Beliefs about Choice Reversals

A pre-registration document with pilot data*

Alexander Coutts† Séverine Toussaert‡

April 14 2025

Abstract

In a dynamic effort choice experiment, we revisit the literature on time inconsistency to better understand the mapping between anticipated and actual choice reversals, and their link with decisions to constrain or expand future choice sets. While many studies document substantial naivety when contrasting an individual's predictions with their actual behavior, they typically require individuals to express degenerate beliefs, thus ignoring any uncertainty they might experience. In this project, we investigate the role of belief uncertainty in accounting for choice reversals, the accuracy of these beliefs, and their power to predict demand for commitment versus flexibility. This document offers a detailed exposition of our experimental design and planned analyses developed on pilot data.

Keywords: reversal, naivety, uncertainty, belief distribution, commitment, flexibility.

*This document outlines the complete set of pre-registered analyses planned for our full experiment (AEARCTR-0015407). For illustration, we present pilot results ($N = 58$), though the final results are likely to differ. Ethics approval was secured from the University of Oxford (ECONCIA20-21-22) and York University (2025-035). We gratefully acknowledge expert research assistance from Abdessamad Ait Mbarek, Axel La Pira, Stefania Merone, Hamid Tahir, and Matthew Tan.

†Schulich School of Business, York University; E-mail: acoutts@schulich.yorku.ca

‡University of Oxford; E-mail: severine.toussaert@economics.ox.ac.uk

1 Executive Summary

Background Many behavioral models assume that agents have *misspecified* beliefs about some preference or economic parameter of interest (e.g., value of some fundamental, own ability, strategic reasoning of other players, etc.). In other words, agents with wrong beliefs are wrong with probability one i.e., the truth is not contained in the support of their beliefs. This is the case for the most popular models of time inconsistency (O'Donoghue and Rabin, 1999; DellaVigna and Malmendier, 2004).

Understanding whether this assumption has bite is important both from a theory and an empirical point of view. Theory predictions sometimes crucially hinge on the fact that beliefs are misspecified. For instance, in learning models, this assumption implies that agents' beliefs might not approach the truth even asymptotically. In the context of time-inconsistent preferences, Heidhues and Kőszegi (2009) show that an agent's welfare when attempting to exercise self-control may dramatically vary depending on whether or not beliefs assign positive weight to true preferences. Relatedly, Spiegler (2011) shows that when firms can price discriminate, whether higher sophistication benefits consumers depends on the modeling of naivety.

On the empirical side, most experiments on dynamic inconsistency have neglected belief uncertainty by eliciting point predictions, in line with the models they aimed to test (Augenblick and Rabin, 2019; Fedyk, 2022). Several of these papers document striking results, with estimates of structural parameters indicating substantial (if not full) naivety. In addition, point beliefs about one's future behavior do not seem to predict demand for commitment devices in the literature, which is often interpreted as a hallmark of sophisticated present bias. This raises the question of whether contrasting point predictions with actual behavior is a sufficient statistic to elicit naivety.

To see the potential concern, consider a situation in which a decision maker (henceforth, DM) is tasked with predicting their own future effort. When prompted to make a point prediction, they report what they think is most likely, but they put only 40% on this outcome. With 60%, they expect a deviation to a lower effort level. Can we then say that this DM is naive? More generally, what can we say about the DM's naivety once their subjective uncertainty is accounted for? Our project aims to examine the extent to which failing to account for an individual's uncertainty might obscure our understanding of the relationship between their own behavior and beliefs.

What we do We conduct an online experiment with a design which closely follows standard protocols for studying time inconsistency. In a first stage, t_1, individuals decide how much work they would like to complete at a future date t_2 (seven days later). Then at t_2, individuals make the same decision for how much work they would like to complete immediately. Our focus is on the careful study of the beliefs that individuals have at t_1. For this reason, after making their initial work decision we elicit both participants' point beliefs about their future work decision as well as a full probability distribution over possible work amounts. We use this information to study (i) the extent to which individuals are uncertain about their choices at t_2; (ii) the interaction between beliefs and time inconsistency; (iii) the relationship between beliefs and preferences for commitment and flexibility. To study the latter, we allow individuals to *customize* the menu of possible work amounts they can choose from at t_2. This provides us with information on whether individuals are willing to pay to remove options (eliciting demand for commitment) and/or add options (eliciting demand for flexibility).

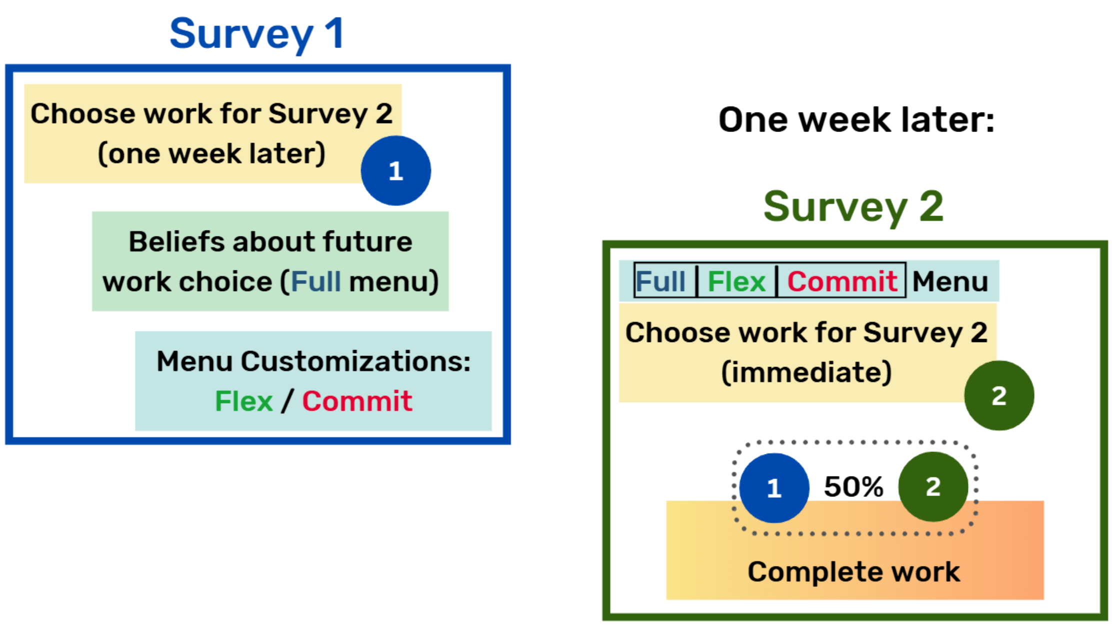
Research plan With this experimental dataset, we will first study whether we replicate prior evidence: focusing on point beliefs only, do individuals appear significantly naive? Second, we will document whether the restriction to point beliefs has significant bite: how certain are individuals about their future choices and do probabilistic beliefs more closely track actual behavior than point beliefs? Third, we will examine the predictive power of belief uncertainty for actual choice reversals and willingness to pay for flexibility vs. commitment. Finally, we will test whether eliciting beliefs (including making subjects consider their own uncertainty) affects their own behavior.

In Section 2 of this pre-analysis plan document, we describe our experimental design, including the methodological choices we made. In Section 3, we introduce the conceptual framework that we will use to formalize the relationship between beliefs, menu preferences, and future behavior. In Section 4, we fully describe the analyses we plan to present, using pilot data to illustrate what we anticipate to be our key results. Finally, Section 5 concludes with a discussion of our approach and additional evidence we plan to leverage to better interpret our main findings.

2 Experimental Design

We conduct an online experiment with a design that follows a standard protocol for studying time inconsistency. The full experimental instructions will be provided in an online appendix. In a first stage t_1, individuals decide how much work they would like to complete at a future date t_2 (exactly seven days later), from a discrete set of 10 options.[1] Then at t_2, participants make the same decision for how much work they would like to complete immediately. Beyond this, the experiment contains two key components. The first is that in our main treatment of interest, we elicit the full probability distribution of their beliefs about choosing every possible work amount. The second is that we study preferences for commitment and flexibility by allowing individuals to *customize* the menu they will see at t_2 regarding different work amounts. Figure 1 shows a simplified representation of the experimental design. Below we describe in turn the work task, the belief elicitation procedure, and the elicitation of preferences for commitment and flexibility.

Figure 1: Experimental design



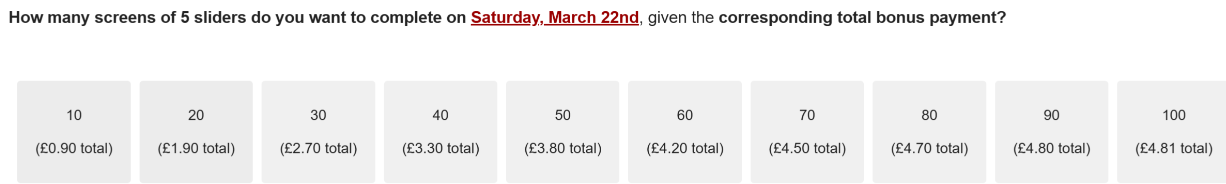
Notes: Simplified representation of the experimental design for the baseline treatment. In Survey 1, all participants entered beliefs and made two customizations to measure flexibility and commitment preferences. In Survey 2, participants make their work decisions for all three menus, with one menu implemented at random (80% = full,10%,10%).

2.1 The effort task

The effort task in question is a "slider task", which involves manually dragging sliders to various pre-specified integers between 0 and 500 (Gill and Prowse, 2019). Participants must choose how

[1]This ensures participation on the same day of the week (Augenblick, Niederle and Sprenger, 2015; Fedyk, 2022)

Figure 2: Effort choices



Notes: Decision screen at t_1 of the effort choices available to participants to complete one week later, at t_2 (here March 22nd). Each button corresponds to a number of screens to complete and the associated total payment.

many screens of 5 sliders they wish to complete for a given fixed payment. Options range from 10 to 100 screens in multiples of 10, amounting to completing between 50 and 500 individual sliders. Figure 2 shows the available options. Below we refer to a generic effort choice as $e \in \mathcal{E} := \{10, 20, ..., 100\}$, with corresponding monetary payment m_e. For instance, completing $e = 40$ screens earns a total payment m_e of £3.30.

Having a discrete choice set reduces the complexity of the instructions and the belief elicitation. Total payments were selected to exhibit decreasing marginal returns in order to shift choices towards an interior solution (Le Yaouanq and Schwardmann, 2022). Before making their choice, participants are required to complete two practice screens to familiarize themselves with the task, and are informed of the average time (in seconds) it took them to complete one screen. Before making their effort choice, they are shown a calculator allowing them to enter their expected time to complete one screen and obtain a (linear) projection of their total time and hourly wage for every possible effort choice.

2.2 Beliefs about future effort

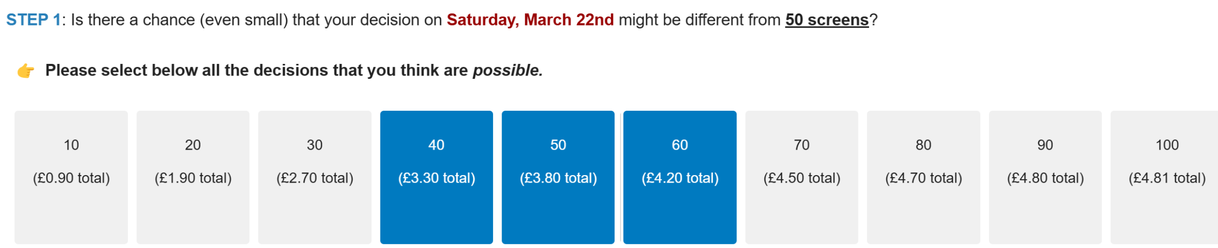
After making their effort choices, participants are asked to make a point prediction of the number of screens they will complete at t_2 by making a selection from the list of 10 choices; we denote this point prediction $\hat{e}_2 \in \mathcal{E}$. This exercise resembles current practices for eliciting beliefs about future effort. As prior work shows no impact of incentives in this context (Augenblick and Rabin, 2019; Fedyk, 2022), we chose not to financially incentivize point predictions.[2]

After the point predictions, we elicit participants' full belief distribution over possible effort levels. To do so, we ask them how certain they are about the number of screens they predict to

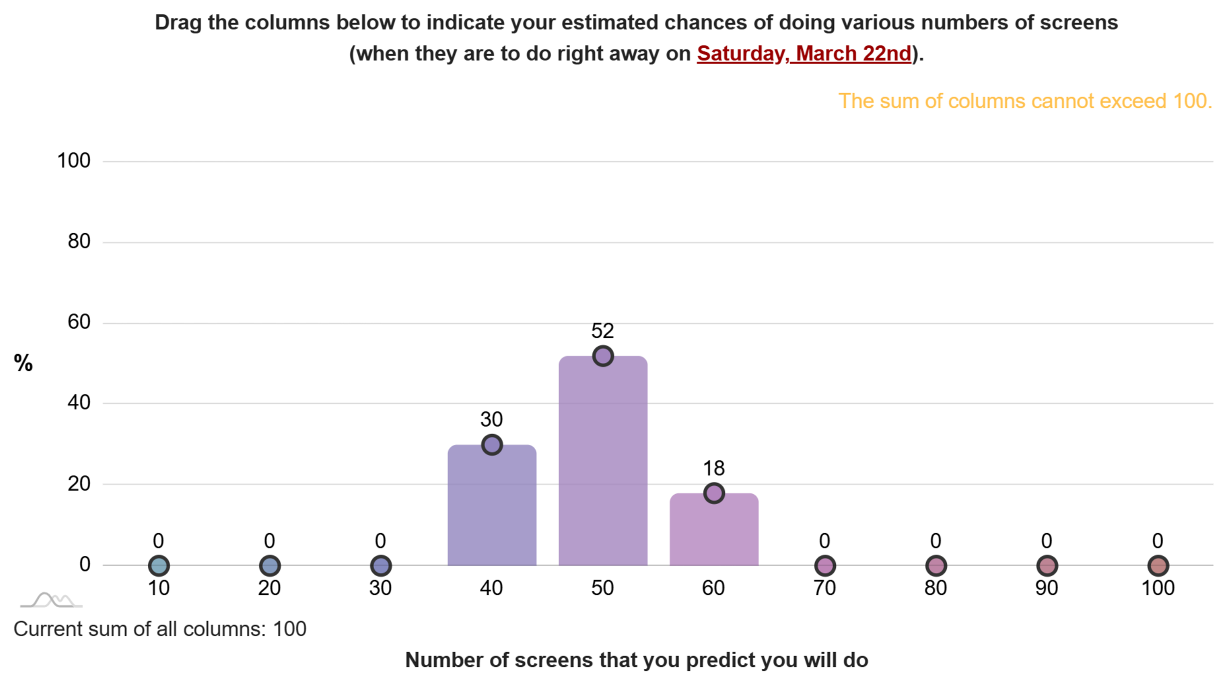
[2]In addition to reducing complexity, a further benefit of not incentivizing point predictions is that we do not have to specify the precise summary statistic (e.g., mean, median, or mode) to elicit. The exact wording of this question is: "When the date of [$t = 2$ *date*] arrives, what do you predict is the number of screens you will choose to do right away?". Given variation in how different participants might interpret this question, and as we also elicit the full distribution, we are able to identify which summary statistic of the distribution best fits the prediction each participant has in mind.

complete at t_2. Our elicitation procedure proceeds in two steps. In Step 1, shown in Figure 3a, participants are asked to select all decisions that they think are *possible*.

Figure 3: Elicitation of belief distribution



(a) Step 1: Choice of support



(b) Step 2: Full distribution

In Step 2, they are shown an interactive chart with different columns corresponding to the 10 possible decisions. They are instructed to indicate the chances that they will choose each decision indicated in Step 1, by dragging the corresponding columns, starting with the most likely decision(s). The interactive chart is displayed in Figure 3b. The elicitation of the belief distribution is incentivized using the Binarized Scoring Rule (BSR). Given the added complexity, here incentives were chosen as a means to engage participants to better understand what is being asked of them.[3] To minimize measurement error, participants are provided with a summary of their

[3]Following advice from Danz, Vesterlund and Wilson (2022), who show that salient information about incentives on the belief decision screen can lead to distortions, full details of the procedure are not shown by default to

estimated chances in a table format after completing the interactive chart; they are permitted to make any changes before final submission. In the following, we let $\hat{f}(e) \in [0, 1]$ denote the probability assigned to each effort e by the respondent.

2.3 Preferences for commitment and flexibility

After eliciting their beliefs, we offer to participants the possibility of customizing their menu of available effort decisions at t_2. Specifically, participants are told that with an 80% chance their choice from the full menu $\mathcal{E} = \{10, 20, ..., 100\}$ will be selected. However, with a 20% chance their choice will be selected from one of two menus that they can customize (10% chance for each).[4]

We refer to the first type of menu, $\mathcal{E}^+$, as eliciting preferences for flexibility. With this menu, participants start from a completely empty menu with no choices, but can choose to add options at a cost of £0.01 per option added. Note that they must add at least one option, so that they have something to choose at t_2. The first added option is always free. An example of menu $\mathcal{E}^+$ is shown in Figure 4a, where $\mathcal{E}^+ = \{50, 60\}$.

We refer to the second type of menu, $\mathcal{E}^-$, as eliciting preferences for commitment. With this menu, participants start from the full menu $\mathcal{E}$, but can choose to remove options at a cost of £0.01 per option removed. Note that they must leave at least one option, so that they have something to choose at t_2. An example of menu $\mathcal{E}^-$ is shown in Figure 4b, where $\mathcal{E}^- = \mathcal{E} \setminus \{30, 40\}$.[5]

At the end of Survey 1, we ask a set of debriefing questions to those who paid to remove options or paid to include options. These questions list a set of possible motives for commitment or flexibility, respectively.[6] In addition, participants who expressed any degree of uncertainty in their predictions are asked to select the reasons why they are uncertain. Finally, we ask questions about their upcoming availability for Survey 2, self-reported effort for customization decisions, and their perceived difficulty of completing the tasks.

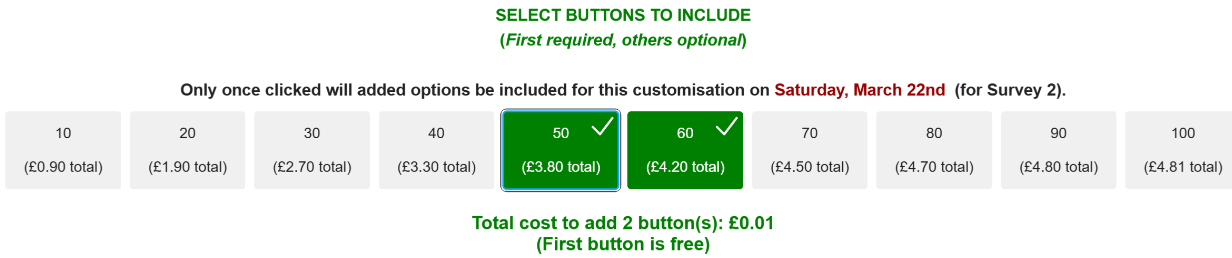
participants, but can be viewed in a pop-up upon clicking a button. Regarding the procedure, one column is selected at random and the BSR is implemented for that column. Payment is set at £0.20. See Hossain and Okui (2013) for further details on the procedure. Since participants are entering their beliefs for an event that they hold control over, this procedure could be in principle manipulated; however, the small payment was chosen to minimize the salience of any such distortions.

[4]Menu customization always occurs after belief elicitation because we were concerned that customization would generate confusion about the probabilities of different effort decisions.

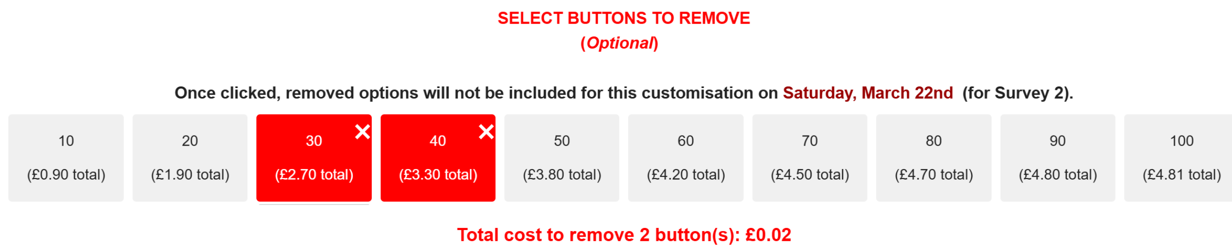
[5]For both types of menus, if a participant either removes their work choice at $t = 1$ or fails to add it, a pop-up message asks whether this was a mistake. The pop-up instructs participants that they can click "proceed anyway" if it was not a mistake, or go back to make changes.

[6]We further flag participants who stated certainty about making their effort choice (i.e., assigning 100% probability to one option), but added multiple options. These participants are asked in an open-ended question to explain this decision. Analogously, for participants who removed an option they assigned a 0% probability of choosing, we ask them why they did so. Finally, we include an open-ended question, which appears if participants added and removed the same option.

Figure 4: Preferences for commitment and flexibility



(a) Menu $\mathcal{E}^+$: Preferences for flexibility



(b) Menu $\mathcal{E}^-$: Preferences for commitment

Notes: Screenshot from experiment at t_1 (here t_1 is one week prior to t_2, i.e., August 30th) showing how participants can customize their menu (if selected) by (a) adding or (b) removing options. In this example, $\mathcal{E}^+ = \{50, 60\}$ and $\mathcal{E}^- = \mathcal{E} \setminus \{30, 40\}$.

2.4 Decisions in Survey 2

Survey 2 is completed exactly one week after Survey 1 and can be started any time between 7:00 am and midnight GMT.[7] The survey begins with a brief refresher of the decision environment. Participants are reminded of their practice time, and shown the same calculator from Survey 1 allowing them to view the total time and hourly wage for each possible work choice.[8] Next, we ask participants to make an effort choice from each of the three menus, with the full menu $\mathcal{E}$ presented first, and the two customized menus $\mathcal{E}^+$ and $\mathcal{E}^-$ shown in a randomized order, effectively implementing the strategy method.[9] After making their choices, we ask participants how their predictions in Survey 1 might have impacted their decisions in Survey 2 and what point prediction they recall having made then.

Following these questions, all the uncertainty is revealed. First, we select which of Survey 1 or

[7]Participants are made aware of this time-frame on the initial recruitment page and reminded again at the end of the survey. Additionally, participants are sent a message once the survey becomes available. They are sent up to three reminders (at 12:00, 17:00, and 22:00) conditional on not having started the survey.

[8]This ensures that participants have the same information as they did when making their t_1 decision. They do not complete additional practice screens in Survey 2, since they were already acquainted with the effort task, and we were concerned that additional learning about the task could bias effort choice.

