctoral studies, Economics University of Cologne Advisors: Axel Ockenfels & Frederik Schwerter *Experimental and behavioral economics*

2017–2019	Master of Science, Economics University of Cologne Thesis supervisor: Jörg Breitung *Statistics and econometrics, behavioral economics*
2015–2017	Bachelor of Science, Business Administration University of Cologne Thesis supervisor: Axel Ockenfels *Finance, microeconomics*

Academic positions

2020–	**Lab Manager**, Cologne Laboratory for Economic Research & **Research Assistant**, Chair of Professor Axel Ockenfels, University of Cologne
2015–2019	**Student Research Assistant**, Chair of Professor Axel Ockenfels, University of Cologne

Publications in peer-reviewed journals

1. Duch, M. L., **Grossmann, M.** & Lauer, T. (2020). *z-Tree unleashed: A novel client-integrating architecture for conducting z-Tree experiments over the Internet.* Journal of Behavioral and Experimental Finance, Volume 28, December 2020. No. 100400.

2. Fišar, M., Greiner, B., Huber, C., Katok, E., Ozkes, A. & **Management Science Reproducibility Collaboration***. (2023). *Reproducibility in Management Science.* Management Science, 70(3):1343-1356.

* Consortium co-authorship.

Appendix F

Eidesstattliche Versicherung

gemäß § 9 Abs. 5 der Promotionsordnung im Promotionsprogramm Research in Management, Economics and Social Sciences der Wirtschafts- und Sozialwissenschaftlichen Fakultät der Universität zu Köln vom 1. August 2022:

Hiermit versichere ich an Eides statt, dass ich die vorliegende Dissertation selbstständig und ohne die Benutzung anderer als der angegebenen Hilfsmittel und Literatur angefertigt habe. Weitere Personen, neben den ggf. in der Arbeit aufgeführten Koautorinnen und Koautoren, waren an der inhaltlich-materiellen Erstellung der vorliegenden Arbeit nicht beteiligt. Alle Stellen, die wörtlich oder sinngemäß aus veröffentlichten und nicht veröffentlichten fremden Werken dem Wortlaut oder dem Sinn nach entnommen wurden, sind als solche kenntlich gemacht. Ich versichere an Eides statt, dass diese Dissertation noch keiner anderen Fakultät oder Universität zur Prüfung vorgelegen hat; dass sie - abgesehen von den angegebenen Teilpublikationen und eingebundenen Artikeln und Manuskripten - noch nicht veröffentlicht worden ist, sowie, dass ich eine Veröffentlichung der Dissertation vor Abschluss der Promotion nicht ohne Genehmigung des Promotionsausschusses vornehmen werde. Die Bestimmungen dieser Ordnung sind mir bekannt. Darüber hinaus erkläre ich hiermit, dass ich die Leitlinien der Universität zu Köln zur Sicherung guter wissenschaftlicher Praxis gelesen und sie bei der Durchführung der der Dissertation zugrundeliegenden Arbeiten und der schriftlich verfassten Dissertation beachtet habe und verpflichte mich hiermit, die dort genannten Vorgaben bei allen wissenschaftlichen Tätigkeiten zu beachten und umzusetzen. Ich versichere, dass die eingereichte elektronische Fassung der eingereichten Druckfassung vollständig entspricht. Ich versichere, dass ich nach bestem Wissen die reine Wahrheit gesagt und nichts verschwiegen habe.

M. R. P. G.