Discussion Paper Series – CRC TR 224

Discussion Paper No. 763
Project A 01

Sticky Models

Paul Grass[1]
Philipp Schirmer[2]
Malin Siemers[3]

June 2026

[1]University of Bonn, Email: paul.grass@uni-bonn.de.
[2]University of Bonn, Email: philipp.schirmer@uni-bonn.de..
[3]University of Bonn,Email: malin.siemers@uni-bonn.de.

Support by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) through CRC TR 224 is gratefully acknowledged.

Sticky Models

Paul Grass, Philipp Schirmer, Malin Siemers

June 19, 2026

Abstract: People often form incomplete mental models, having to revise them as new relevant variables become observable. We show experimentally that models are 'sticky': revised models remain strongly influenced by earlier models formed using a subset of variables. Sticky models occur across three different data-generating processes and across heterogeneous reasoning types of subjects. Guided by a simple framework of dynamic model formation, we investigate cognitive effort allocation as a key mechanism: across three DGPs, we find that stickiness is driven by subjects who exert relatively less cognitive effort during the model revision relative to the initial model formation.

JEL Codes: D83, D91

Keywords: mental models, learning dynamics, attention, mental representation, bounded rationality

---

Contact: Paul Grass, University of Bonn, paul.grass@uni-bonn.de. Philipp Schirmer, University of Bonn, philipp.schirmer@uni-bonn.de. Malin Siemers, University of Bonn, malin.siemers@uni-bonn.de. The order of authors is alphabetical. **Acknowledgments**: We thank Botond Kőszegi, Chris Roth, and Florian Zimmermann for their invaluable guidance and support. We thank Chiara Aina, Peter Andre, Kai Barron, Benjamin Enke, Nicola Gennaioli, Duarte Gonçalves, Thomas Graeber, Luca Henkel, Ulrike Malmendier, Robin Musolff, Ryan Oprea, Joshua Schwartzstein, Ran Spiegler, Mark Toth, Emanuel Vespa, Sevgi Yuksel, seminar and conference participants at the Berlin Micro Theory and Behavioral Economics PhD Conference, ESA European Meeting, ESA North American Meeting, Maastricht University, M-BEES, the MPI Workshop on Misperceptions, Mental Models, and Polarization, SITE Experimental Economics, the University of Bonn, the Workshop on Beliefs, Narratives and Memory, and the Zeuthen Workshop on Cognitive Foundations of Economic Behavior for helpful comments and discussions. Maren Bermúdez Böckle and Cansu Toraman provided excellent research assistance. **Funding**: Support by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) through CRC TR 224 (Project A01) and ECONtribute (Germany's Excellence Strategy - EXC 2126/1-390838866), by the Bonn Graduate School of Economics (BGSE) as well as by the Joachim Herz Foundation is gratefully acknowledged. **Ethics**: We did not receive IRB approval since the University of Bonn does not have its own Institutional Review Board for economic experiments. The studies involved no deception and we obtained informed consent from all participants. **Preregistration**: Our data collections were preregistered at AsPredicted.org (numbers #174,937, #281,544 and #281,804). For more details, see Appendix C.

1 Introduction

Mental models are the subjective frameworks through which individuals perceive and interpret their environment, forming the foundation for inference, prediction, and decision-making in economic contexts. In settings where not all relevant variables are initially observable or attended to, economic agents make decisions using mental models that are formed on a subset of relevant variables. As additional variables become available, rational agents should revise their models to reflect the *joint* importance of both new and previously considered variables.

Consider, for example, a venture capitalist who needs to revisit her model of investment success as new big data analytics reveal additional insights on the determinants of startup success. Similarly, a stock market analyst predicting expected asset returns may form a preliminary model based on a set of risk factors and, when additional factors are proposed, fail to integrate them—instead shifting between simple models (Hong et al., 2007). Finally, a hiring manager who uses past work experience to infer qualification should evaluate this applicant characteristic differently once learning about systemic discrimination by other managers (Bohren et al. 2025; Pager 2003).

A common feature of all these dynamic learning environments is that irrespective of the number of observations they have seen so far, optimizing agents may need to fundamentally revise their models once data on additional variables become available. Such model revision operates jointly on two margins. First, the model's *structure*—the set of variables it includes (extensive margin)—must be updated by integrating newly predictive variables or discarding existing ones that become redundant. Second, the model's *parametrization*—the effect attributed to each included variable (intensive margin)—must be re-estimated conditional on all other included variables.

While an extensive literature provides evidence on how people update their beliefs within a fixed model structure, less is known about how people revise their models when the set of available variables expands. This raises two questions: How do people revise their mental models when confronted with new economic variables? Which cognitive mechanisms determine the revised models that individuals ultimately hold?