[9]After submitting their choice from $\mathcal{E}$, respondents report how difficult they found making such a decision. Our data analysis focuses on choices from the full menu $\mathcal{E}$ (comparable across subjects and time), which we use to study choice reversals between t_1 and t_2.

Survey 2 decisions counts, with 50% chance of either. If Survey 2 is selected, the choice from one of the three menus is drawn at random, with 80% chance of the full menu $\mathcal{E}$, 10% chance of the flexible menu $\mathcal{E}^+$ and 10% chance of the commitment menu $\mathcal{E}^-$. Then the participant proceeds to do the number of screens selected from that menu. At the end of the survey, participants who exhibited choice reversals are asked to select the reasons for these reversals. We also ask a question to those participants who started the survey at a different time than they planned based on their reported availability in Survey 1.

2.5 Treatment condition without beliefs

Our main treatment of interest relates beliefs about future effort to (i) preferences for flexibility and commitment at t_1 and (ii) actual effort at t_2. Since predictions concern an event that participants have control over, one question is whether the very elicitation of beliefs could influence future behavior e.g., by creating a reference point for the participant. In addition, being asked to explicitly consider one's uncertainty could affect preferences for flexibility and commitment. To assess this, we consider a condition in which beliefs are not elicited. In other words, in Survey 1, participants only make their effort choice for Survey 2 and their customization decisions. We test whether making participants explicitly form beliefs has an impact on their subsequent decisions.

2.6 Sample size and recruitment

The experiment will be conducted with 1,000 UK respondents, 700 for the main treatment and 300 for the treatment without beliefs, using the online platform Prolific.[10] Survey 1 contains 5 mandatory comprehension questions and failure to answer any question correctly within two attempts will result in survey termination.[11] To receive any payment, participants are required to complete both surveys. Payments will be scheduled for one week after the completion of Survey 2 and include a show-up fee of £4. All our analyses will be restricted to the sample of individuals who completed both surveys.

Sample size justification We estimate that with $N = 1,000$ participants, 700 for the main treatment and 300 for the treatment without beliefs, we will have 80% power to detect a difference of 9 percentage points in the number of downward reversals from a baseline of 30%, which we hypothesized based on pilot data described in more details in Section 4. Based on this same pilot data, we anticipate between 10 and 15% attrition between Survey 1 and Survey 2. As a result, we

[10]We will require a minimum Prolific approval rate of 99%, a minimum of 10 previous submissions, and a maximum of 1,000 previous submissions. We will enable Prolific's option to balance participant gender distribution.

[11]Comprehension checks ensure understanding of core elements of the design such as the timing of the slider task and the importance of every decision.

will aim to recruit a total of 1,150 participants.[12] A summary of our piloting strategy and how it informed our design and analytical decisions will be provided in an appendix.

2.7 Expert forecasts

To benchmark some of our results, we collect forecasts from academic experts on the Social Science Prediction Platform (https://socialscienceprediction.org/). Potential forecasters are invited to participate in a 15-minute survey. Respondents have to meet the following inclusion criteria: (i) being a Ph.D. candidate, postdoctoral researcher, or (assistant/associate/full) professor; (ii) working in the field of economics and/or psychology; (iii) not having heard about the study yet. The forecast data collection started on February 27, 2025 i.e., before our main data collection and we will leave the survey open until April 30, 2025. Access to the forecast data is restricted during this collection period. We will use the forecasts as long as we managed to collect answers from at least 30 respondents; we set this parameter based on a review of several recent papers, which document that using crowds of 5 to 10 forecasters already provides informative signals on the current state of knowledge by significantly reducing the influence of extreme individual forecasts (DellaVigna and Pope, 2018; Otis, 2022; Iacovone, McKenzie and Meager, 2023).

Forecasters are first introduced to the experimental design and presented with the actual decisions faced by participants, which they can themselves experience before submitting their forecasts. We collect a total of 21 forecasts of the following quantities: (i) proportion of actual upward and downward reversals; (ii) proportion of predicted upward and downward reversals, based on point predictions and based on belief distributions; (iii) proportion of participants who express uncertainty and average support size of the belief distribution; (iv) proportion who assigned 0% chance to their actual choice (by reversal status); (v) proportion of reversals for participants who expressed vs. did not express uncertainty; (vi) proportion of participants demanding flexibility and commitment; (vii) fraction of added (removed) options that participants assigned a positive chance to choosing; (viii) impact of belief elicitation on Survey 2 effort (categorical variable for statistical significance in two-sided t-test of equality between belief and no-belief groups, with categories $p < 0.01$, $p \in [0.01, 0.05)$, $p \in [0.05, 0.1)$, $p \geq 0.1$).

To incentivize forecasting accuracy, we will randomly select one of the 21 forecasts and corresponding outcome; the 5 forecasters closest to our sample estimate will have a prize of £100 donated to a charity of their choice (donations on a voluntary basis and ties broken randomly).

We will report forecasts descriptively as well as conduct one-sample proportion tests (for

[12] If attrition exceeds 20%, we will present sensitivity analyses for our main results in an appendix assuming that all attrited subjects would have reversed their choice (upper bound) or none would have done so (lower bound). We will also test whether non-response to Survey 2 is predicted by belief uncertainty (support size of the belief distribution), anticipated reversals ($1 - \hat{f}(e_1)$), and demand for flexibility/commitment (intensive and extensive margin)

dichotomic outcomes) or t-tests (for continuous outcomes) of the null hypothesis that the experimental results coincide with the mean forecast treated as a constant. In so doing, we treat our sample of forecasters as our population of interest.

3 Conceptual framework

3.1 General setup

Following models of menu choice (Dekel, Lipman and Rustichini, 2001; Dekel et al., 2007; Gul and Pesendorfer, 2001), we consider a decision maker (DM henceforth) who must decide at t_1 on the effort option(s) $E \subseteq \mathcal{E}$ available to them at some future period t_2. The DM faces two considerations. First, the DM faces uncertainty about future shocks $s \in S$ that could affect their utility of completing the effort task. For concreteness, we assume that these shocks affect the cost $c_s(e)$ of producing effort level e. At t_1, the DM forms beliefs $\mathbf{p} \in \Delta(S)$ about these future shocks, where p_s is the probability of state s. Subjective uncertainty about future shocks creates a preference for flexibility at t_1, that is, the DM would prefer (ceteris paribus) for a larger set E of options to be available at t_2 in order to tailor their effort level e to the realized shock s. At the same time, the DM may face the temptation at t_2 to complete a different number of tasks than they would like at t_1 (regardless of the state of the world). Temptation creates a preference for commitment at t_1, i.e., the DM would prefer (ceteris paribus) for a smaller set E of options to be available at t_2. Importantly, we allow the DM to be tempted either by lower effort (at the expense of a lower payment) or by larger payments (requiring a higher effort).

Utility function We assume that the DM's preferences at t_1 over the menu E of effort levels available at t_2 admit the following representation

$$W(E) = \sum_{s \in S} p_s \left[\max_{e \in E} [U_s(e) + V_s(e)] - \max_{e' \in E} V_s(e') \right]$$

The first utility term U_s is the "normative" (free-of-temptation) utility U_s of choosing a given effort level $e \in E$ if state s is realized; we assume that it is given by $U_s(e) = m_e - c_s(e)$ (i.e., the difference between the monetary payment and effort cost). In line with models of costly self-control (Gul and Pesendorfer, 2001; Dekel, Lipman and Rustichini, 2009), we interpret the second utility term V_s as the "temptation utility" of choosing effort $e \in E$. We assume it is given by $V_s(e) = m_e - \gamma c_s(e)$ where $\gamma > 0$ may magnify the effort cost relative to the monetary gain ($\gamma > 1$), capturing a temptation to do less, or weaken it ($\gamma < 1$), capturing a temptation to do more.[13] The

[13]As an alternative framing, one could assume $V_s(e) = \beta_m m_e - \beta_e c_s(e)$ where $\beta_m, \beta_e \in [0, 1]$ are present bias

difference $V_s(e) - \max_{e' \in E} V_s(e') < 0$ corresponds to the self-control cost of resisting $e' \in E$ when choosing e instead. For simplicity and to avoid having to deal with ties, we assume that for each $s \in S$, the utilities U_s, V_s, and $U_s + V_s$ have all a unique maximizer. To allow uncertainty to play a non-trivial role, we require that there is no statewise-dominant option $\bar{e}_U = \text{argmax}_{e \in \mathcal{E}} U_s(e)$ or $\bar{e}_{U+V} = \text{argmax}_{e \in \mathcal{E}} U_s(e) + V_s(e)$ for all $s \in S$ (i.e., choices must be responsive to the state).

Effort chosen at t_1 vs. t_2 At t_1, the DM must determine what effort level e_1 they would like to complete at t_2, before any uncertainty is realized. In the language of the model, the DM is asked to commit to a singleton option $\{e_1\}$ where $e_1 \in \mathcal{E}$ and $\{e_1\}$ maximizes $W(\{e\}) = \sum_{s \in S} p_s U_s(e)$ (since $V_s(e) - \max_{e' \in E} V_s(e') = 0$ for all $s \in S$ when $E = \{e\}$). In the experiment, we compare e_1 to the effort level e_2 chosen at t_2 from the full set of options $\mathcal{E}$. The DM's choice e_2 is the effort which solves $\max_{e \in \mathcal{E}} U_s(e) + V_s(e)$ under the realized state s. We say that a *choice reversal* between t_1 and t_2 occurs if $e_2 \neq e_1$; otherwise, we say that choice is *time consistent* ($e_2 = e_1$).

In this framework, choice reversals may occur for two reasons. First, in the presence of uncertainty ($|S| > 1$), e_1 could be the subjective expected utility maximizing choice without being ex post optimal, given the realized state s.[14] Second, temptation may create choice reversals depending on the size of γ. When $\gamma > 1$, *downward reversals* ($e_2 < e_1$) may occur with positive probability, while when $\gamma < 1$, *upward reversals* may occur ($e_2 > e_1$).

Optimal menu(s) The DM's preferences over menus depend on the size of the state space $|S|$ and probability distribution $\mathbf{p}$ over states, and the influence of temptation γ. To see the role of these various forces, consider the case where $|S| = 1$ (no uncertainty) and $\gamma = 1$ (no temptation), so that $W(E) = \max_{e \in E} U_s(e)$. In this case, the DM faces no flexibility or commitment concerns and will value equally any menu that contains the effort level e_1 (maximizing U_s over $\mathcal{E}$). Thus, $\mathcal{E}^+ = \{e_1\}$, $\mathcal{E}^- = \mathcal{E}$, and $\{e_1\} \sim \mathcal{E}$. Now suppose that $\gamma = 1$ but $|S| > 1$, so that $W(E) = \sum_{s \in S} p_s \max_{e \in E} U_s(e)$. In addition, suppose that $\mathbf{p}$ has full support on S and that each $e \in \mathcal{E}$ is the unique maximizer of some utility in $\{U_s\}_{s \in S}$. In this case, $\mathcal{E}^+ = \mathcal{E}^- = \mathcal{E}$, i.e., the DM strictly

parameters; in this case, $\beta_m < 1$ and $\beta_e = 1$ would correspond to $\gamma > 1$, while $\beta_m = 1$ and $\beta_e < 1$ would correspond to $\gamma < 1$. For parsimony, we take γ as deterministic. A generalization of the model would allow the temptation parameter to be stochastic, $\gamma \sim F_{[\underline{\gamma}, \overline{\gamma}]}$ for some distribution function F with $\underline{\gamma} < 1$ and $\overline{\gamma} > 1$, allowing the DM to be tempted by lower and higher effort levels at different times.

[14] In fact, not only might the DM deviate from e_1 with positive probability, they might do so with probability one, even with perfectly well-calibrated beliefs. This is the case if e_1 acts as a compromise option that, although not optimal under any state at t_2, minimizes expected losses at t_1 from the inability to tailor choices to the state. For instance, suppose $S = \{s_1, s_2\}$, $p = \text{P}\{s = s_1\}$, $e = 10$ maximizes $U_s(e)$ under $s = s_1$, while $e = 90$ maximizes $U_s(e)$ under $s = s_2$. The DM will choose $e_1 = 50$ (that is, prefer $\{50\}$ to both $\{10\}$ and $\{90\}$ at t_1) provided that $\frac{U_{s_1}(10) - U_{s_1}(50)}{U_{s_2}(50) - U_{s_2}(10)} < \frac{1-p}{p} < \frac{U_{s_1}(50) - U_{s_1}(90)}{U_{s_2}(90) - U_{s_2}(50)}$, but then deviate at t_2 to $e_2 = 10$ with probability p and to $e_2 = 90$ with probability $1 - p$.

prefers to keep all options in the set.[15] More generally, there is a direct relationship between the DM's subjective uncertainty and preference for flexibility, with the size of $\mathcal{E}^+$ growing as S grows (Dekel, Lipman and Rustichini, 2001). Next, suppose that $|S| = 1$ and $\gamma \neq 1$, i.e., the DM only faces temptation concerns, with $W(E) = \max_{e \in E}[U_s(e) + V_s(e)] - \max_{e' \in E} V_s(e')$ as in Gul and Pesendorfer (2001). In this case, $\mathcal{E}^+ = \{e_1\}$ and $\mathcal{E}^- \subseteq \mathcal{E}$. Which options are removed to form $\mathcal{E}^-$ depends on whether $\gamma > 1$ (lower effort levels removed) or $\gamma < 1$ (higher effort levels removed). Finally, suppose $|S| > 1$ and $\gamma \neq 1$. In this case, the DM may exhibit both a preference for flexibility ($|\mathcal{E}^+| > 1$) and a preference for commitment ($|\mathcal{E}^-| < |\mathcal{E}|$).

3.2 Example

To illustrate the model mechanics, we consider a simple parametrized example. Suppose that $S = \{s_1, s_2\}$, $p = \mathrm{P}\{s = s_1\}$ and $c_s(e) = \frac{1}{1000}e^s$, so that s governs the convexity of the effort cost function. We assume $s_1 = 2$ and $s_2 = 1.6$. Table 3 shows how the chosen e_1, $\mathcal{E}^+$, $\mathcal{E}^-$ and e_2 change as a function of p and the temptation parameter γ.

In the absence of any uncertainty ($p \in \{0, 1\}$) and any temptation ($\gamma = 1$), the DM simply chooses the effort level $e^* = e_1 = e_2$ that maximizes their utility U_s under the unique state s. If $s = s_1$, the effort cost is fairly high at $c_s(e) = \frac{1}{1000}e^2$, yielding $e^* = 30$; if $s = s_2$, the effort cost is only $c_s(e) = \frac{1}{1000}e^{1.6}$ and $e^* = 70$. Because $p \in \{0, 1\}$, the DM has no preference for flexibility i.e., $\mathcal{E}^+ = \{e_1\}$, and since $\gamma = 1$, they have no preference for commitment i.e., $\mathcal{E}^- = \mathcal{E}$. When $p = 0.5$, the DM has flexibility motives, $\mathcal{E}^+ = \{30, 70\}$; in addition, when forced to select an option at t_1, the DM chooses $e_1 = 40$, a compromise between 30 and 70. Now consider the impact of temptation. When $\gamma = 3$, the DM overweights the cost of effort, choosing $e_2 = 20$ when $s = s_1$ and $e_2 = 60$ when $s = s_2$ i.e., a downward deviation by 10 tasks in both states relative to the no-temptation case ($\gamma = 1$). To minimize the impact of temptation, the DM chooses to remove (a subset of) lower effort levels. By contrast, when $\gamma = 0.05$, the DM underweights the cost of effort relative to the monetary payment, choosing $e_2 = 50$ when $s = s_1$ and $e_2 = 80$ when $s = s_2$ i.e., an upward deviation by 20 tasks when $s = s_1$ and 10 tasks when $s = s_2$. As a result, the DM now prefers to exclude higher effort levels. For the intermediate cases $\gamma = 1.7$ and $\gamma = 0.9$, the DM exhibits a weaker preference for commitment (removing fewer options) and choices at t_2 exhibit no deviation due to temptation in either state. When uncertainty is low (e.g., $p = 0.9$), the DM favors at t_1 the effort level that is most likely to be chosen at t_2 (i.e., $e_2 = 30$ under $s = s_1$) and choices are time consistent with high probability (i.e., $e_2 = e_1$ with probability $p = 0.9$).

[15] This assumes that the cost of adding options is sufficiently low relative to the perceived benefits of flexibility; otherwise, the DM may choose to only add a subset of options. In addition, note that if $|S| < |\mathcal{E}|$ (i.e., the size of the state space is smaller than the set of options), then $|\mathcal{E}^+| = |S|$ (again, assuming that no two subjective states have the same maximizer); in particular, the number of subjective states that can be revealed in the utility representation of

Table 3: Chosen menus and effort for different parameter values

Parameter values		**Menu and effort choices**			
p	γ	e_1	$\mathcal{E}^+$	$\mathcal{E}^-$	e_2 (from $\mathcal{E}$)
1	1	30	$\{30\}$	$\mathcal{E}$	30
0	1	70	$\{70\}$	$\mathcal{E}$	70
0.5	1	40	$\{30, 70\}$	$\mathcal{E}$	30 with $p = 0.5$, else 70
1	3	30	$\{30\}$	$\mathcal{E} \setminus \{10, 20\}$	20
0	3	70	$\{70\}$	$\mathcal{E} \setminus \{30, 40, 50, 60\}$	60
0.5	3	40	$\{30, 60\}$	$\mathcal{E} \setminus \{10, 20, 40, 50\}$	20 with $p = 0.5$, else 60
1	0.05	30	$\{30\}$	$\{10, 20, 30\}$	50
0	0.05	70	$\{70\}$	$\mathcal{E} \setminus \{80, 90, 100\}$	80
0.5	0.05	40	$\{40\}$	$\{10, 20, 30, 40\}$	50 with $p = 0.5$, else 80
1	1.7	30	$\{30\}$	$\mathcal{E} \setminus \{20\}$	30
0	1.7	70	$\{70\}$	$\mathcal{E} \setminus \{60\}$	70
0.9	1.7	30	$\{30, 70\}$	$\mathcal{E} \setminus \{20, 60\}$	30 with $p = 0.9$, else 70.
1	0.9	30	$\{30\}$	$\mathcal{E}$	30
0	0.9	70	$\{70\}$	$\mathcal{E} \setminus \{80\}$	70
0.9	0.9	30	$\{30, 70\}$	$\mathcal{E} \setminus \{80\}$	30 with $p = 0.9$, else 70.

Notes: The menu that maximizes W is not unique since the DM is unaffected (whether positively or negatively) by the presence of certain options. Within the set of maximizers, $\mathcal{E}^+$ corresponds to the smallest menu and $\mathcal{E}^-$ to the largest one (since adding/removing options presents no additional benefits but is costly). The menus $\mathcal{E}^+$ and $\mathcal{E}^-$ are optimal assuming that the cost of commitment/flexibility (1 cent per option removed/added) is sufficiently low relative to the added benefits; if not, then $\mathcal{E}^+ = \{e_1\}$ and $\mathcal{E}^- = \mathcal{E}$.

3.3 Link between menu preferences and beliefs

We now articulate testable implications that emerge from the model relating preferences over menus $\mathcal{E}^+$ and $\mathcal{E}^-$ to anticipated choice $\hat{f}(e)$ from $\mathcal{E}$.

Preference for flexibility and uncertainty In this model, the value of flexibility is purely instrumental: it may allow the DM to achieve better material consequences by conditioning their effort choice on the realized state. In other words, there is no intrinsic value to choosing from a larger choice set e.g., because the DM has a pure preference for autonomy or decision rights (Sen, 1988; Bartling, Fehr and Herz, 2014; Ferreira, Hanaki and Tarroux, 2020). One implication is that if the DM prefers to retain access to multiple options, they must expect to choose each of these options in some circumstance (Ahn and Sarver, 2013; Dean and McNeill, 2020). Therefore, if a DM

the DM's preferences is bounded above by the number of choices available to the DM.

has a preference for flexibility ($|\mathcal{E}^+| > 1$), they must have non-degenerate beliefs ($|\text{supp}(\hat{f})| > 1$); in particular, they must expect a choice reversal with positive probability ($\hat{f}(e_1) \neq 1$).[16] A partial converse is also true: a DM who faces uncertainty will pay to add options to their choice set if (i) the cost of customization (£0.01 per added option) is judged low enough; (ii) temptation concerns are sufficiently mild (otherwise, they may counterbalance flexibility motives e.g., when $p = 0.5$ and $\gamma = 0.05$ in Table 3). This leads to the following hypothesis:

H1: *Preference for flexibility is positively associated with subjective uncertainty and anticipated choice reversals.*

We will test this hypothesis in three ways. First, we will conduct a test of proportions to assess whether those who choose to pay for additional options ($|\mathcal{E}^+| > 1$) are more likely to report non-degenerate beliefs ($|\text{supp}(\hat{f})| > 1$), and thus to anticipate a choice reversal ($\hat{f}(e_1) \neq 1$), than those who don't. Second, to examine the intensive margin, we will test in an exploratory analysis whether there is a positive correlation between the number of options added, $|\mathcal{E}^+|$, and the size of the support of the DM's belief distribution, $|\text{supp}(\hat{f})|$, as well as the perceived probability of reversal, $1 - \hat{f}(e_1)$.[17] Finally, we will test whether the following property holds

$$\text{If } e \in \mathcal{E}^+ \text{ then } \hat{f}(e) > 0.$$

In words, this property states that a DM who chooses to include a given option e in their customized menu $\mathcal{E}^+$ must expect to choose it with positive probability from the full set $\mathcal{E}$.[18] We coin this property linking preference for flexibility to expected consequences ĉ-flex. We note that ĉ-flex may be violated in this framework if temptation is sufficiently large (e.g., when $p = 0.5$ and $\gamma = 3$ in Table 3). We will test this property both at the individual and aggregate levels.

Preference for commitment and uncertainty Unlike choice set expansions, choice set restrictions ($\mathcal{E}^- \subset \mathcal{E}$) may happen in this framework for two reasons. First, as with flexibility concerns, the DM may choose commitment because they anticipate the possibility of a choice reversal ($\hat{f}(e_1) \neq 1$). This is the only commitment motive in models of present bias (Laibson,

[16] Note that this is true as long as choices at t_2 are responsive to the state; if not, one could have a case where $|\mathcal{E}^+| > 1$ (revealing that the DM perceives at least two states $s_1, s_2 \in S$ and effort levels $e, e' \in \mathcal{E}^+$ such that $U_{s_1}(e) > U_{s_1}(e')$ and $U_{s_2}(e') > U_{s_2}(e)$), but the DM expects to choose some $\bar{e} \in \mathcal{E}$ at t_2 (maximizing $U_s + V_s$) regardless of $s \in S$. In such a case, $|\mathcal{E}^+| > 1$ but $\hat{f}(\bar{e}) = 1$.

[17] We refer to this analysis as "exploratory" because theoretical predictions depend on the precise probability assigned to each e.

[18] This property is related, but different, from Axiom A1 of Ahn and Sarver (2013) and the consequentialism property tested in Dean and McNeill (2020). Using the notations from this paper, their axiom states that if $E \cup \{e\} \succ E$ then $f^{E\cup\{e\}}(e) > 0$, where $f^{E\cup\{e\}}$ is the actual choice probability of e from $E \cup \{e\}$. By contrast, our property relates a DM's preferred menu $\mathcal{E}^+$ to anticipated choice $\hat{f}(e)$ from the full menu $\mathcal{E}$ (where $\hat{f}(e)$ is understood as $\hat{f}^{\mathcal{E}}(e)$).

1997; O'Donoghue and Rabin, 1999), where utility solely depends on material consequences. As noted earlier, such reversals may happen if the temptation parameter γ sufficiently deviates from 1 (e.g., when $\gamma = 3$ or $\gamma = 0.05$ in Table 3). Second, the DM may demand commitment despite anticipating their choice to be time consistent so as to eliminate the cost of resisting tempting options (Gul and Pesendorfer, 2001; Toussaert, 2018). This may happen if temptation is mild (e.g., when $\gamma = 1.7$ or $\gamma = 0.9$ in Table 3). An important implication is that, although positive, the link between anticipated choice reversals and commitment demand is likely to be weaker than with demand for flexibility. The experience of subjective uncertainty ($|\text{supp}(\hat{f})| > 1$) itself interacts with demand for commitment: when the DM expects to make different choices in different states, the options that are particularly tempting change with the state; as a result, the number of excluded options may increase with the amount of uncertainty experienced, i.e., $|\mathcal{E}^-|$ may go down (e.g., comparing cases $p \in \{0, 1\}$ vs. $p = 0.9$ in Table 3 when $\gamma = 1.7$).[19]

H2: *Preference for commitment is positively associated with subjective uncertainty and anticipated choice reversals, albeit less strongly than preference for flexibility.*

We will perform the symmetric counterpart of all tests formulated for the case of flexibility and contrast our findings. In particular, we propose to test the following counterpart to ĉ-FLEX linking choice removal to expected consequences

$$\text{If } e \notin \mathcal{E}^- \text{ then } \hat{f}(e) > 0$$

which we coin ĉ-COMMIT for symmetry. In words, this property requires that a DM who chooses to exclude an option e from their customized menu $\mathcal{E}^-$ must expect to choose it with positive probability from the full set $\mathcal{E}$. This property should be satisfied by consequentialist models such as models of present bias, but may be violated under costly self-control.

3.4 Link between beliefs and actual behavior

We now turn to the consistency between choice expectations $\hat{f}(e)$ and actual behavior $f(e)$, a connection that is necessary to identify naivety.

Naivety and model misspecification We allow for the possibility that the DM holds inaccurate beliefs about some aspect of the model. For instance, the DM might misperceive the distribution of future states, i.e., $\hat{\mathbf{p}} \neq \mathbf{p}$. Alternatively, the DM might form inaccurate beliefs about γ.

[19]This is not a general feature however, because temptation and flexibility concerns enter in conflict. For instance, when $\gamma = 0.05$ in Table 3, $|\mathcal{E}^-|$ is higher (the number of excluded options is smaller) when $p = 0.5$ than when $p = 1$. In such a case, the DM has no preference for flexibility ($\mathcal{E}^+ = \{40\}$ when $p = 0.5$). This also illustrates a positive link between preference for flexibility and preference fo commitment.

Regardless of the precise nature of this misperception, the consequence is that the DM's beliefs $\hat{f}(e)$ about effort at t_2 may differ from the actual choice probability $f(e)$. We propose to examine two classes of tests to assess the extent to which choice reversals may be mispredicted by the DM i.e., *misspecification* and *calibration* tests.

First, we say that the DM's beliefs are *misspecified* if $\hat{f}(e_2) = 0$, i.e., they assign zero probability to the effort level that is eventually chosen. We perform misspecification tests at the individual level and assess the extent to which choice reversals are associated with misspecified beliefs. For instance, in deterministic models of present bias with partial naivety, the DM is assumed to have misspecified beliefs $\hat{\beta}$ about their present bias parameter β (i.e., $\hat{\beta} \neq \beta$ with probability one), in turn implying a possible misspecification of beliefs about actual behavior.

Second, we say that the DM's beliefs are *miscalibrated* if $\hat{f}(e) \neq f(e)$ for some $e \in \mathcal{E}$, i.e., actual choice probabilities differ from expected choice probabilities. Since we observe only one decision at t_2 for each individual, we cannot estimate individual choice probabilities; therefore, we perform calibration tests at the aggregate level. If beliefs are well-calibrated, then among all individuals who assign $x\%$ to choosing a given effort level e, the fraction who end up choosing this effort level should be approximately $x\%$. Looking at it another way, among well-calibrated individuals, the proportion who choose a given effort level e should match their average belief about choosing this effort level. In practice, we perform the latter test by taking single iid draws from each participant's belief distribution, which we compare to actual effort choices, thus allowing us to make comparisons between similar objects (actual vs. predicted effort).[20]

Even if somewhat miscalibrated, probabilistic beliefs could be highly predictive of choice reversals, suggesting a certain degree of sophistication. We assess this by testing whether the proportion of actual choice reversals differs depending on whether the possibility of choice reversal was anticipated or not ($\hat{f}(e_1) \neq 1$ vs. $\hat{f}(e_1) = 1$). We also compare the distribution of $\hat{f}(e_1)$ among those who reversed their choice downward ($e_2 < e_1$) or upward ($e_2 > e_1$) to those who did not ($e_2 = e_1$).

Menu preferences and actual choice behavior Building on the above calibration tests, we test the counterparts of ĉ-FLEX and ĉ-COMMIT for actual choice behavior:

$$\text{If } e \in \mathcal{E}^+ \text{ then } f(e) > 0.$$

$$\text{If } e \notin \mathcal{E}^- \text{ then } f(e) > 0.$$

[20]Note that performing calibration tests at the aggregate level is a valid approach under the assumption that the shocks faced by each DM in the population are drawn iid e.g., they do not face correlated shocks leading all of them to revise their choice downwards, a reasonable assumption in our experimental design.

which we call C-FLEX and C-COMMIT. We conduct these tests at the aggregate level by calculating among all individuals who chose to include (exclude) a given effort level e the proportion who ended up choosing it from $\mathcal{E}$.

4 Results

Below we present a sketch of our main results using data from a pilot conducted on Prolific in November 2023 with a total of 81 participants. Our analysis sample is restricted to the 58 respondents for whom we have complete data.[21] We make use of this small sample to generate our main tables and figures and provide tentative results, which will be updated upon collection of our main dataset. Information to be completed or amended manually is indicated below in square brackets in gray (some of the statistics will be automatically updated). The full replication code is available on OSF at `https://osf.io/79x6v/`. Additional analyses not presented here will be clearly labeled as unregistered and exploratory in the final version of the paper.

We document three classes of empirical findings, focusing on the relationship between: point predictions and choice reversals (Section 4.1); belief uncertainty and choice reversals (Section 4.2); beliefs, actual choices, and preference for commitment vs. flexibility (Section 4.3).

4.1 Choice reversals and point predictions

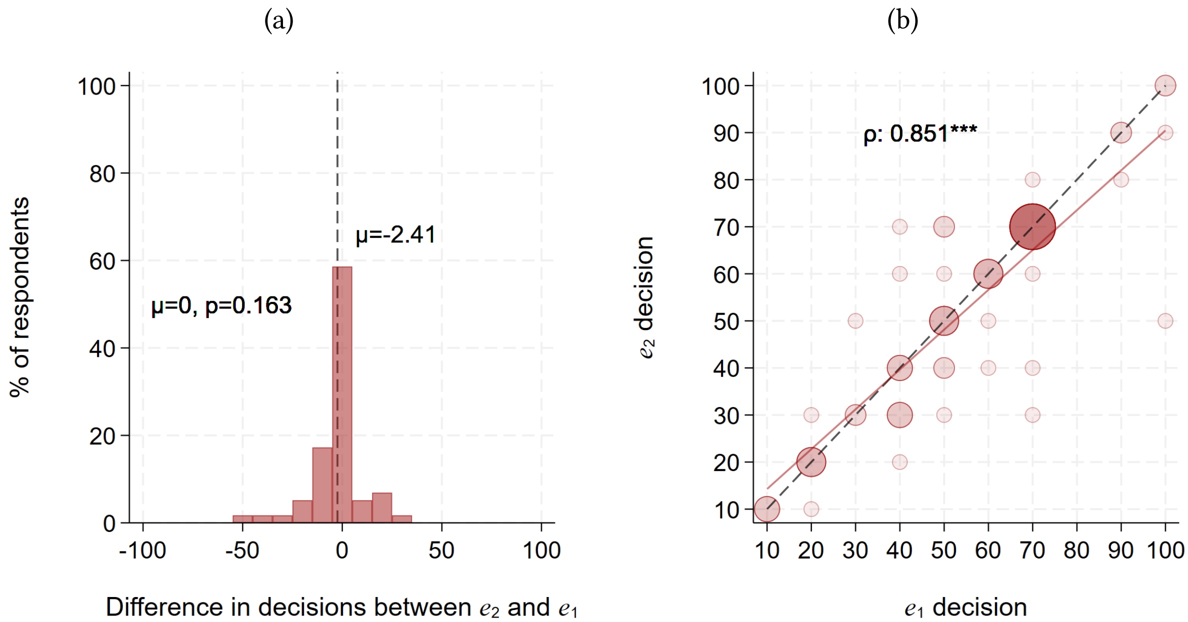
We first examine the extent to which participants exhibit choice reversals between t_1 and t_2, and mispredict their future behavior when considering point predictions only. Appendix Figure B.1 presents the full distribution of effort choices at both dates.

Result 1. *In line with the literature, participants exhibit a (small) aggregate tendency to lower down their effort choices between t_1 and t_2. In particular, 59% of participants make time consistent choices, while 28% (14%) revise their choice downward (upward).*

Figure 5a shows that the size of deviations is generally small. The mean number of tasks chosen at t_1 is 53.6, compared to 51.2 at t_2, an average deviation of 2.4 (Paired t-test p-value = 0.163). Figure B.2 shows the average effort level at t_2, conditional on every possible effort level chosen at t_1.

[21] A total of 71 participants returned for Survey 2 (attrition of 12%). Table A.1 presents comparisons of effort and beliefs for participants who did not return for survey 2, versus those who completed both surveys. All our analyses are restricted to the sample of individuals who (i) completed both surveys; (ii) made a choice from the full menu in Survey 2 (58). Restriction (ii) will not apply for our main data collection because we will use the strategy method to obtain a choice from the full menu for each respondent in Survey 2, while respondents in the pilot made a choice from only one menu (full menu or one of the two customizations), using the direct response method. As a reference point, Survey 1 & 2 took an average of 19 and 35 minutes respectively. Total earnings were £7.72 (£8.58 per hour), including a show-up fee of £4.

Figure 5: Deviations of decision e_2 from decision e_1



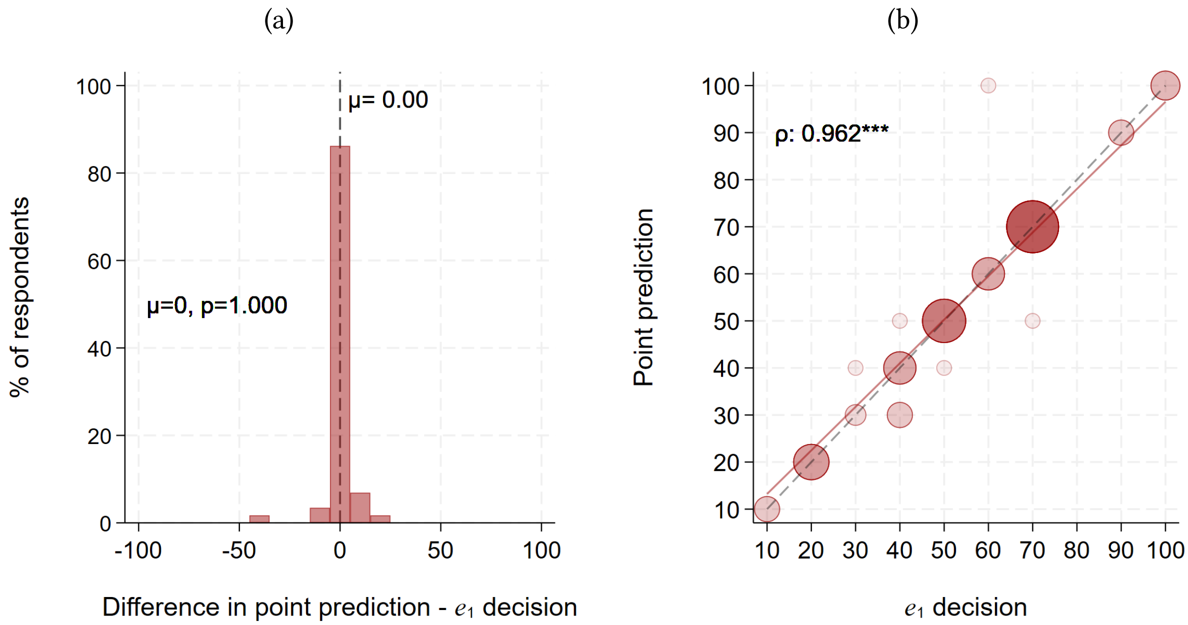
Notes: (a) Histogram of the distribution of differences $e_2 - e_1$ with a vertical bar for the mean and p-value of paired t-test $H_0 \ : \ e_2 - e_1 = 0$. (b) Scatter plot of e_2 against e_1 with Pearson correlation coefficient. Smallest bubble refers to N=1 respondent(s). Largest bubble refers to N=10 respondent(s).

Deviations are the largest for [effort levels to be specified]. As shown in Figure 5b, effort decisions display a high level of stability between the two weeks. However, 41% of participants deviate from the choice made at t_1 and 67% of these deviations are below the 45° line, corresponding to a downward reversal. [Comment on whether experts anticipate relatively more downward reversals and whether they are accurate in levels.]

Result 2. *Point predictions are closely aligned with effort decisions made at t_1: 86% of participants predict $\hat{e}_2 = e_1$, while only 59% actually choose e_1 at t_2. As a result, point predictions exhibit a slight upward bias i.e., participants tend to overestimate their future effort.*

As shown in Figure 6a, the mean deviation of the point prediction $\hat{e}_2$ from the effort decision e_1 is very close to, and not significantly different from, zero. The correlation between the two is nearly equal to 1 (Figure 6b). Combining Results 1 & 2, this means that point predictions are upwardly biased, with a mean over-prediction of 2.4 (Figure B.3a) (95% CI: [-1.01, 5.84]); the correlation between $\hat{e}_2$ and e_2 is weaker than the one between $\hat{e}_2$ and e_1 and only slightly above the correlation between e_2 and e_1 (Figure B.3b). One question is how predictions of reversals (comparing $\hat{e}_2$ to e_1) relate to actual reversals (comparing e_2 to e_1). Figure 7 shows that among those who predicted being time consistent, roughly 62% made a correct prediction, with a relatively even number who ended up reversing their choice downward and upward. Among those who ended

Figure 6: Deviations of point prediction $\hat{e}_2$ from decision e_1

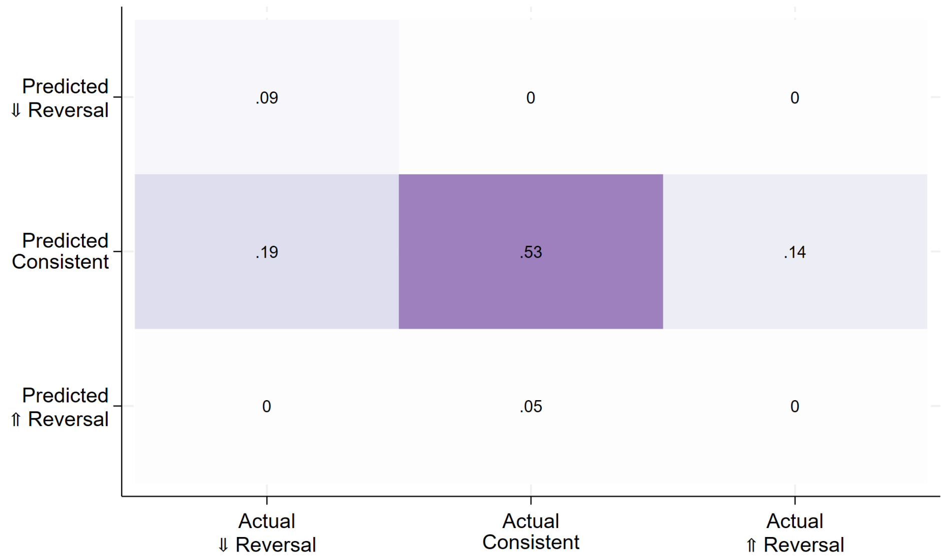


(a) Histogram of the distribution of differences $\hat{e}_2 - e_1$ with a vertical bar for the mean and p-value of paired t-test $H_0 : \hat{e}_2 - e_1 = 0$. (b) Scatter plot of $\hat{e}_2$ against e_1 with Pearson correlation coefficient. Smallest bubble refers to N=1 respondent(s). Largest bubble refers to N=13 respondent(s).

being time consistent, 91% behaved in line with their beliefs, while actual reversals appear mostly unanticipated. Interestingly, predictions of downward reversals show a high degree of accuracy. [To add: discussion of whether the experts predict an underestimation of downward/upward reversals from respondents.]

Summary [Tentative] We largely replicate previous findings in the literature. First, on average, we find a downward bias in the effort chosen at t_2 relative to t_1. Second, these reversals appear largely unanticipated, with point predictions equaling the effort chosen at t_1 for most participants, suggesting near complete naivety as in Augenblick and Rabin (2019) and Fedyk (2022). In the next section, we ask whether these inferences continue to hold after accounting for participants' uncertainty about their future effort choice.

Figure 7: Heatmap of predicted vs. actual reversals



Notes: Heatmap for the estimated probability mass function of the joint distribution P(predicted reversal = j, actual reversal = k) for j and k in $\{-1, 0, 1\}$, where -1 refers to a downward reversal ($\hat{e}_2 < e_1$ for predicted and $e_2 < e_1$ for actual), 0 to a time consistent choice ($\hat{e}_2 = e_1$ for predicted and $e_2 = e_1$ for actual), and 1 refers to an upward reversal ($\hat{e}_2 > e_1$ for predicted and $e_2 > e_1$ for actual). Numbers in each square refer to the respective frequencies.

4.2 Accounting for subjective uncertainty

4.2.1 Can subjective uncertainty be ignored?

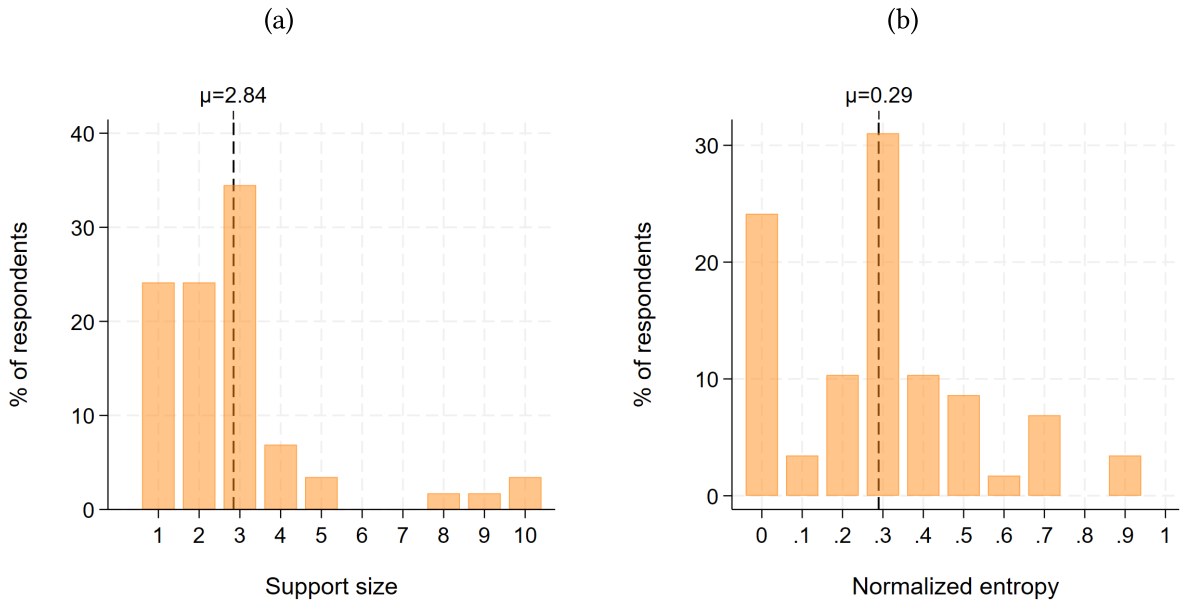
Result 3. *When asked to form predictions using a belief distribution, only 24% express certainty. Point predictions are thus not a sufficient statistic for measuring expectations.*

Figure 8a shows the number of effort levels that participants deemed possible in the sense that $\hat{f}(e) > 0$.[22] On average, participants included 2.8 effort levels in the support of their beliefs (median of 3.0). Figure 8b shows that participants exhibited moderate uncertainty regarding the effort level they might choose among those belonging to the support. Among those who reported degenerate beliefs (i.e., $|\text{supp}(\hat{f})| = 1$), 93% assign all the mass to their chosen effort at t_1 (i.e., $\hat{f}(e_1) = 1$), which is what theory would predict provided that temptation is not perceived to be too large (i.e., for sufficiently small departures of $\hat{\gamma}$ from 1 - see Table 3).

Importantly, non-degeneracy means that the vast majority of participants do anticipate a potential reversal (although they might underestimate the likelihood or size of these reversals).

[22]The support is calculated on the basis of the belief distribution (which corresponds to Step 2 in the design). 5% of participants selected an effort in Step 1, but not Step 2. On the other hand, 19% of participants selected an effort in Step 2, but not Step 1. [We will present robustness analyses using Step 1 to calculate the support. See Section 5 for a more detailed exposition.]

Figure 8: Belief Uncertainty



Notes: (a) Histogram of the distribution of support size $|\text{supp}(\hat{f})|$ with a vertical bar for the mean. (b) Histogram of the distribution of normalized entropy $\left(-\frac{1}{\ln 10}\sum_{j=1}^{10}\hat{f}(e_j)\ln\hat{f}(e_j)\right)$ with a vertical bar for the mean.

As a benchmark, we test whether academic experts anticipate the level of uncertainty expressed by participants: on average, experts predicted that [X%] would express any uncertainty (support larger than one), with a predicted mean support size of [X%] [Appendix figure to be added]. In our subsequent analyses, we use support size as our main measure of subjective uncertainty, which we correlate with choice reversals and demand for flexibility and commitment; we discuss the corresponding analyses using normalized entropy in an appendix.[23]

Result 4. *The point prediction receives 67% mass on average. However, 28% of participants assign less than 50% chance to their point belief.*

Figure B.4 shows a quantile plot of $\hat{f}(\hat{e}_2)$, the probability mass assigned to $\hat{e}_2$ across all participants. The point prediction corresponds to (one of) the mode(s) for 88% of participants, to the median for 81% of participants, and to the mean (rounded to the nearest effort choice) for 74% of participants.

Result 5. *The expression of subjective uncertainty is [not] a significant predictor of choice reversals.*

[23] Although clearly distinct, these two measures are highly correlated with each other, and with alternative measures of uncertainty such as the range or the standard deviation of the belief distribution, see Table B.1. See also Table B.2 for summary statistics about the belief distributions.

Table 2 shows how the likelihood that a participant exhibits a reversal changes depending on whether they expressed uncertainty about their future behavior ($|\text{supp}(\hat{f})| = 1$ or > 1). The proportion of time consistent participants is [(not) significantly lower] among those who expressed uncertainty. [Discussion of whether our sample of forecasters anticipated a difference and test of the hypothesis that the observed difference is equal to the forecasted difference.] We observe [no/some] asymmetry in the propensity to exhibit upward vs. downward reversals depending on belief uncertainty.[24] Figure B.5 examines whether the chance of choice reversal depends not only on the extensive margin of belief uncertainty, but also on the intensive margin, showing [no/some] evidence of the latter.

To go one step further, we examine how the distribution of $\hat{f}(e_1)$ (perceived chances of a time consistent choice) differs for participants who ended up being time consistent ($e_2 = e_1$) compared to those who reversed their choice either downward ($e_2 < e_1$) or upward ($e_2 > e_1$). Figure 9a shows quantile plots of $\hat{f}(e_1)$ for each reversal type. We observe a [near first-order stochastic dominance] relationship in the distribution of probability weights assigned to e_1, with time consistent participants generally showing higher confidence ex ante that they will choose e_1 (Kolmogorov-Smirnov test p-value: 0.161). The point-biserial correlation between $\hat{f}(e_1)$ and time consistency at t_2 (indicator for $e_2 = e_1$) is 0.29. Thus, probabilistic beliefs are highly predictive of reversal status.[25]

Table 2: Conditional probabilities of reversal

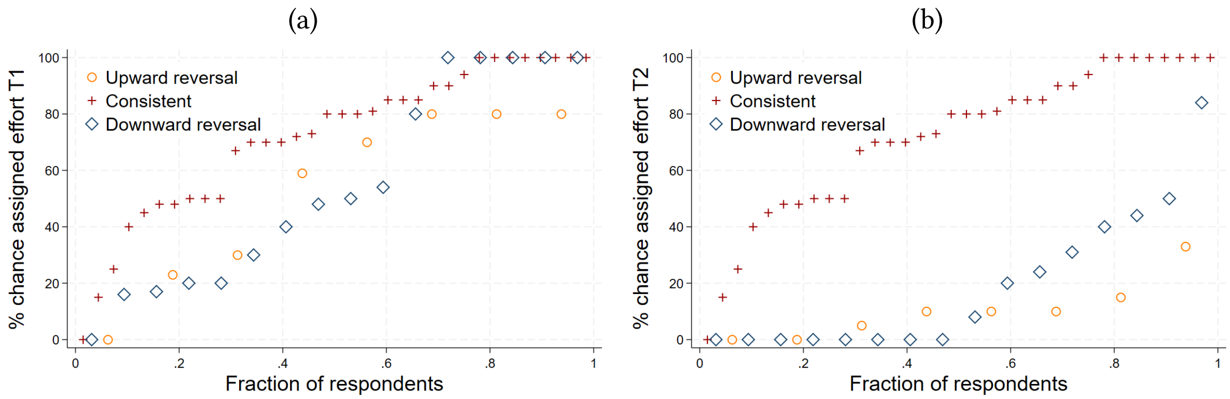
	Consistent	**Upward reversal**	**Downward reversal**
No belief uncertainty	64.3%	0.0%	35.7%
Belief uncertainty	56.8%	18.2%	25.0%
p-value	0.621	0.086	0.435

Notes: Chances of a consistent choice ($e_2 = e_1$), upward reversal ($e_2 > e_1$), or downward reversal ($e_2 < e_1$) conditional on exhibiting either no belief uncertainty ($|\text{supp}(\hat{f})| = 1$) or some uncertainty ($|\text{supp}(\hat{f})| > 1$); exact p-values from proportions tests. $N = 58$.

[24] Since $\hat{f}(e_1) = 1$ for virtually all participants with degenerate beliefs, findings are nearly the same for reversal probabilities conditional on whether $\hat{f}(e_1) = 1$ or $\hat{f}(e_1) \neq 1$.

[25] To help with the interpretation of the size of this correlation, we examine the statistic we would obtain if we took a single draw from participants' belief distributions and assigned this draw as their realized effort level, thus assuming that participants behave in a way that is fully compatible with their model of the world. This correlation is 0.56.

Figure 9: Distributions of the belief probability mass assigned to effort decisions e_1 and e_2



Notes: (a) Quantile plot across all participants of the belief distribution probability mass $\hat{f}(e_1)$ assigned to e_1, by reversal type. (b) Corresponding quantile plot for $\hat{f}(e_2)$.

4.2.2 Naivety after accounting for subjective uncertainty

Result 6. *(Misspecification): Only 19% of participants have misspecified beliefs, meaning that they assign zero probability to their actual effort. However, while only 3% of time consistent individuals are misspecified, the proportion for downward and upward reversals is 50% and 25% respectively.*

Figure 9b shows the distribution of $\hat{f}(e_2)$ for each reversal type. The probability mass assigned to the effort eventually chosen at t_2 is generally low for participants who exhibited either a downward or an upward reversal. This finding suggests that a non-trivial fraction of individuals have a misspecified model of the world.[26] We find that forecasters [under/over/correctly] estimate the proportion of participants with misspecified beliefs. [quantile plot of expert predictions by reversal status in appendix]

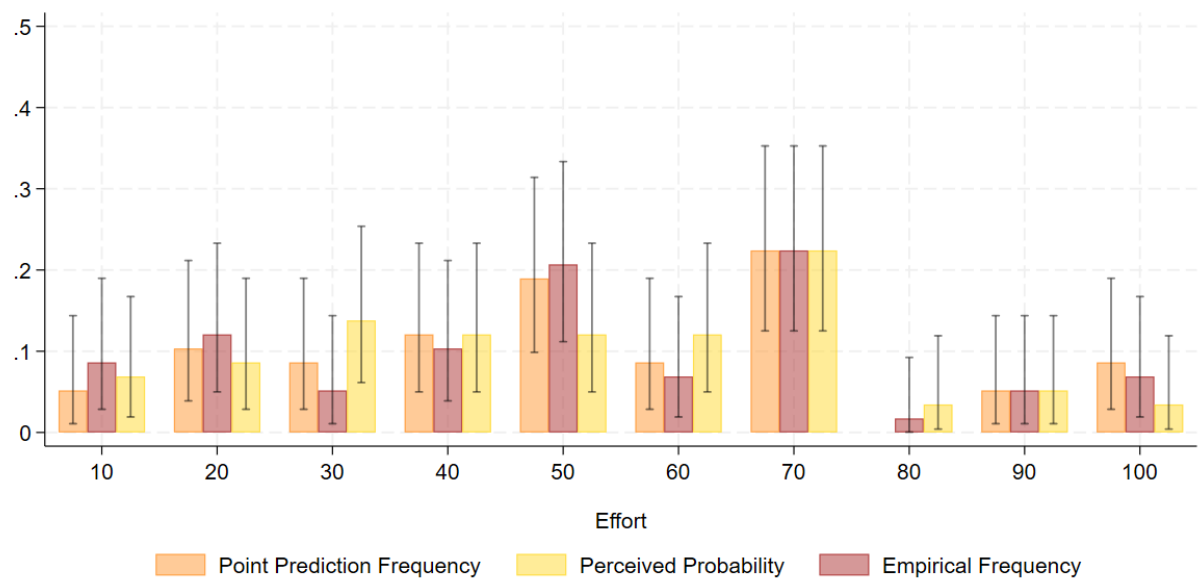
Result 7. *(Miscalibration): In the aggregate, participants' beliefs are miscalibrated to a varying degree across effort levels. Accounting for subjective uncertainty entails a [significant/small] reduction in the gap between predictions and actual effort.*

Figure 10 compares the fraction of participants who expected to choose a given effort level according their point belief (yellow bars), the corresponding fraction inferred when taking one draw from their belief distribution (orange bars), and the actual fraction who chose this effort level (maroon bars). The amount of miscalibration is [significant/small], especially for [effort levels to be inserted]. Nevertheless, as with point predictions, predictions based on central features of the

[26]Examining statistical differences by type, we find that the fraction of misspecified individuals is significantly different for downward and upward reversals compared to consistent individuals (proportions test p-values = 0.029, 0.000, respectively). Comparing misspecification for individuals with downward versus upward reversals, the corresponding proportions test p-value is 0.242.

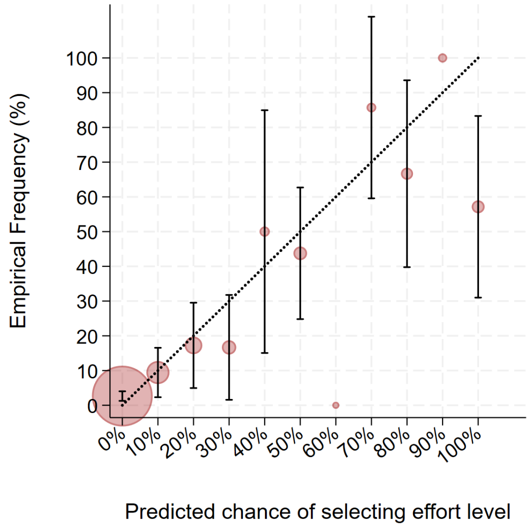
elicited belief distributions e.g., the mean, median or mode(s) remain slightly biased upward, with the bias being less pronounced for [least biased predictor to be added] (see Table B.2). Figure 11 shows that issues of miscalibration are more frequent for effort levels that were deemed [X, Y, Z%] likely, suggesting particular distortions for events perceived as [low/medium/high] probability.

Figure 10: Examining calibration over effort levels



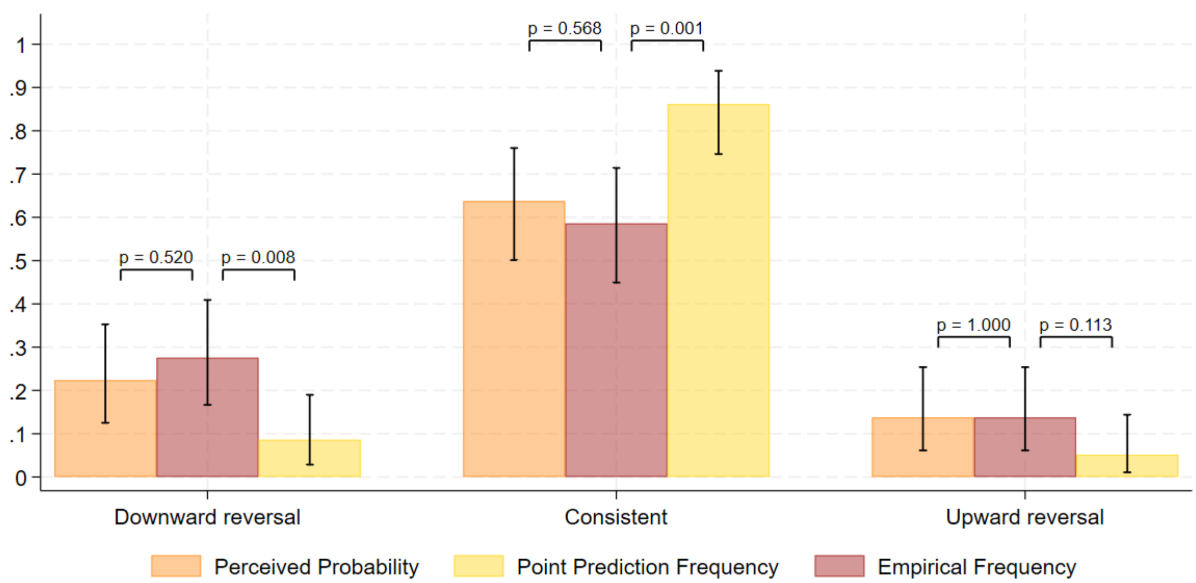
Notes: Predicted frequency of given effort level using random draw from belief distribution (orange), and point predictions (yellow) vs. empirically observed frequency (maroon). Binomial exact 95% confidence intervals shown.

Figure 11: Calibration of beliefs against effort



Notes: Empirically observed frequency of choosing a given effort level, conditional on the probability assigned to choosing that effort level. GLM-estimated binomial model with cluster-robust 95% confidence intervals shown. Each predicted chance is ±5 percentage points (e.g., 20% refers to $[15, 25]$%), excluding the end points of 0% and 100% which are exact; hence predictions in the intervals $(0, 5)$ and $(95, 100)$ are dropped (applies to 1% of all predictions). Bubble size is proportional to sample size.

Figure 12: Predicted vs. Actual Frequency of Reversals



Notes: Predicted frequency of reversal using random draw from belief distribution (orange), and point predictions (yellow) vs. empirically observed frequency of reversal (maroon). Binomial exact 95% confidence intervals shown. Two-sided p-values are calculated from tests of proportions.

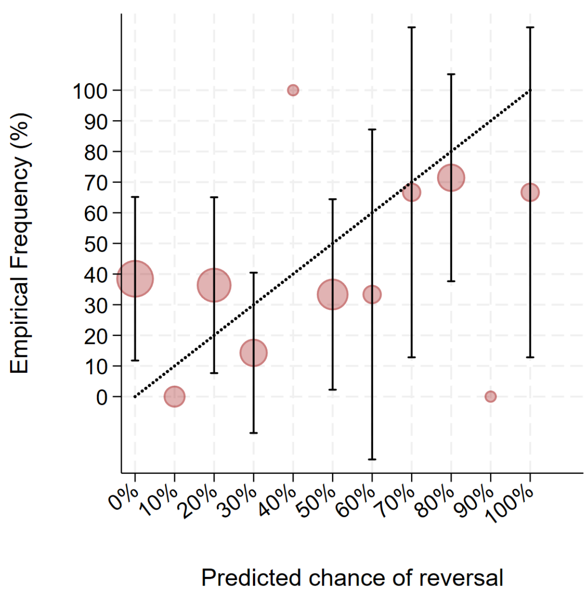
Result 8. *(Prediction of reversals) Despite the above distortions, beliefs about the possibility of re-*

versal are well-calibrated after accounting for subjective uncertainty.

Figure 12 contrasts the proportion of actual choice reversals (maroon bars) with beliefs about choice reversals when using (i) point predictions (yellow bars); (ii) predictions based on draws from the elicited belief distributions (orange bars). While point predictions provide clear underestimates of the proportion of both downward and upward reversals, the proportions based on the elicited belief distributions are very close to the observed proportions. [To add: discussion of corresponding figure for forecasters and assessment of whether they are well-calibrated.]

To gain further insights, Figure 13 contrasts $1 - \hat{f}(e_1)$ with $1 - f(e_1)$. For instance, among participants whose belief distribution $\hat{f}$ assigned 0% chance to choosing e_1, 38% actually chose e_1 at t_2. Overall, beliefs about (non)-reversals appear [well calibrated]. As an exploratory analysis, Figure B.6 splits this analysis into upward and downward reversals, finding [no/some] differences.

Figure 13: Predicted vs. Actual Frequency of Reversal



Notes: Empirically observed frequency of reversal $1 - f(e_1)$, conditional on the probability assigned to any reversal $1-\hat{f}(e_1)$. GLM-estimated binomial model with cluster-robust 95% confidence intervals shown. Each predicted chance is ±5 percentage points (e.g., 20% refers to $[15, 25]$%), excluding the end points of 0% and 100% which are exact; hence predictions in the intervals $(0, 5)$ and $(95, 100)$ are dropped (applies to 0% of participants). Bubble size is proportional to sample size.

Summary [Tentative] Accounting for participants' subjective uncertainty about their future effort decisions modifies the picture painted about their naivety. First, the amount of uncertainty expressed is non-trivial and positively correlated with observed reversals. Second, once accounting for uncertainty, beliefs about the possibility of reversal are well-calibrated. However, the measurement of full belief distributions also reveals that a substantial fraction of those who end up reversing their choice have misspecified beliefs.

4.3 Preferences for commitment and flexibility

We now examine participants' preferences for customizing their future choice set, which we relate to their choice expectations and actual choice behavior: what are the characteristics of those who pay to keep some of their options open? How do they differ from those who pay to restrict their choice set in some way?

4.3.1 Demand for future choice sets

Result 9. *(Overall take-up rates) About 36% choose to restrict their future choice set, while 45% choose to add options. 16% do both.*

Table 3 shows the percentage of participants who paid to remove at least one option from their choice set ($\mathcal{E}^- \subset \mathcal{E}$) and those who paid to add at least one option ($|\mathcal{E}^+| > 1$). Demand for commitment is somewhat less prevalent than demand for flexibility (diff = 0.09, proportions test p-value = 0.344), but the observed prevalence rate is well in line with similar studies of this type e.g., Toussaert (2018). [To add: discussion of whether the observed proportions match the expectations of forecasters.] A non-trivial fraction of participants reveal a preference for both removing and adding options, a pattern that can be rationalized in the presence of uncertainty (see Section 3).[27] Figure B.7 examines the intensive margin of customization decisions. Those who choose to restrict their choice set remove an average of 5.0 options and those who pay to expand it add an average of 2.1 options (diff=2.9 , p=0.000). The correlation between number of options included and excluded is 0.103 (p-value: 0.441).[28]

Figure B.10 shows the distribution of effort levels that were removed vs. added, conditional only on those participants who paid for commitment or flexibility, respectively. Participants were more likely to remove extreme effort levels, with 10 and 100 being excluded respectively by 62% and 76% of participants showing preference for commitment. These findings are compatible with some participants expecting to be tempted by lower effort ($\hat{\gamma} > 1$) and some by larger payments

[27] We have several reasons to believe that this finding is unlikely to be due to noise as in Carrera et al. (2022). First, to ensure that participants do not select commitment by mistake, the interface requires a correct answer to a comprehension question clarifying that customization was optional. Second, virtually no participant chose to both remove and add a given effort option (0% of participants). Finally, very few participants mentioned confusion as a reason for their customization decisions (see Figures B.8 and B.9).

[28] Looking at the robustness of our results to randomizing the order of customization decisions, we find a slight difference for commitment decisions, but not for decisions to acquire flexibility (see Table B.3).

Table 3: Demand for commitment and flexibility

	Did not add	Added	Total
Did not remove	34.5%	29.3%	63.8%
Removed	20.7%	15.5%	36.2%
Total	55.2%	44.8%	100%

Notes: Percentage of participants who added at least one option ($|\mathcal{E}^+| > 1$) and/or removed at least one option ($\mathcal{E}^- \subset \mathcal{E}$). $N = 58$.

($\hat{\gamma} < 1$).[29] On the other hand, participants exhibiting preference for flexibility were more likely to want to keep intermediate effort levels, with 50 and 70 being included 54% and 42% of the time. Among participants who chose not to pay for additional options, 97% included only the effort chosen at t_1 (i.e., $\mathcal{E}^+ = \{e_1\}$), in line with the model outlined in Section 3.

4.3.2 Link between demand for choice sets and beliefs

Result 10. *Preference for flexibility is positively associated with subjective uncertainty, but preference for commitment is not.*

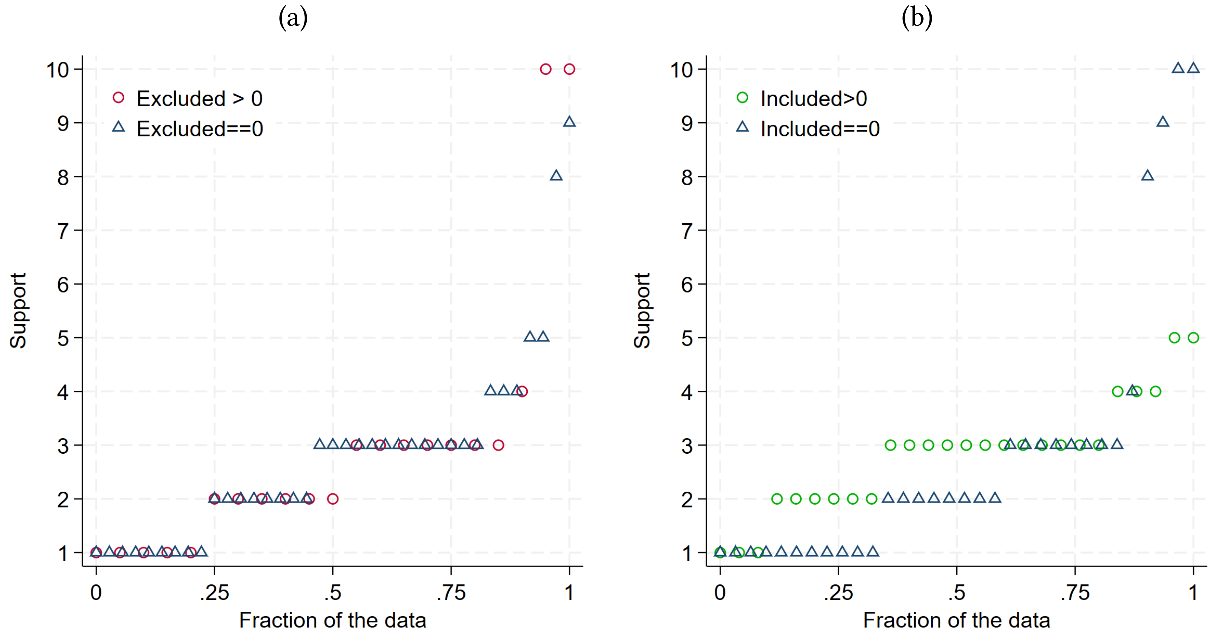
Figure 14 shows how the support size of participants' belief distributions, $|\text{supp}(\hat{f})|$, varies depending on whether they exhibited (i) a preference for commitment ($\mathcal{E}^- \subset \mathcal{E}$ vs. $\mathcal{E}^- = \mathcal{E}$); (ii) a preference for flexibility ($|\mathcal{E}^+| > 1$ vs. $|\mathcal{E}^+| = 1$). According to this uncertainty metric, participants who paid to exclude options do not differ in their subjective uncertainty compared to those who did not. In particular, the proportion expressing degenerate beliefs ($|\text{supp}(\hat{f}) = 1|$) is nearly identical in both cases (24% vs. 24%, p=0.966). On the other hand, preference for flexibility is positively correlated with support size: while 34% express certainty among those who chose not to include additional options, this proportion goes down to 12% for those who do (p = 0.044). The mean number of effort levels included in the support of $\hat{f}$ is 2.81 (2.88) for those with (without) a preference for flexibility (p=0.902). We observe a similar pattern when using the (normalized) entropy of $\hat{f}$ as an alternative measure of subjective uncertainty (see Figure B.11).[30]

Looking at the intensive margin of only those who paid to add options, we observe a positive correlation between the number of options added, $|\mathcal{E}^+|$, and the size of the support of the DM's belief distribution, $|\text{supp}(\hat{f})|$ (rho = 0.363 , p=0.068; see Figure B.12). On the other hand, the correlation between support size and number of options removed is negative (rho=-0.447, p=0.042; see Figure B.13). Figure B.14 shows how the average number of options removed vs. added changes

[29] In total, 14% of participants removed both 10 and 100, which could be rationalized in the framework if beliefs about temptation put positive weight on both $\gamma > 1$ and $\gamma < 1$.

[30] Our conclusions are also similar if we instead look at the proportion of participants for whom $\hat{f}(e_1) < 1$ vs. $\hat{f}(e_1) = 1$ depending on their customization preferences.

Figure 14: Support size by demand for commitment and flexibility



Notes: (a) Quantile plot of support of belief distribution $|\text{supp}(\hat{f})|$ by whether or not the participant exhibited a demand for commitment (excluded > 0 vs. = 0). (b) Quantile plot of support of belief distribution by whether or not the participant exhibited a demand for flexibility (included > 0 vs. = 0).

as a function of perceived chances of reversal, $1 - \hat{f}(e_1)$. We find [no clear/some positive/some negative] association with anticipations of choice reversals (also see Panel B of Table B.4).

These observations bring us to the question of whether the customization decisions of participants reflect a desire to alter their own material outcomes: do participants add or remove options from their choice set because they expect to choose them with positive probability if available? How do their preferences over choice sets relate to their choice expectations? To answer this question, we now conduct aggregate- and individual-level tests of our (perceived) consequentialism property ĉ-FLEX (ĉ-COMMIT), which specifies that if the DM chooses to include (exclude) a given option $e \in \mathcal{E}^+$ ($e \notin \mathcal{E}^-$), then $\hat{f}(e) > 0$.

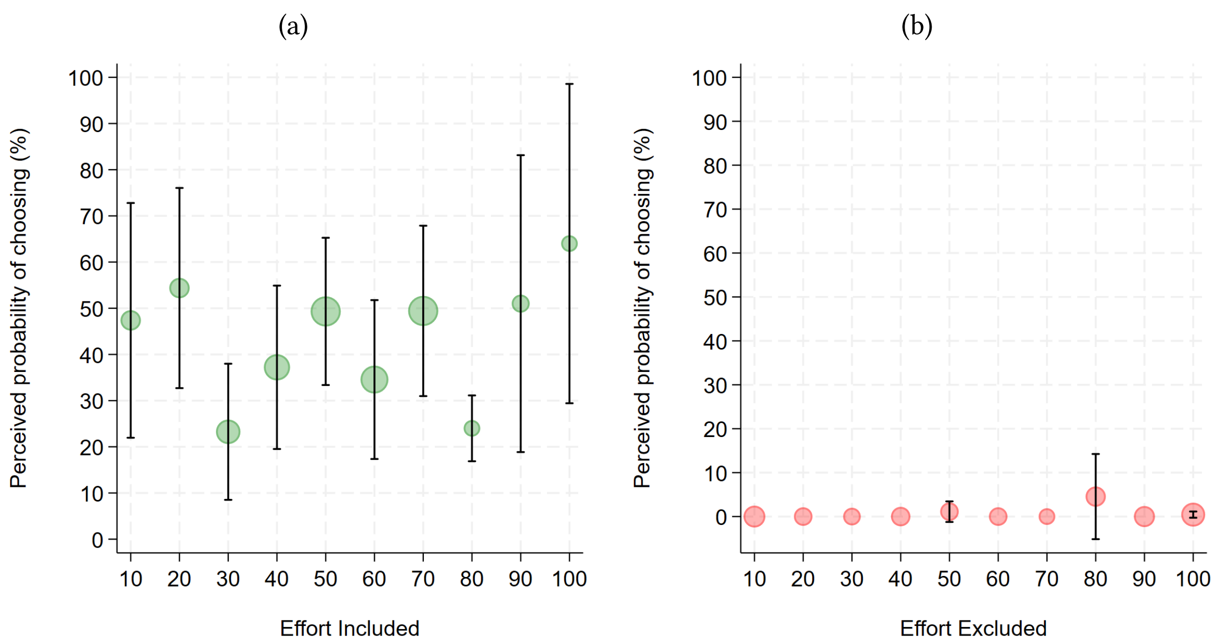
Result 11. *Participants pay to add effort levels that they expect to choose with positive probability; however, they pay to remove options that they do not expect to choose, a finding incompatible with purely consequentialist models of temptation.*

For each of the 10 effort levels, Figure 15a (resp. 15b) shows the average choice probability assigned to this effort level among all participants who included it in (excluded it from) their choice set. For included options, the unweighted average choice expectation across the 10 effort levels is 43.5%, and ranges from 23.3% to 64.0%, well above zero. This is not the case for excluded options, where the analogous unweighted average is 0.6%, and ranges from 0.0% to 4.5%, suggesting a strong violation of perceived consequentialism. Pooling the inclusion (exclusion) decisions of all participants, ĉ-FLEX is violated in 13% of cases where it had bite (14/112 cases), while ĉ-COMMIT is violated in 96% of all cases (100/104 cases); see Figure B.15 for a breakdown by effort level. [To add: discussion of whether forecasters anticipate this.]

Individual-level tests provide a similar picture: individual behavior is fully compatible with ĉ-FLEX for 69% of the participants who expressed some preference for flexibility (N=26); on the other hand, 90% of the participants who expressed some preference for commitment violate ĉ-COMMIT at least once (N=21). Figure B.16 shows how the share of violations varies with the number of included and excluded options: participants are far more likely to violate consequentialism for choice set restrictions than for choice set expansions regardless of the number of options included/excluded. Going one step further, Figure B.17 shows that there is a monotonically increasing relationship between a participant's expectation of choosing a given effort e, $\hat{f}(e)$, and their propensity to include it in $\mathcal{E}^+$. There is no such relationship between $\hat{f}(e)$ and options excluded.

Taken together, these findings show that one needs to step outside of purely consequentialist models of temptation (e.g., models of present bias) to rationalize the observed choice patterns.

Figure 15: Relationship between choice expectation and commitment/flexibility



Notes: (a) Average of $\hat{f}(e)$, for each given e included. (b) Average of $\hat{f}(e)$, for each given e excluded. OLS-estimated linear model with cluster-robust 95% confidence intervals shown. Bubble size is proportional to sample size.

Table 4: Reversals and demand for commitment and flexibility

Actual reversals	**Did not add**	**Added**	**Total**
Did not remove	25%	53%	38%
Removed	42%	56%	48%
Total	31%	54%	41%

Notes: Percentage of participants who exhibited choice reversals, conditional on preferences for commitment and flexibility. $N = 58$.

4.3.3 Link between demand for choice sets and choice reversals

We conclude this results section by investigating how preferences over choice sets relate to actual choices. After all, individuals who anticipated not choosing a given option if available could nevertheless succumb to it when the time comes, consistent with naivety.

Result 12. *Demand for commitment and demand for flexibility [both predict] choice reversals.*

Table 4 shows how the proportion of actual choice reversals changes depending on the customization decisions of participants.[31] Time consistent participants were generally less inclined to pay for customization and preferences over choice sets are highly predictive of reversals. While our estimate of the unconditional probability of reversal was 0.41, the estimated posterior conditional on demanding flexibility, $\mathrm{P}(e_2 \neq e_1 \mid |\mathcal{E}^+| > 1)$, is equal to 0.54, and the corresponding posterior for commitment demand, $\mathrm{P}(e_2 \neq e_1 \mid \mathcal{E}^- \subset \mathcal{E})$, is 0.48. As shown in Panel C of Table B.4, demand for flexibility [is/is not] significantly associated with actual reversals, while demand for commitment [is/is not]. Looking at the intensive margin of customization decisions, the average number of added and excluded options also varies by reversal type: participants with a stronger [demand for commitment and/or flexibility] are more likely to reverse their choice (see Figure B.18).[32] In general, preference for flexibility appears to be a stronger and more robust predictor of anticipated and actual choice reversals than preference for commitment, both on the extensive and intensive margins (see all panels in Table B.4).

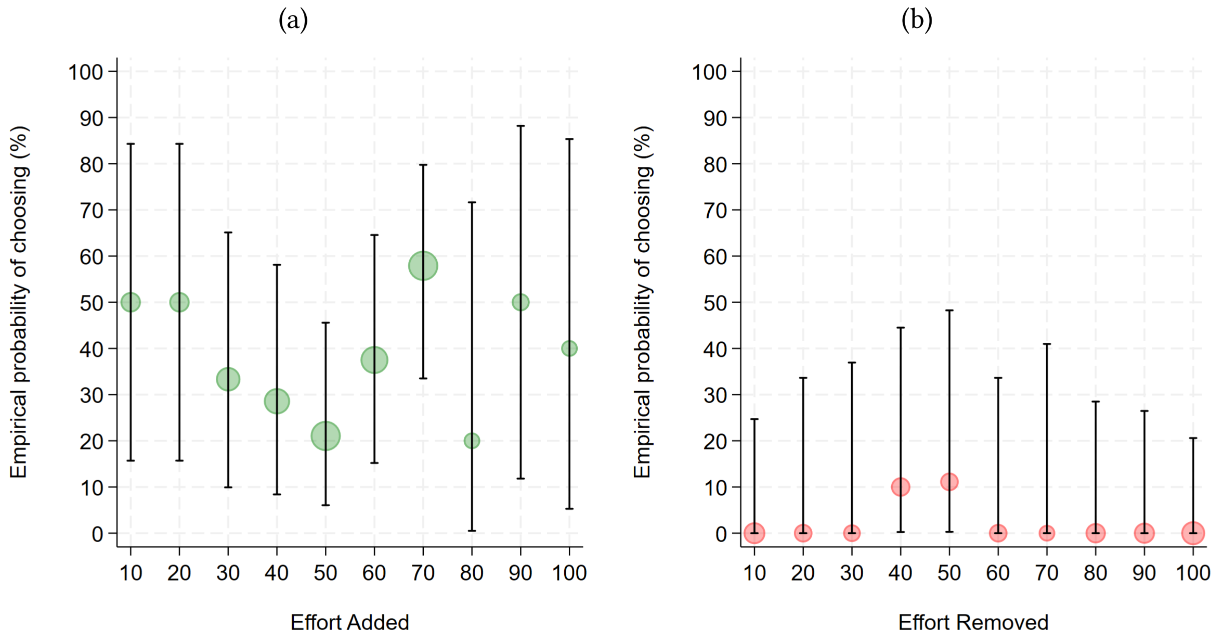
As discussed in Section 3, demand for commitment could be correlated with (anticipations of) choice reversals driven by uncertainty over the realized state, i.e., the expectation of overwhelming temptation is not necessary to generate this correlation. If that is the main mechanism, then

[31] Table B.5 shows the analogous table for predicted reversals, using participants' point predictions as well as their belief distributions, respectively.

[32] [To be updated based on main data: Interestingly, Figure B.19 shows that the specific effort levels added or removed differ by reversal type: while the distribution of added and removed effort levels is balanced for time consistent participants, those who end up reversing their choice upwards are more likely to remove higher effort levels and more likely to add lower effort levels. The reverse is not true for those who exhibit a downward reversal; in this case, the added options are more likely to be intermediate and the removed options more likely to be the extreme ones.]

removed choices should not be chosen. We thus examine whether consequentialism holds for actual choices by testing c-FLEX and c-COMMIT at the aggregate level. Figure 16 shows that c-FLEX is broadly satisfied, but c-COMMIT is violated.

Figure 16: Effort decisions and demand for choice sets



Notes: Binomial exact 95% confidence intervals shown. Bubble size is proportional to sample size. (a) Empirically observed frequency of chosen effort level at t_2, conditional on including option in menu. (b) Empirically observed frequency of chosen effort level at t_2, conditional on removing option in menu.

5 Discussion

In this paper, we revisit a classical experimental framework for studying the consistency of effort decisions in intertemporal choice situations. Relative to traditional designs, we observe not only effort choices and point predictions about future behavior, we also measure uncertainty about these future choices by eliciting a full probability distribution over outcomes from each participant. Doing so allows us to study the extent to which individuals' beliefs might be misspecified and/or miscalibrated. We pair this data with information on preferences for commitment and flexibility, which we measure by allowing participants to customize their future choice sets in any way they want. Combining all these pieces of data, we clarify the theoretical and empirical relationship between demand for commitment and flexibility, predictions of future choices, and actual choices when the time comes.

[Conclusions to be updated based on main data] In line with past research, we find a downward bias on average in the effort chosen when decisions have immediate consequences, and point predictions are almost fully in line with initial decisions, suggesting substantial naivety. The picture however changes once subjective uncertainty is accounted for. First, about [three quarters] of participants express some uncertainty about their future choices and the expression of uncertainty is correlated with actual choice reversals, suggesting that it cannot be ignored if one wants to make statements about naivety. Second, after accounting for uncertainty, predictions of choice reversals are well calibrated relative to actual choice reversals. Nevertheless, a substantial fraction of participants appear to have misspecified beliefs. Third, we find that demand for flexibility is strongly positively associated with subjective uncertainty and actual choice reversals, while the relationship is [inexistent/weaker] for preference for commitment. Furthermore, while flexibility considerations seem to be closely tied to choice expectations, this is not the case for commitment: options removed from the choice set are virtually always assigned zero probability of being chosen. This strong violation of consequentialism is in contradiction with predictions of models of sophisticated present bias and favors models where choice set restrictions are driven by non-consequentialist motives such as self-control costs.

Below we leverage additional information to discuss alternative interpretations of our data and articulate directions for future research.

True naivety or simply measurement error? We find that a substantial fraction of those who reverse their choice assigned zero chance to choosing the effort level they eventually selected at t_2 (i.e., $\hat{f}(e_2) = 0$). Our favored interpretation is that those participants have a misspecified model of the world, i.e., upon carefully considering their options, they failed to foresee the possibility of some contingencies. However, another interpretation is that participants have a perfectly correct model of the world but their beliefs were inaccurately measured, thus showing apparent misspecification and/or miscalibration. Since expressing beliefs by providing a full distribution might be a difficult exercise, we paid careful attention to the design of this task. We asked participants to proceed in steps, by first selecting the support of their belief distribution (Step 1) and then assigning the weights (Step 2). We also allowed them to readjust the assigned probabilities after selecting them on the chart, which should have minimized measurement error. As a result, we find limited evidence of inconsistencies between these two steps and with point predictions, with only 5% failing to include their point belief in the support.[33] To further assess whether measurement error might be an issue, we asked participants who did not assign 100% probability to their chosen effort e_2 to explain why. Among those who assigned zero probability, [X%] cited confusion about the belief elicitation task, with the main reasons being [X, Y].

Nature of misperceptions If a non-trivial fraction of participants indeed have misspecified and/or miscalibrated beliefs about their future effort e_2, one question is what exactly they are wrong about. Within the theoretical framework outlined in Section 3, we highlight that the DM might misperceive the distribution of shocks influencing their effort cost e.g., by assigning zero weight to the state s at which e_2 is the maximizer of $U_s + V_s$. Alternatively, the DM could misperceive their temptation parameter ($\hat{\gamma} \neq \gamma$), by failing to acknowledge the distortionary role of concerns for immediate gratification. Without imposing additional restrictions on the structure of these beliefs, it is difficult to draw inferences about the nature of the misspecification from the comparison of predictions with actual choices. Future work could seek to develop new identification strategies to enable this separation e.g., via an information provision experiment (Haaland, Roth and Wohlfart, 2023). Short of that, we examine self-reported assessments at the end of the second survey. We find that participants explain their (unanticipated) choice reversals by a mixture of [distribution of motives], suggesting that [X,Y] might be primary sources of misperception.

Non-neutrality of belief elicitation As data architects, we constructed our dataset by connecting behavior with beliefs directly elicited from respondents (instead of inferring them indirectly from behavior). One concern about this exercise, besides its difficulty and the associated risks of measurement error, is that the act of eliciting someone's beliefs about their future behavior could alter what they do when the time comes e.g., if the expectations formed by the

[33] [To add: We will investigate whether our main findings on the relationship between belief uncertainty, choice reversals, and demand for choice set restrictions vs. expansions continue to hold if (i) we use Step 1 of the belief elicitation procedure to calculate the support $|\text{supp}(\hat{f})|$, instead of Step 2; (ii) we remove all subjects who were inconsistent between Step 1 and Step 2; (iii) we remove all subjects who reported being confused about the prediction task at the end of Survey 2. To further assess the perceived difficulty of the belief elicitation task, we will report descriptively ratings of the belief distribution interface and belief difficulty ratings.]

individual create an internal reference point, possibly serving as a commitment device. Although the available evidence suggests that these concerns are likely minor (Augenblick and Rabin, 2019; Le Yaouanq and Schwardmann, 2022; Fedyk, 2022), we test whether beliefs influence actual choice behavior by comparing the findings of our main treatment with data obtained from a treatment that was identical in all respects except for the absence of any belief elicitation. If belief elicitation created a desire for consistency, one should observe a higher proportion of choice reversals when beliefs are not elicited. As indicated in Appendix Table [X], we find [no/limited/some] evidence in this direction, with a [similar] proportion of choice reversals in both treatments [proportion test results]. We also find [no/minor/significant] differences in the distribution of effort at t_2 across conditions. [To add: Appendix figure presenting quantile plots of effort by condition and result from Kolmogorov-Smirnov tests.] Finally, since beliefs were elicited before customization decisions at t_1, we test whether being asked to pro-actively consider the possibility of making various effort decisions affects demand for commitment or flexibility. We find [no/some] differences in the extensive or intensive margin of the demand for choice set restrictions or expansions.[To add: Appendix table reporting the fraction of participants who chose to remove/add at least one option and the average number of options added or removed.] Future work should examine the circumstances under which formulating expectations may affect future behavior and help individuals achieve higher self-discipline.[34]

Interpretation of commitment and flexibility decisions Choice set expansions generally satisfy our consequentialism property, but not choice set restrictions: while participants who add an option to their choice set expect to choose it with positive probability, excluded options are rarely expected to be chosen if available. To rationalize the latter, other reasons beyond altering one's material consequences must be accounted for. The framework proposed in this paper captures such non-material concerns by allowing the DM's utility to be negatively affected by the presence of unchosen alternatives as in Gul and Pesendorfer (2001). While we propose self-control costs as a possible rationalization for the observed failures of consequentialism, other complementary explanations may include decision costs or attentional costs (Maćkowiak, Matějka and Wiederholt, 2023), fear of choice overload (Dean, Ravindran and Stoye, 2024), guilt (Kopylov, 2012), regret (Sarver, 2008), image concerns (Dillenberger and Sadowski, 2012), to name a few. Figures B.8 and B.9 present a breakdown of the interpretations provided by respondents for their customization decisions. We find that demand for customized choice sets reflects a range of mo-

[34] One conjecture is that the passage of time may weaken the strength of expectations e.g., if individuals simply fail to recall later on what they anticipated they would do. [To examine this interpretation, we investigate whether participants have a tendency to recall making point predictions that are closer to the actual decisions they made than to the predictions they truly made at t_1.] For recent work on the interplay between memory and self-control, see Sial, Sydnor and Taubinsky (2024).

tives, with the avoidance of temptation being one of them. Future research could explore new ways of separating these various motives and understanding their behavioral implications.

Link between commitment demand, sophistication, and reversals In this paper, we find [no/positive/negative] correlation between demand for choice set restrictions and choice revisions. Our findings [contradict/align with] Sadoff, Samek and Sprenger (2020) who find a negative correlation. They interpret this negative correlation as being due to an interplay between present bias and naivety: if those most inclined to reverse their choices are the least likely to anticipate them, they will also be the least likely to seek strategies to alter their future choices. In Section 4.3.2, we document that anticipations of choice reversals ($1-\hat{f}(e_1)$) do not seem to predict whether participants choose any restrictions. To investigate this interpretation further, we also examine the relationship between demand for commitment ($\mathcal{E}^- \subset \mathcal{E}$) and belief misspecification ($\hat{f}(e_2) = 0$), or more generally anticipations of one's actual choice $\hat{f}(e_2)$. As shown in Figure B.20, we find [limited/no/some] support for this conjecture: [if anything, those who chose to remove at least one option were less likely to anticipate their eventual choice]. Further, 13.5% of individuals who did not exclude options were misspecified, as compared to 28.6% of individuals who did exclude options (proportions test p-value = 0.160).[35]

[35]Comparing the probability assigned to choosing actual effort e_2, $\hat{f}(e_2)$, this was 0.51 for individuals who did not exclude any option from the choice set, and 0.45 for those who did (t-test p-value = 0.547).

References

Ahn, David S., and Todd Sarver. 2013. "Preference for flexibility and random choice." *Econometrica*, 81(1): 341–361.

Augenblick, Ned, and Matthew Rabin. 2019. "An Experiment on Time Preference and Misprediction in Unpleasant Tasks." *The Review of Economic Studies*, 86(3): 941–975.

Augenblick, Ned, Muriel Niederle, and Charles Sprenger. 2015. "Working over Time: Dynamic Inconsistency in Real Effort Tasks *." *The Quarterly Journal of Economics*, 130(3): 1067–1115.

Bartling, Björn, Ernst Fehr, and Holger Herz. 2014. "The Intrinsic Value of Decision Rights." *Econometrica*, 82(6): 2005–2039.

Carrera, Mariana, Heather Royer, Mark Stehr, Justin Sydnor, and Dmitry Taubinsky. 2022. "Who Chooses Commitment? Evidence and Welfare Implications." *The Review of Economic Studies*, 89(3): 1205–1244.

Danz, David, Lise Vesterlund, and Alistair J. Wilson. 2022. "Belief Elicitation and Behavioral Incentive Compatibility." *American Economic Review*, 112(9): 2851–2883.

Dean, Mark, and John McNeill. 2020. "Preference for Flexibility and Random Choice: an Experimental Analysis." Working Paper.

Dean, Mark, Dilip Ravindran, and Jörg Stoye. 2024. "A Better Test of Choice Overload." Working Paper.

Dekel, Eddie, Barton L. Lipman, Aldo Rustichini, and Todd Sarver. 2007. "Representing Preferences with a Unique Subjective State Space: A Corrigendum." *Econometrica*, 75(2): 591–600.

Dekel, Eddie, Barton L. Lipman, and Aldo Rustichini. 2001. "Representing Preferences with a Unique Subjective State Space." *Econometrica*, 69(4): 891–934.

Dekel, Eddie, Barton L. Lipman, and Aldo Rustichini. 2009. "Temptation-Driven Preferences." *Review of Economic Studies*, 76(3): 937–971.

DellaVigna, Stefano, and Devin Pope. 2018. "Predicting Experimental Results: Who Knows What?" *Journal of Political Economy*, 126(6): 2410–2456.

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

Dillenberger, David, and Philipp Sadowski. 2012. “Ashamed to be Selfish.” *Theoretical Economics*, 7(1): 99–124.

Fedyk, Anastassia. 2022. “Asymmetric Naiveté: Beliefs about Self-Control.” *Management Science, forthcoming.*

Ferreira, João V., Nobuyuki Hanaki, and Benoît Tarroux. 2020. “On the roots of the intrinsic value of decision rights: Experimental evidence.” *Games and Economic Behavior*, 119: 110–122.

Gill, David, and Victoria Prowse. 2019. “Measuring costly effort using the slider task.” *Journal of Behavioral and Experimental Finance*, 21: 1–9.

Gul, Faruk, and Wolfgang Pesendorfer. 2001. “Temptation and Self-Control.” *Econometrica*, 69(6): 1403–1435.

Haaland, Ingar, Christopher Roth, and Johannes Wohlfart. 2023. “Designing Information Provision Experiments.” *Journal of Economic Literature*, 61(1): 3–40.

Heidhues, Paul, and Botond Kőszegi. 2009. “Futile Attempts at Self-Control.” *Journal of the European Economic Association*, 7(2-3): 423–434.

Hossain, Tanjim, and Ryo Okui. 2013. “The Binarized Scoring Rule.” *The Review of Economic Studies*, 80(3): 984–1001.

Iacovone, Leonardo, David McKenzie, and Rachael Meager. 2023. “Bayesian Impact Evaluation with Informative Priors: An Application to a Colombian Management and Export Improvement Program.” The World Bank 10274.

Kopylov, Igor. 2012. “Perfectionism and Choice.” *Econometrica*, 80(5): 1819–1843.

Laibson, David. 1997. “Golden Eggs and Hyperbolic Discounting.” *Quarterly Journal of Economics*, 112(2): 443–478.

Le Yaouanq, Yves, and Peter Schwardmann. 2022. “Learning About One’s Self.” *Journal of the European Economic Association*, 20(5): 1791–1828.

Maćkowiak, Bartosz, Filip Matějka, and Mirko Wiederholt. 2023. “Rational Inattention: A Review.” *Journal of Economic Literature*, 61(1): 226–273.

O'Donoghue, Ted, and Matthew Rabin. 1999. "Doing It Now or Later." *American Economic Review*, 89(1): 103–124.

Otis, Nicholas. 2022. "Policy Choice and the Wisdom of Crowds."

Sadoff, Sally, Anya Samek, and Charles Sprenger. 2020. "Dynamic Inconsistency in Food Choice: Experimental Evidence from Two Food Deserts." *The Review of Economic Studies*, 87(4): 1954–1988.

Sarver, Todd. 2008. "Anticipating Regret: Why Fewer Options May Be Better." *Econometrica*, 76(2): 263–305.

Sen, Amartya K. 1988. "Freedom of choice: Concept and content." *European Economic Review*, 32: 269–94.

Sial, Afras, Justin Sydnor, and Dmitry Taubinsky. 2024. "Biased Memory and Perceptions of Self-Control." Working paper.

Spiegler, Ran. 2011. "Dynamically Inconsistent Preferences III: Partial Naivete." In *Bounded Rationality and Industrial Organization.* Oxford University Press.

Toussaert, Séverine. 2018. "Eliciting Temptation and Self-Control Through Menu Choices: A Lab Experiment." *Econometrica*, 86(3): 859–889.

Appendix

A Attrition

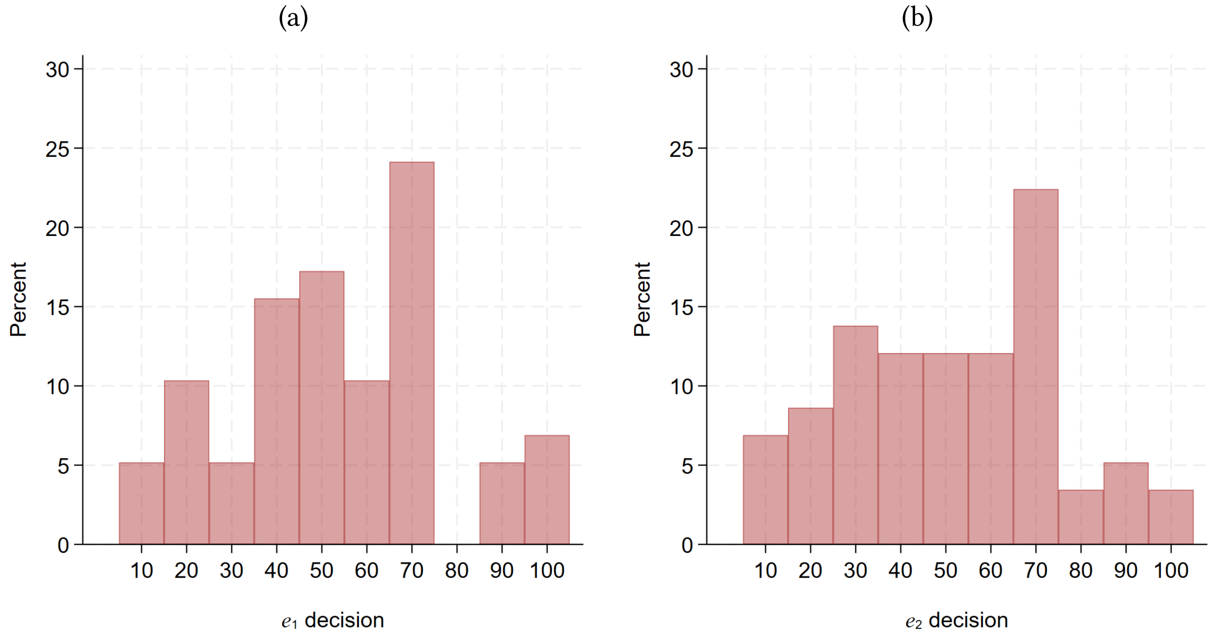
Table A.1: Attrition - Effort and Beliefs

	Completed only S1 Mean	**Completed both** Mean	Diff	P-value
Effort				
e_1	38.0	53.4	-15.4	0.110
Predictors				
$\hat{e}_2$	38.0	53.5	-15.5	0.108
$E_{\hat{f}_i}(e)$	37.7	52.2	-14.5	0.117
$\text{mode}_{\hat{f}_i}(e)$	38.0	52.6	-14.6	0.128
$\hat{F}_i^{-1}(0.5)$	38.0	52.8	-14.8	0.122
Uncertainty				
$\min(\text{supp}(\hat{f}_i))$	32.0	43.0	-11.0	0.136
$\max(\text{supp}(\hat{f}_i))$	42.0	63.2	-21.2	0.040
range(supp($\hat{f}_i$))	10.0	20.3	-20.5	0.085
$\sigma_{\hat{f}_i}$	2.29	6.18	-3.89	0.002
entropy($\hat{f}_i$)	0.09	0.30	-0.21	0.000
$\lvert\text{supp}(\hat{f}_i)\rvert$	1.80	2.99	-1.19	0.009
FP-skew($\hat{f}_i$)	-0.25	0.24	-0.49	0.401
Commitment/Flexibility				
Added any	0.20	0.45	-0.25	0.110
Number added (if any)	3.0	2.5	0.5	0.690
Removed any	0.30	0.32	-0.02	0.886
Number removed (if any)	3.7	4.7	-1.0	0.688
N	10	71		

Notes: Table shows average values of variables derived from effort and beliefs, for those who completed Survey 1 (but did not return for Survey 2), versus those completed both surveys. P-values are derived from unpaired, unequal variance, t-tests. Predictors panel shows summary statistics for point predictions and predictors based on belief distributions. Uncertainty panel provides additional information about participants' belief distributions, $\hat{f}_i$. entropy($\hat{f}_i$) refers to normalized entropy $-\frac{1}{\ln 10}\sum_{j=1}^{10}\hat{f}(e_j)\ln\hat{f}(e_j)$. FP-Skew refers to the Fisher-Pearson skewness coefficient.

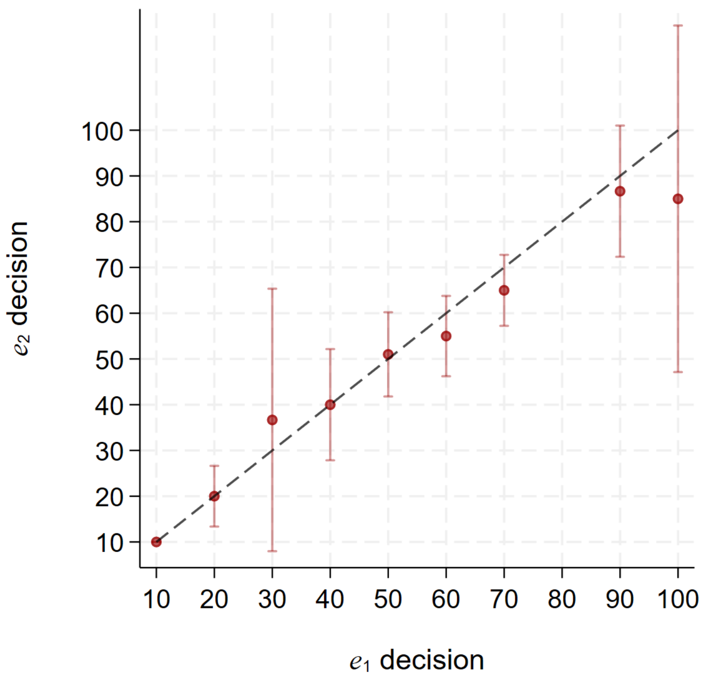
B Additional Figures and Tables

Figure B.1: Histogram of effort decisions



Notes: (a) Histogram of e_1. (b) Histogram of e_2.

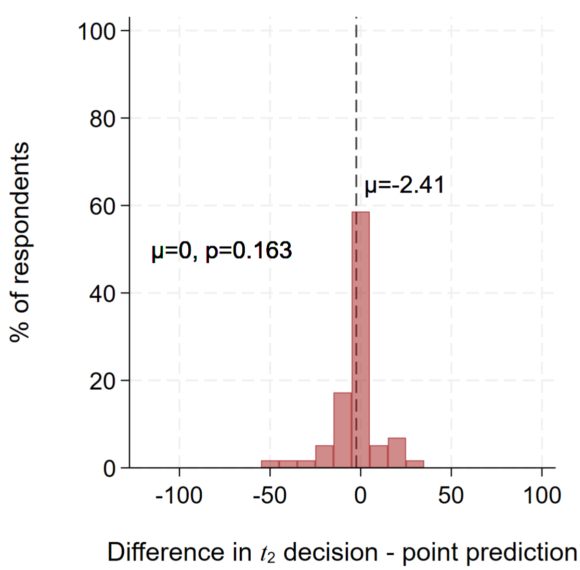
Figure B.2: Choice reversals by t_1 decision



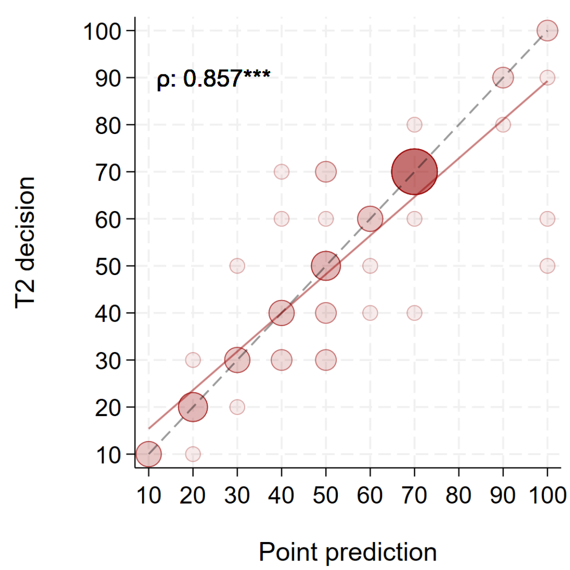
Notes: Mean of actual effort at time t_2, e_2, for each possible effort level at time t_1, e_1. Shown with 95% confidence intervals. $N = 58$.

Figure B.3: Deviations of decision e_2 from point prediction $\hat{e}_2$

(a)



(b)



Notes: (a) Histogram of the distribution of differences $e_2 - \hat{e}_2$ with a vertical bar for the mean and p-value of paired t-test $H_0 : e_2 - \hat{e}_2 = 0$. (b) Scatter plot of e_2 against $\hat{e}_2$ with Pearson correlation coefficient. Smallest bubble refers to N=1 respondent(s). Largest bubble refers to N=10 respondent(s).

Table B.1: Correlation of belief uncertainty measures

	$\lvert\text{supp}(\hat{f}_i)\rvert$	entropy$(\hat{f}_i)$	range(supp$(\hat{f}_i)$)	$\sigma_{\hat{f}_i}$
$\lvert\text{supp}(\hat{f}_i)\rvert$	1.000			
entropy$(\hat{f}_i)$	0.903***	1.000		
range(supp$(\hat{f}_i)$)	0.990***	0.894***	1.000	
$\sigma_{\hat{f}_i}$	0.933***	0.949***	0.951***	1.000

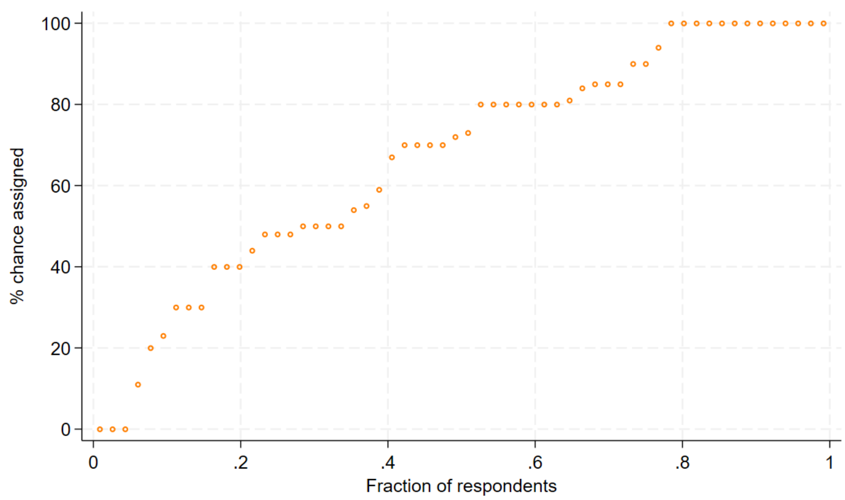
Notes: This table shows the correlation coefficients between belief uncertainty measures. Significance levels are indicated by * (10%), ** (5%), and *** (1%). $N = 58$.

Table B.2: Summary statistics - Effort and Beliefs

	Mean	SD	Min	Q25	Q50	Q75	Max	Diff e_2	P-value e_2
e_1	53.6	23.9	10.0	40.0	50.0	70.0	100.0	-2.4	0.163
e_2	51.2	23.8	10.0	30.0	50.0	70.0	100.0	-	-
Predictors									
$\hat{e}_2$	53.6	24.8	10.0	40.0	50.0	70.0	100.0	-2.4	0.163
$E_{\hat{f}_i}(e)$	52.7	23.7	10.6	40.0	52.8	70.0	100.0	-1.5	0.379
$\text{mode}_{\hat{f}_i}(e)$	53.4	25.9	10.0	30.0	50.0	70.0	100.0	-2.2	0.227
$\hat{f}_i^{-1}(0.5)$	53.4	24.7	10.0	40.0	50.0	70.0	100.0	-2.2	0.216
Uncertainty									
min(supp$(\hat{f}_i)$)	44.3	25.1	10.0	20.0	45.0	60.0	100.0	-	-
max(supp$(\hat{f}_i)$)	63.3	25.0	20.0	50.0	70.0	80.0	100.0	-	-
range(supp$(\hat{f}_i)$)	19.0	20.6	0.0	10.0	20.0	20.0	90.0	-	-
$\sigma_{\hat{f}_i}$	5.95	5.65	0.00	2.37	5.00	7.75	26.93	-	-
entropy$(\hat{f}_i)$	0.29	0.24	0.00	0.10	0.28	0.43	0.98	-	-
$\lvert\text{supp}(\hat{f}_i)\rvert$	2.84	2.05	1.00	2.00	3.00	3.00	10.00	-	-
FP-skew$(\hat{f}_i)$	0.32	1.16	-2.03	-0.16	0.00	0.99	3.71	-	-

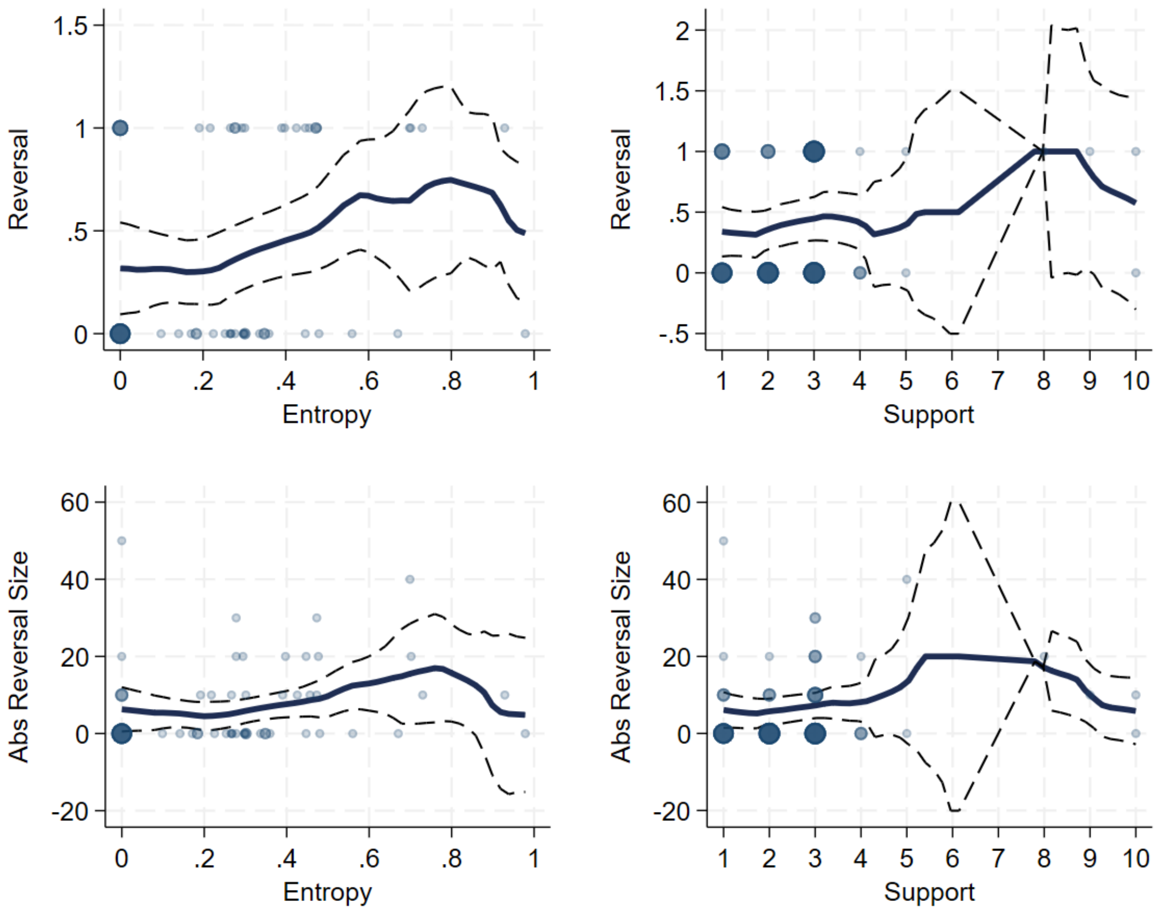
Notes: Predictors panel shows summary statistics for point predictions and predictors based on belief distributions, compared with actual effort at t_2, e_2. P-values are derived from Wilcoxon matched-pairs signed-rank tests (row variable vs. e_2). Uncertainty panel provides additional information about participants' belief distributions, $\hat{f}_i$. entropy$(\hat{f}_i)$ refers to normalized entropy $-\frac{1}{\ln 10}\sum_{j=1}^{10}\hat{f}(e_j)\ln\hat{f}(e_j)$. FP-Skew refers to the Fisher-Pearson skewness coefficient. $N = 58$.

Figure B.4: Distribution of the belief probability mass assigned to point prediction $\hat{e}_2$



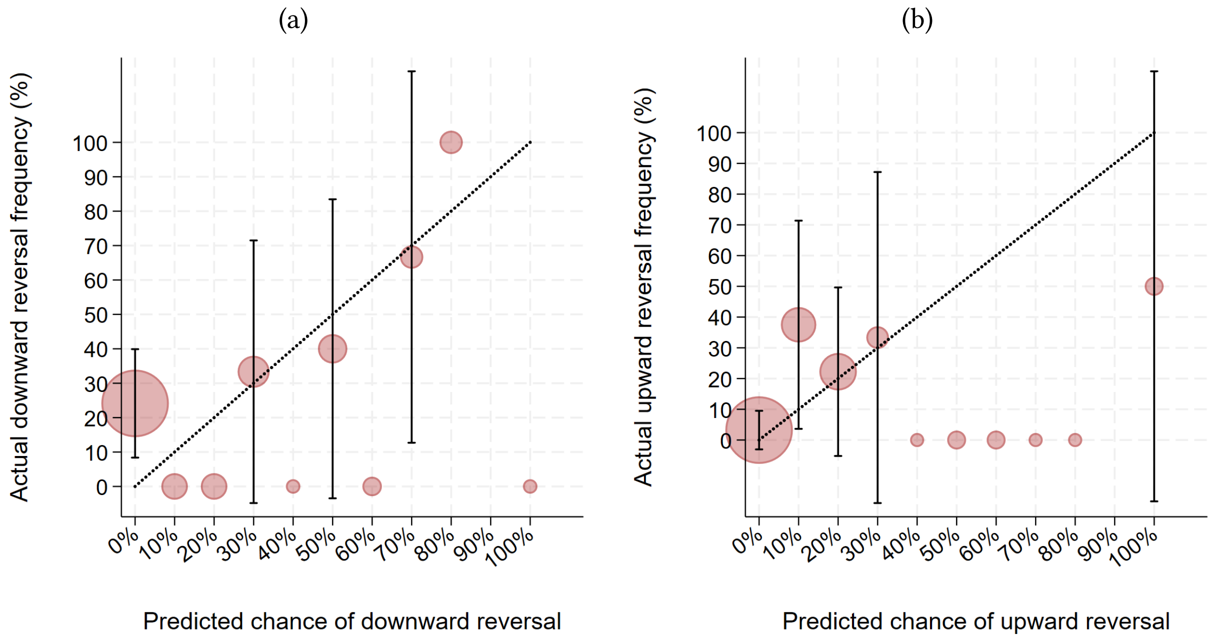
Notes: Quantile plot of the belief distribution probability mass $\hat{f}(\hat{e}_2)$ assigned by each respondent to their point prediction $\hat{e}_2$. $N = 58$.

Figure B.5: Uncertainty and Reversals I



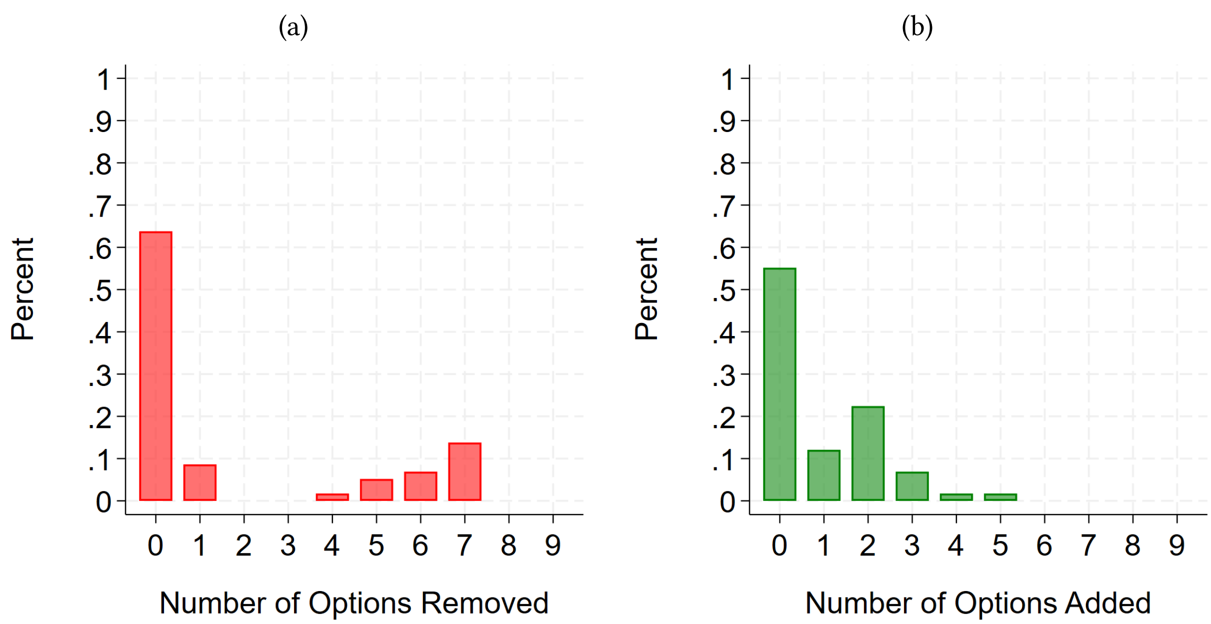
Notes: Likelihood of reversal and absolute size of reversal by belief uncertainty measure: normalized entropy $-\frac{1}{\ln 10}\sum_{j=1}^{10}\hat{f}(e_j)\ln\hat{f}(e_j)$ for left column and support of belief distribution $|\mathrm{supp}(\hat{f})|$ for right column. Each panel displays a local polynomial regression of either (i) a binary reversal indicator (top row) or (ii) the absolute size of the reversal in effort (bottom row) on the respective uncertainty measure. Dashed lines show 95% confidence intervals based on standard errors from local polynomial smoothing. $N = 58$.

Figure B.6: Uncertainty and reversals II



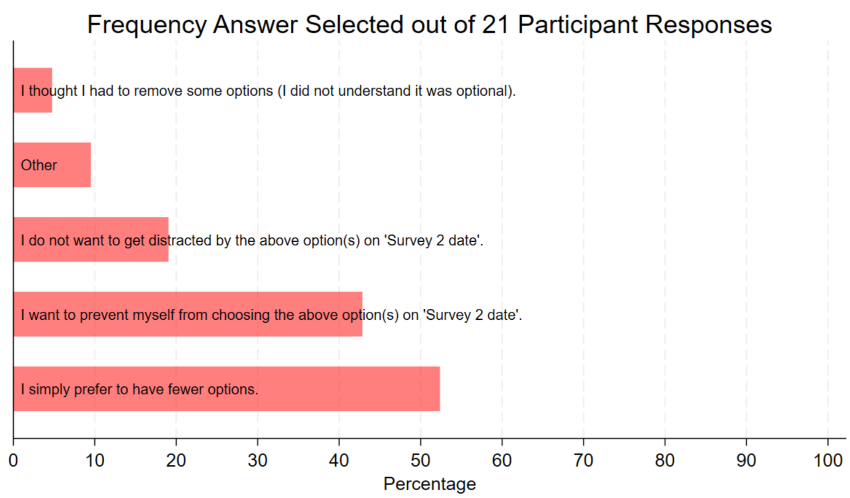
Notes: Empirically observed frequency of respective reversal type, conditional on the probability assigned to that reversal. GLM-estimated binomial model with cluster-robust 95% confidence intervals shown. Each predicted chance is ±5 percentage points (e.g. 20% refers to $[15, 25]$%), excluding the end points of 0% and 100% which are exact; hence predictions in the intervals $(0, 5)$ and $(95, 100)$ are dropped. Applies to (a) 0% and (b) 0% of participants, respectively. Bubble size is proportional to sample size.

Figure B.7: Number of options removed and added



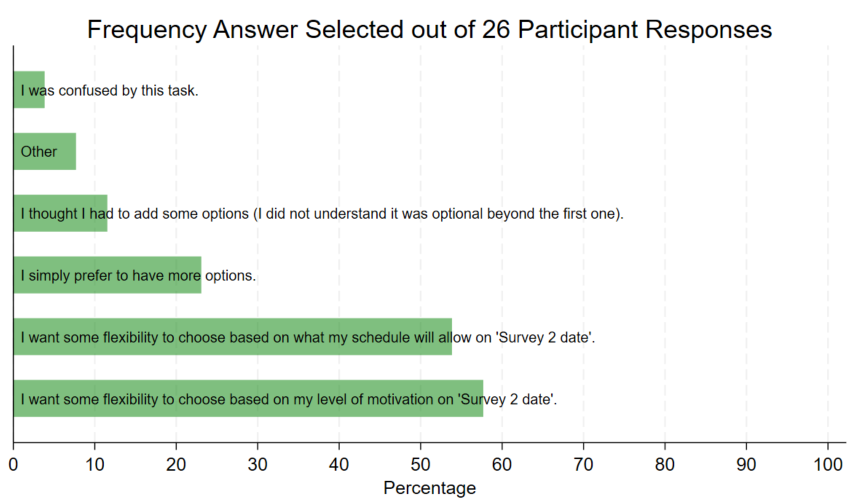
Notes: (a) Histogram of the number of options removed. (b) Histogram of the number of options added.

Figure B.8: Motives for excluding alternatives



Notes: Motives provided to exclude options, among those who paid to remove at least one.

Figure B.9: Motives for including alternatives



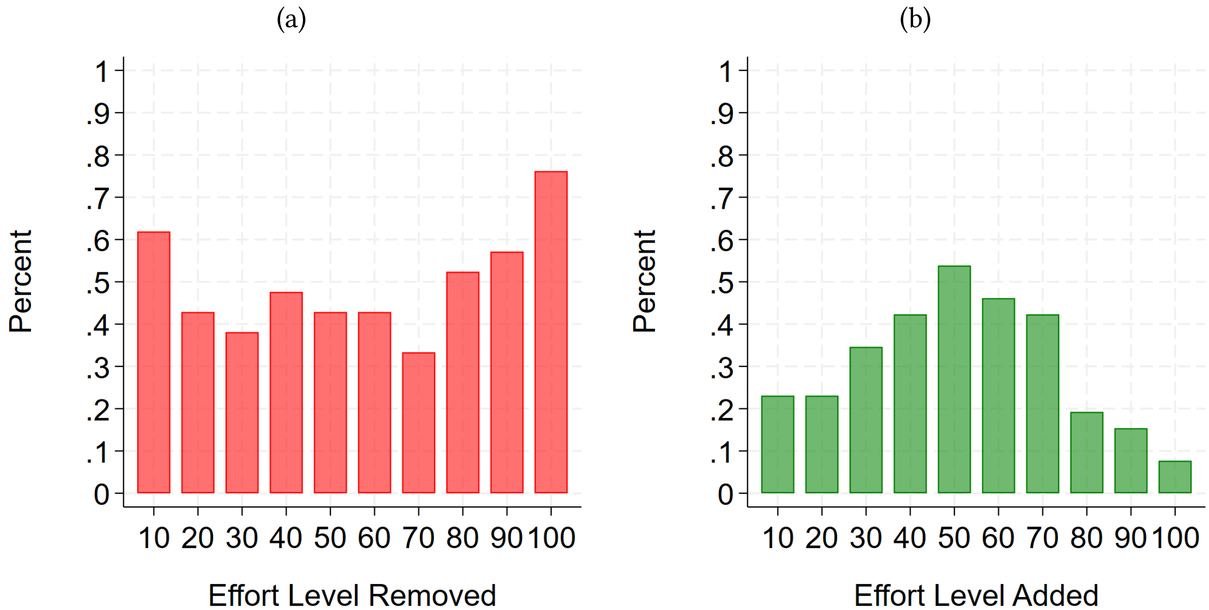
Notes: Motives provided to include options, among those who paid to add at least one.

Table B.3: Examining Order Effects - Flexibility and Commitment menus

Measure	**Commitment first**	**Commitment Second**	**Difference**	**p-value**
Paid to add	0.44	0.45	-0.01	0.914
Paid to remove	0.44	0.30	0.14	0.291

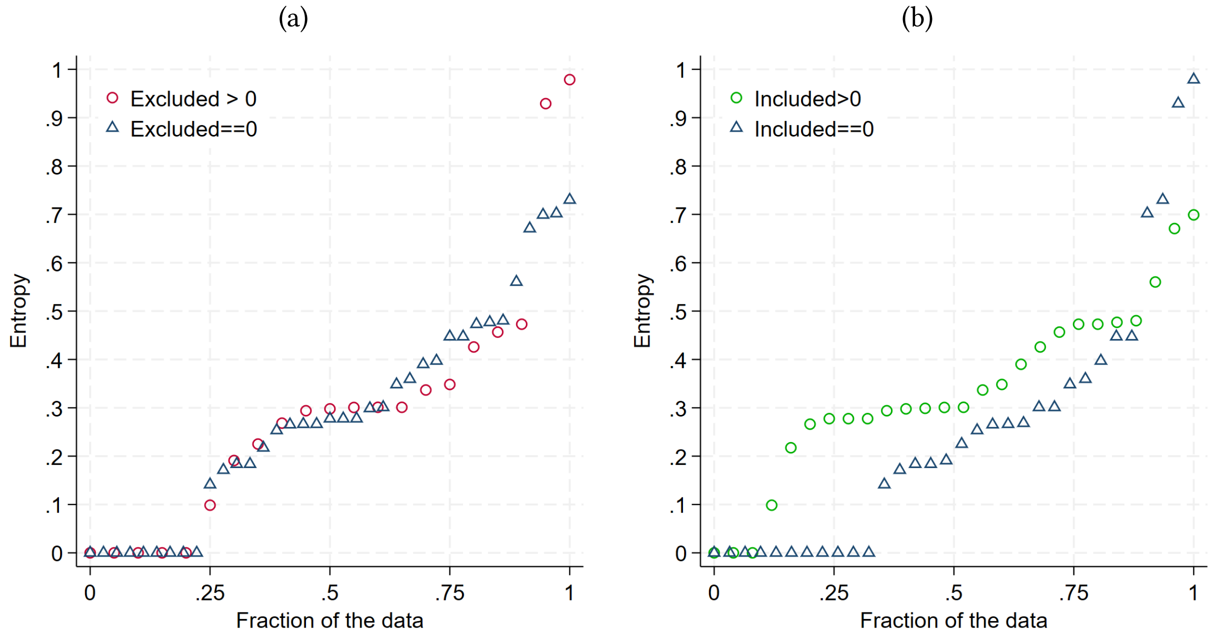
Notes: Proportion of individuals who exhibited preferences for flexibility (paid to add) or commitment (paid to remove), by the (random) order in which the two menu customization options appeared (commitment first or second). The "Difference" column shows the difference between commitment first and commitment second. $N = 58$.

Figure B.10: Proportion who chose to remove vs add a given effort level



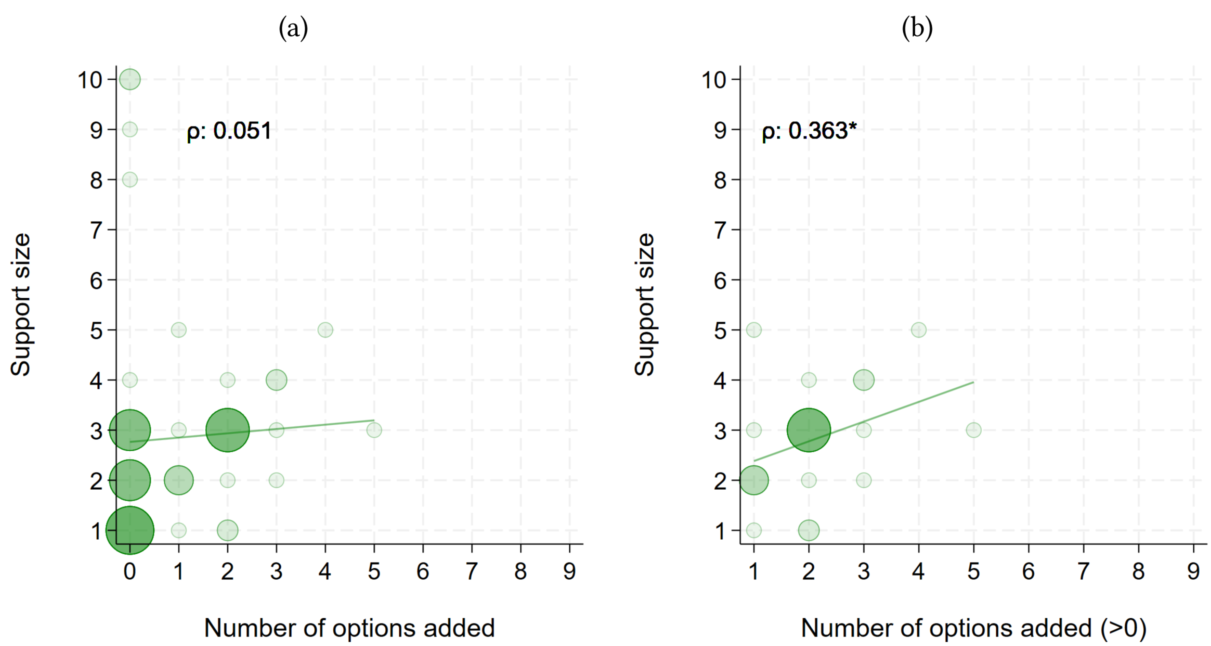
Notes: (a) Bar graph of the proportion of participants who removed a given effort level, conditional on paying to remove at least one option ($N = 21$). (b) Bar graph of the proportion of participants who added a given effort level, conditional on paying to add at least one additional option ($N = 26$).

Figure B.11: Entropy by demand for commitment and flexibility



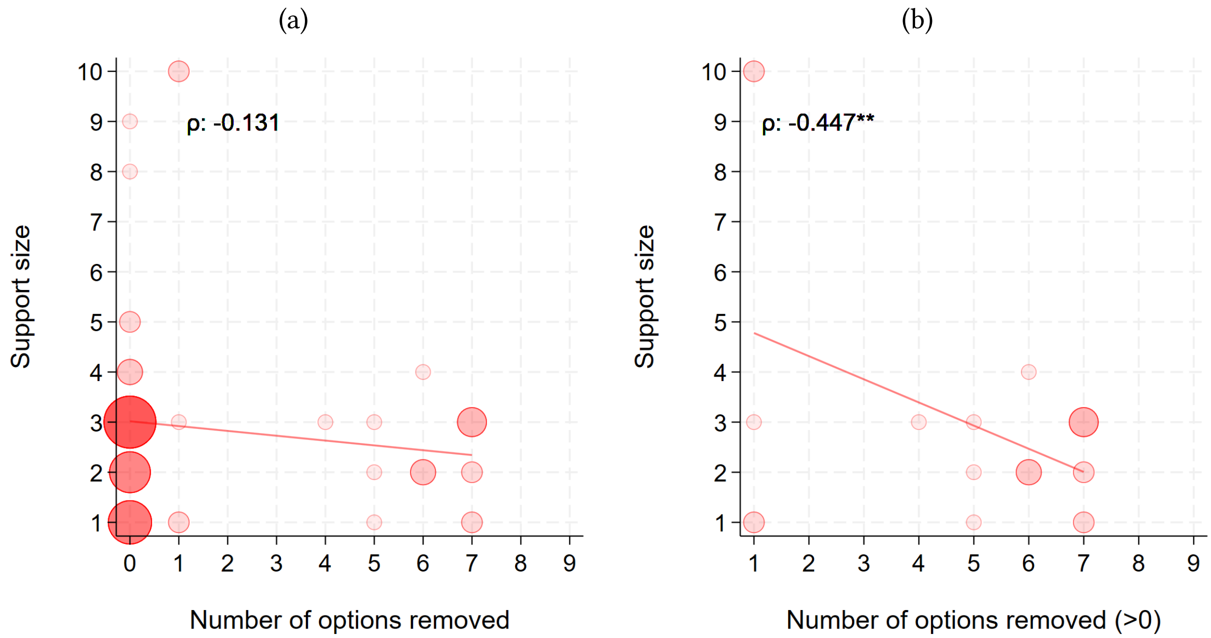
Notes: (a) Quantile plot of the normalized entropy of a participant's belief distribution by whether or not the participant exhibited a demand for commitment (excluded > 0 vs. = 0). (b) Quantile plot of belief entropy by whether or not the participant exhibited a demand for flexibility (included > 0 vs. = 0).

Figure B.12: Support size and demand for flexibility



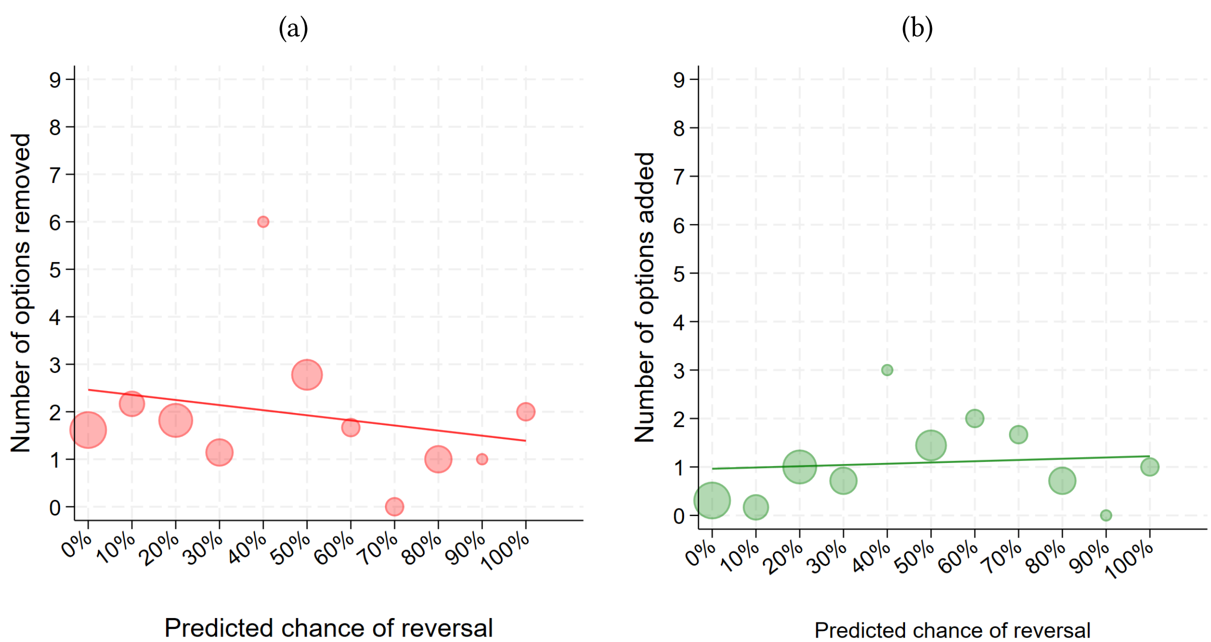
Notes: (a) Scatter plot of support of belief distribution by number of options added (all participants). (b) Scatter plot of support of belief distribution by number of options added (sample restricted to those who added > 0). Bubble size is proportional to sample size.

Figure B.13: Support size and demand for commitment



Notes: (a) Scatter plot of support of belief distribution by number of options removed (all participants). (b) Scatter plot of support of belief distribution by number of options removed (sample restricted to those who removed > 0). Bubble size is proportional to sample size.

Figure B.14: Predicted reversal and flexibility/commitment



Notes: (a) Average number of options removed, conditional on the probability assigned to any reversal $1 - \hat{f}(e_1)$. (b) Average number of options included, conditional on the probability assigned to any reversal $1 - \hat{f}(e_1)$. Each predicted chance is ±5 percentage points (e.g. 20% refers to $[15, 25]$%), excluding the end points of 0% and 100% which are exact; hence predictions in the intervals $(0, 5)$ and $(95, 100)$ are dropped (applies to 0% of participants). Bubble size is proportional to sample size.

Table B.4: Reversals and preference for commitment/flexibility

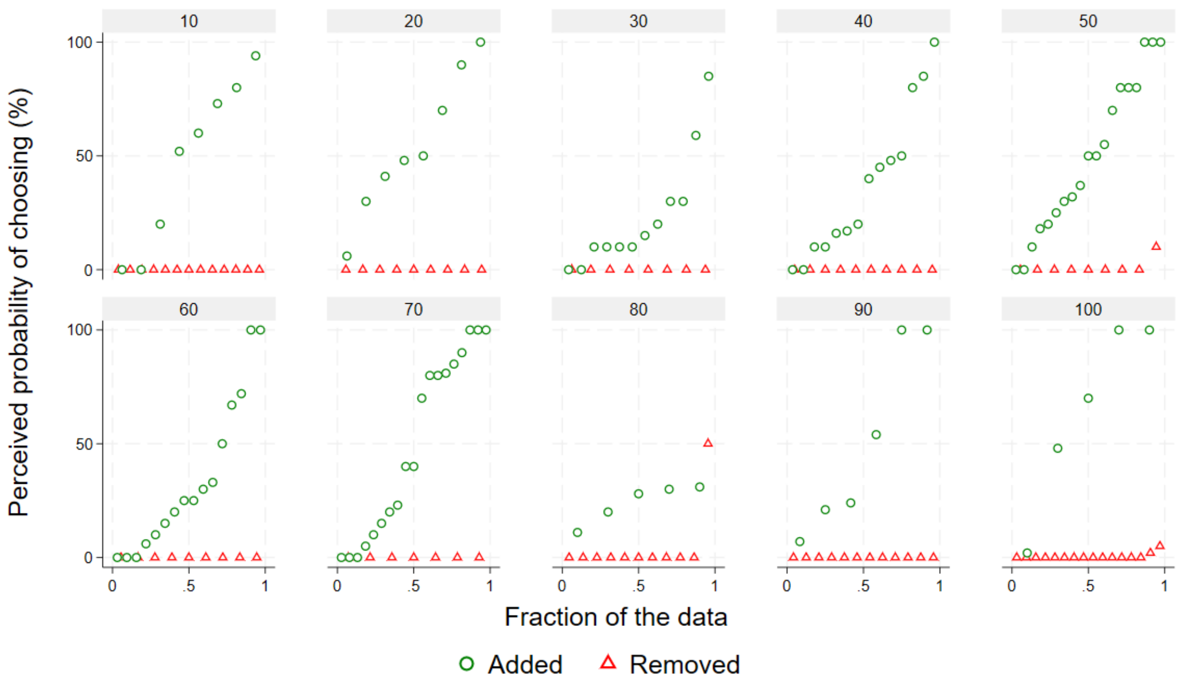
Panel A: Predicted Reversal (point prediction)						
	(1)	(2)	(3)	(4)	(5)	(6)
Added Any	0.029		0.027			
	(0.093)		(0.092)			
Removed Any		-0.067	-0.066			
		(0.090)	(0.090)			
Number Added				0.006		0.011
				(0.041)		(0.041)
Number Removed					-0.019	-0.019
					(0.012)	(0.013)
Constant	0.125^{**}	0.162^{**}	0.150^{**}	0.132^{**}	0.172^{***}	0.162^{**}
	(0.059)	(0.062)	(0.064)	(0.058)	(0.059)	(0.064)

Panel B: Predicted Reversal (belief distribution)						
	(1)	(2)	(3)	(4)	(5)	(6)
Added Any	0.156^{*}		0.156^{*}			
	(0.079)		(0.079)			
Removed Any		0.014	0.019			
		(0.086)	(0.086)			
Number Added				0.060^{**}		0.061^{**}
				(0.027)		(0.028)
Number Removed					-0.002	-0.005
					(0.014)	(0.014)
Constant	0.285^{***}	0.350^{***}	0.278^{***}	0.299^{***}	0.359^{***}	0.307^{***}
	(0.056)	(0.051)	(0.056)	(0.052)	(0.049)	(0.055)

Panel C: Actual Reversal						
	(1)	(2)	(3)	(4)	(5)	(6)
Added Any	0.226^{*}		0.229^{*}			
	(0.130)		(0.130)			
Removed Any		0.098	0.105			
		(0.137)	(0.136)			
Number Added				0.124^{**}		0.124^{**}
				(0.048)		(0.048)
Number Removed					0.009	0.003
					(0.024)	(0.022)
Constant	0.312^{***}	0.378^{***}	0.273^{***}	0.298^{***}	0.398^{***}	0.293^{***}
	(0.083)	(0.081)	(0.093)	(0.077)	(0.077)	(0.084)

Notes: Robust standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. All specifications have $N = 58$ observations.

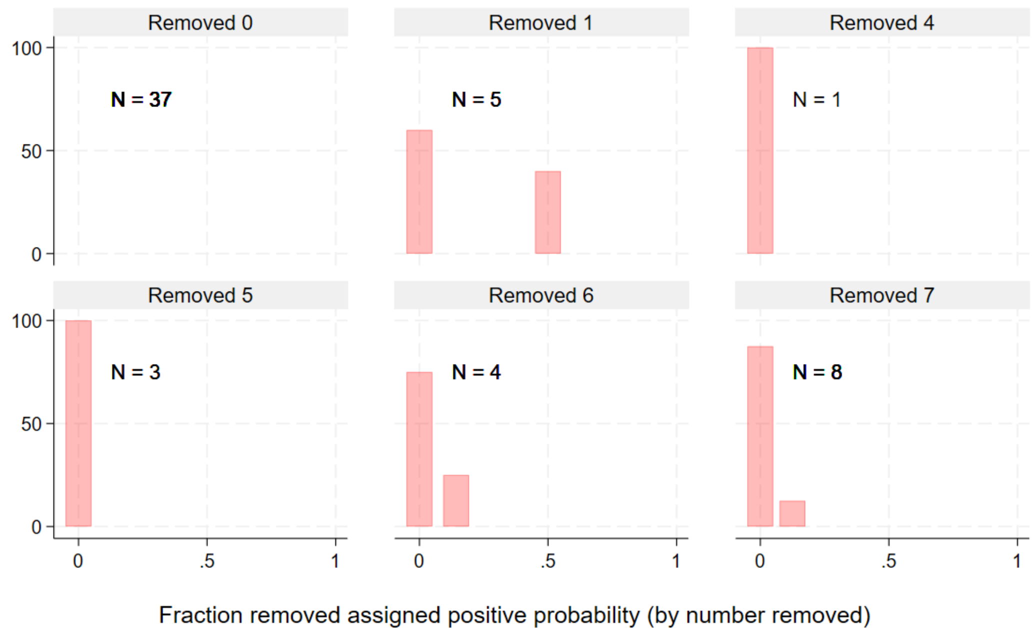
Figure B.15: Relationship between choice expectation and commitment/flexibility



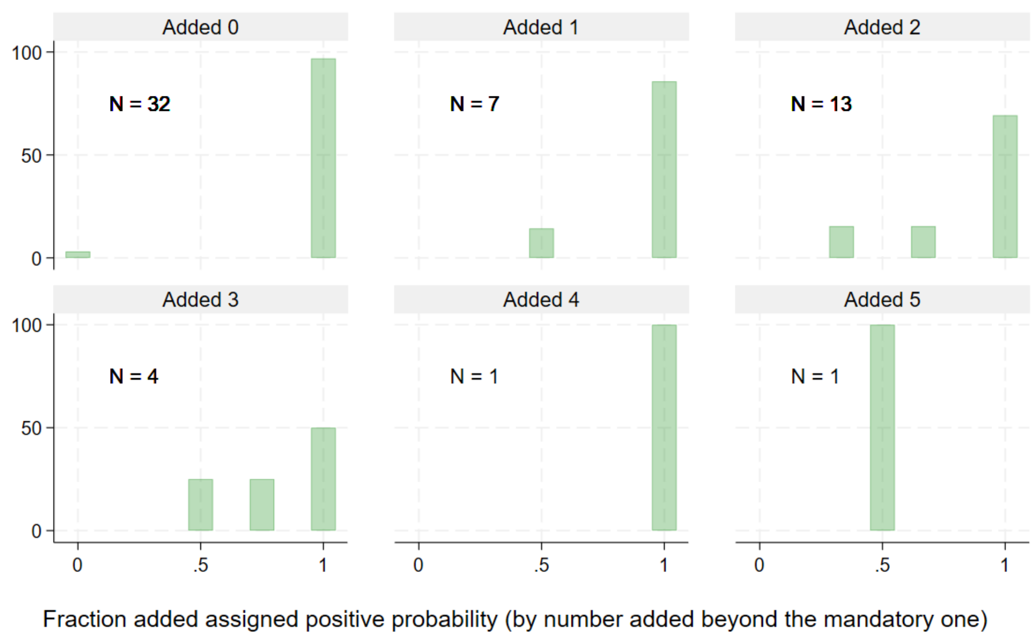
Notes: Quantile plot of percent chance, $\hat{f}(e)$, assigned to effort level $e \in \{10, 20, ..., 100\}$, if e was added vs. removed.

Figure B.16: Violations of ĉ-COMMIT (top panel) and ĉ-FLEX (bottom panel)

(a)

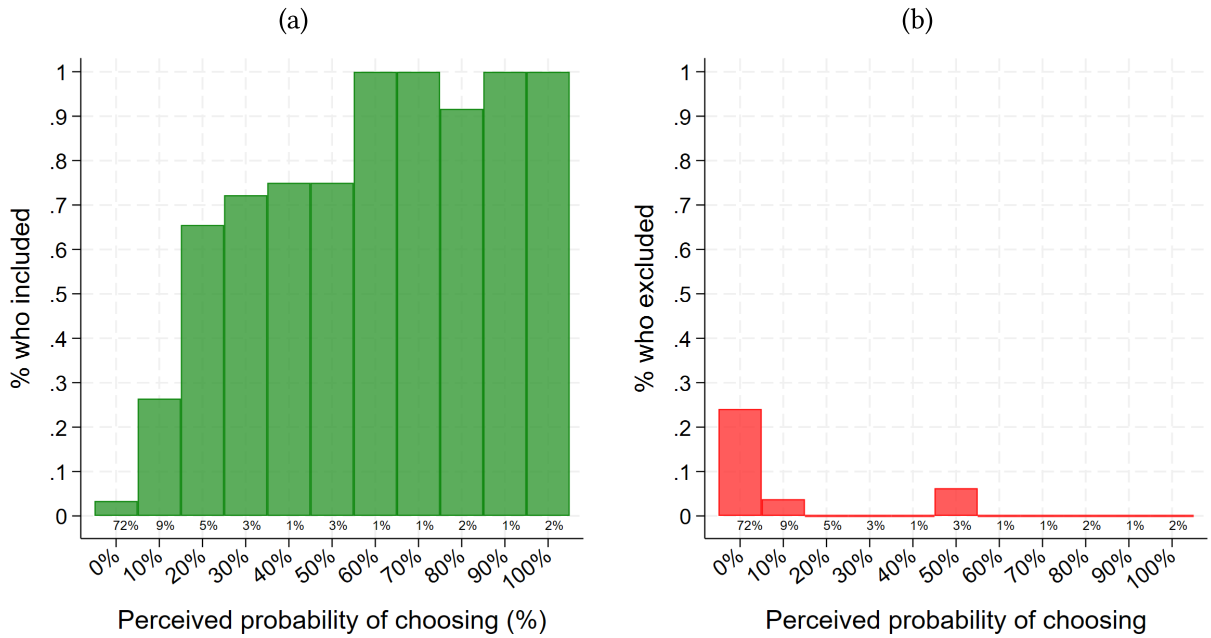


(b)



Notes: (a) Among all effort levels removed, $e \notin \mathcal{E}^-$, figure shows the fraction assigned positive probability, $\hat{f}(e) > 0$ (by number of effort levels removed). (b) Among all effort levels added, $e \in \mathcal{E}^+$, figure shows the fraction assigned positive probability, $\hat{f}(e) > 0$ (by number of effort levels added beyond the mandatory one). Fractions calculated at the participant level.

Figure B.17: Relationship between choice expectation and choice inclusion/exclusion



Notes: Percentage who included (a) or removed (b) effort levels, conditional on assigning different probabilities of choosing them. Each predicted chance is ±5 percentage points (e.g. 20% refers to $[15, 25]$%), excluding the end points of 0% and 100% which are exact; hence predictions in the intervals $(0, 5)$ and $(95, 100)$ are dropped (applies to 1% of all predictions). Percentages below the bars show the fraction of observations (belief about a given effort level) for each perceived probability of choosing e.g., 72% of all effort levels were assigned 0% probability of being chosen.

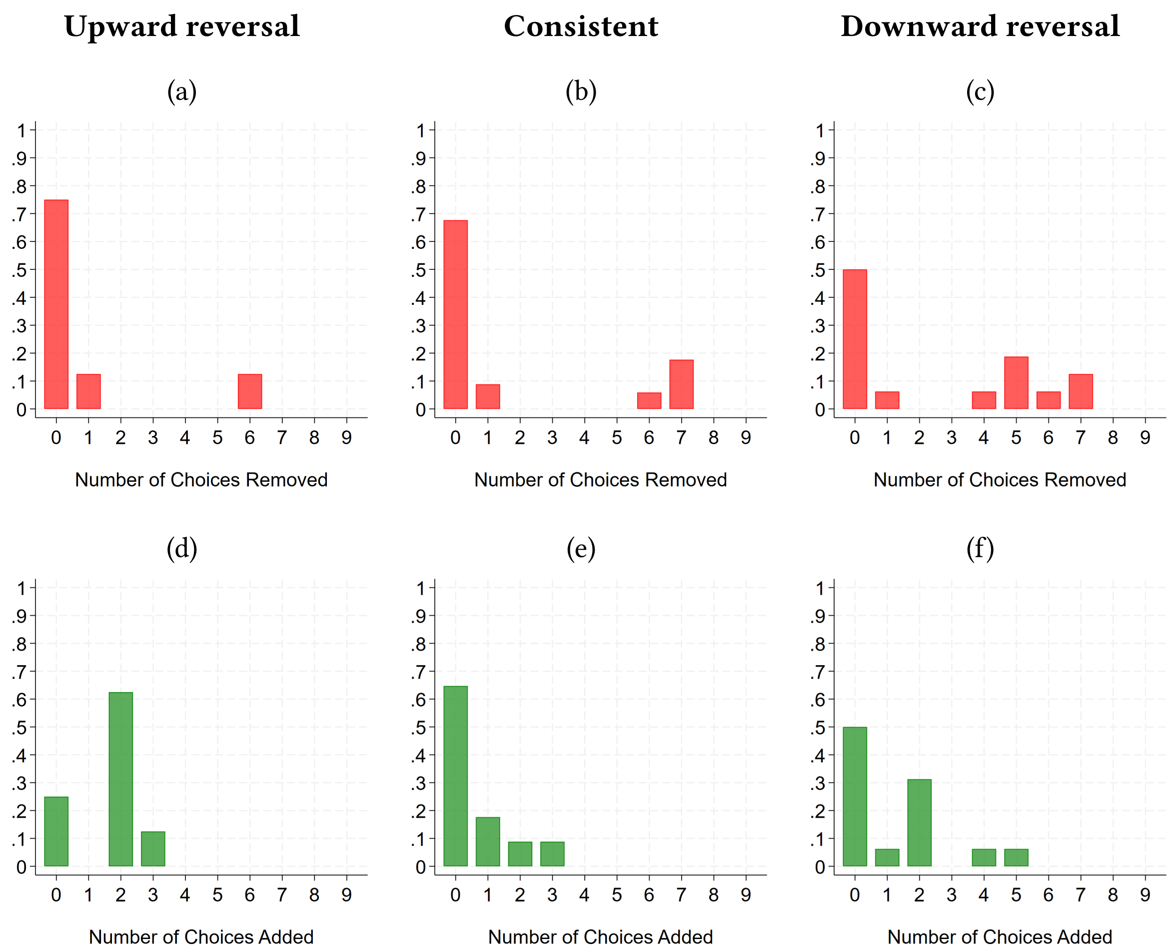
Table B.5: Predicted reversals and demand for commitment and flexibility

Panel A: Predicted Reversal (point prediction)	**Did not add**	**Added**	**Total**
Did not remove	10%	24%	16%
Removed	17%	0%	10%
Total	13%	15%	14%

Panel B: Predicted reversals (belief distribution)	**Did not add**	**Added**	**Total**
Did not remove	30%	47%	38%
Removed	25%	44%	33%
Total	28%	46%	36%

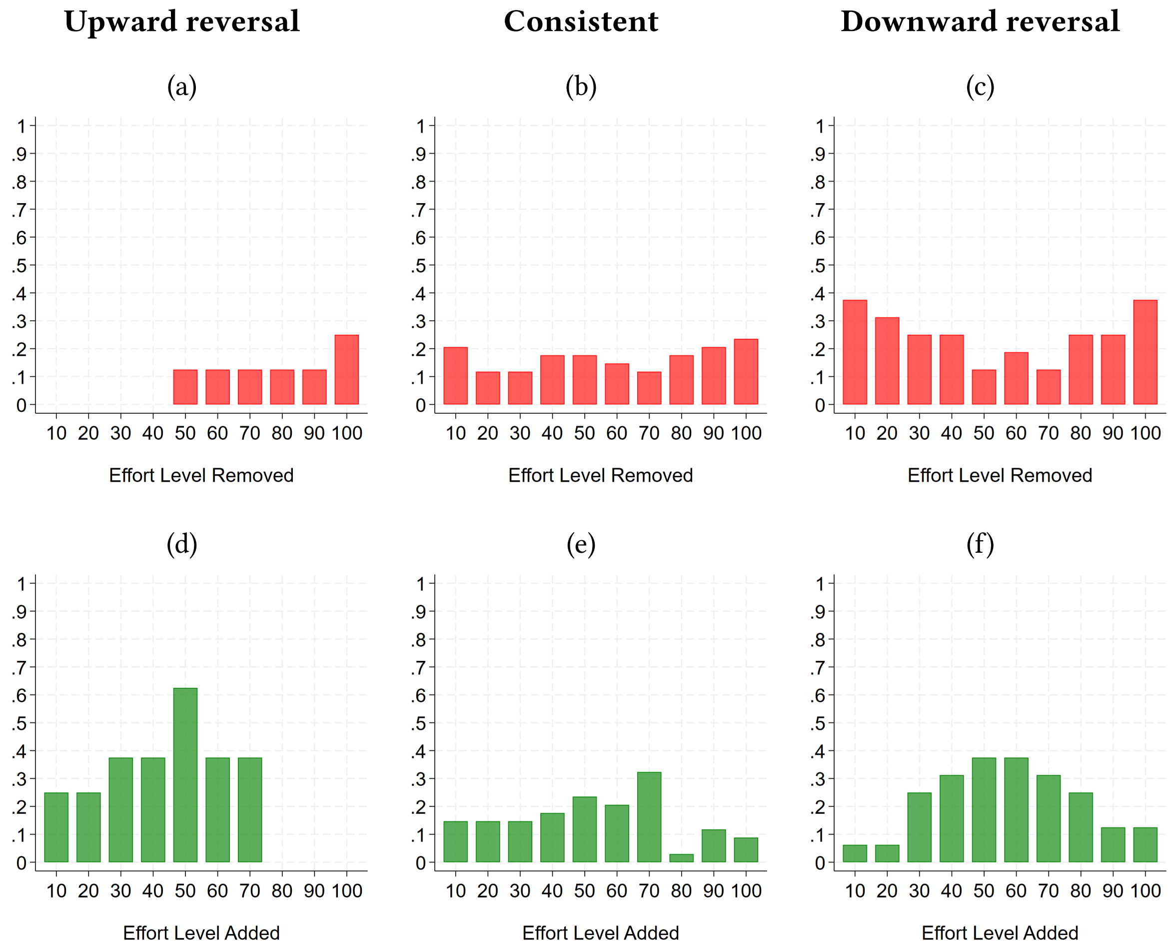
Notes: First panel shows predicted reversals based on participants' point predictions. Second panel of table shows percentage of participants who predicted reversals, where predictions are based on one draw from each participant's belief distribution. $N = 58$.

Figure B.18: Extent of demand for commitment and flexibility by reversal



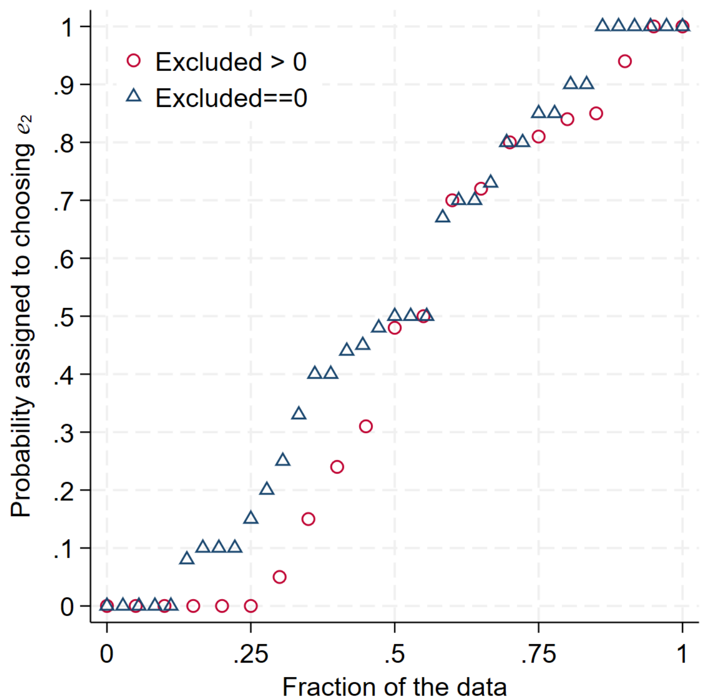
Notes: Panels (a)-(c) refer to the number of effort levels removed by reversal type (upward reversal, consistent, downward reversal). Panels (d)-(f) refer to the number of effort levels added by reversal type (upward reversal, consistent, downward reversal).

Figure B.19: Commitment and flexibility choices by reversal



Notes: Panels (a)-(c) refer to the distribution of specific effort levels removed by reversal type (upward reversal, consistent, downward reversal). Panels (d)-(f) refer to the distribution of specific effort levels added by reversal type (upward reversal, consistent, downward reversal).

Figure B.20: Anticipations of actual choice vs. demand for commitment



Notes: Quantile plot of the probability assigned to choosing actual e_2, $\hat{f}(e_2)$, by whether or not the participant exhibited a demand for commitment (excluded > 0 vs. = 0).