We address these questions by experimentally investigating how people revise mental models that capture statistical relationships in data when information on additional variables becomes available. Our focus is on path dependence in model formation, that is, the extent to which final models depend on models initially formed using only a subset of potentially relevant variables. Specifically, we hypothesize that mental models are *sticky*, meaning that individuals insufficiently revise their initial models in such dynamic learning environments.

We test this hypothesis using an experimental design that isolates path dependence

in model formation by varying the order in which variables are revealed while ensuring that all subjects ultimately observe the same complete dataset. In our Baseline experiment, we find that models are sticky. In two additional experiments that vary key features of the data-generating process, we corroborate this finding, pointing towards the generality of the phenomenon.

Understanding how people revise their models and identifying the mechanisms driving the identified learning dynamics is important for at least two reasons. First, from a theoretical perspective, most economic models assume that beliefs are updated seamlessly; however, cognitive frictions such as selective attention and inertia may prevent the correct integration of new information in models (e.g., Schwartzstein 2014), suggesting that theoretical frameworks of model learning could benefit from explicitly accounting for such updating distortions. Second, from a policy perspective, clarifying why individuals sometimes cling to outdated models can help design interventions that promote the adoption of new, model-relevant information.

Studying model revision in the field is difficult. Initial models are endogenous, and the set of variables an individual considers is unobserved. Even when revision does occur, we usually cannot tell whether it reflects learning about a previously neglected variable or instead updating based on additional observations on variables already in the model. Our two-stage experiment controls both confounds while capturing the essence of dynamic model learning.

In both stages of our experiments, individuals form prediction models about how the independent variables (or 'predictors') X and Z relate to the outcome variable Y (*Success* or *Failure*). In the first stage, subjects form a stochastic model based on a dataset describing the relationship between one randomly drawn variable and the outcome. We denote the two treatment groups by *X-first* and *Z-first*, respectively. In the second stage, the existing dataset expands by an additional column for the unobserved variable in the first stage, i.e., Z for the *X-first* group and X for the *Z-first* group. The key identifying idea is that treatment assignment affects the initial model subjects form, but not the information set available to them by the end of the experiment. Since the information revealed in the first stage is a subset of the information available in the second stage, a rational model-revising agent should arrive at identical second-stage models regardless of treatment.

In each stage, subjects see 40 data points that are initially revealed one by one so that subjects 'experience' the data. Subjects then encounter pairs of new projects with unknown outcomes that differ in the value of exactly one predictor and, as a result, potentially differ in their likelihood of success as well. They decide which project they prefer and report their willingness to pay (WTP) to switch projects. We subsequently elicit beliefs about the conditional success probabilities of projects across all possible

combinations of predictor values. This set of beliefs constitutes a complete statistical model of the relationship between the predictor and the outcome which closely maps to a parametrized linear regression. To corroborate our measure for subjects' statistical models, we additionally ask subjects at the end of the second stage whether they expect a ceteris paribus variation in X or Z to affect the outcome at all.

Beyond identification, the design has three further strengths. First, subjects begin by forming models using a single predictor, allowing us to study subsequent model revision without deception, as the induced change in the optimal model is brought about purely by an expansion in the model space. Second, the design provides benchmarks for both the unconditional marginal effect in the first stage and the conditional marginal effect in the second stage, enabling us to measure the extent to which subjects rely on each benchmark when forming and revising their models. Third, the two-stage structure partitions cognitive activity into formation and revision phases, allowing us to use stage-level response times to test mechanism predictions about effort allocation.

To guide our analyses and derive hypotheses, we introduce a simple framework of dynamic model formation. In this framework, the first-stage beliefs about success probabilities conditional on the initially observed predictor serve as defaults when subjects form beliefs conditional on both predictors in the second stage. Consequently, beliefs about the marginal effects of predictors remain partially anchored to the initial model. The framework predicts path dependence in model formation across a broad range of DGPs, provided that the variable observed in the first stage is predictive of the outcome. This takes the form of subjects who observe a predictor in the first stage assigning larger marginal effects to this predictor than subjects who observe it later.

We test our central hypothesis by examining how the predictor observed in the first stage shapes second-stage models. Specifically, we compare second-stage beliefs about the marginal effects of predictors at both the intensive and extensive margin across treatment groups. Beyond this reduced-form comparison, we estimate a structural specification derived from our framework, in which second-stage beliefs are decomposed into a weighted combination of the first-stage benchmark, which conditions only on the initially observed predictor, and the rational second-stage benchmark, which conditions on both predictors. Stickiness corresponds to continued reliance on the first-stage benchmark once the second-stage benchmark becomes available.

In our *Baseline* experiment, we consider a data-generating process that provides a particularly transparent setting for detecting stickiness. A key property of the baseline dataset is that, in the first stage, both predictors X and Z are correlated with the outcome, yet in the second stage, conditional on Z, the predictor X is uncorrelated with the outcome. The data shown to the *X-first* group in the first stage thus exhibits omitted variable bias, requiring subjects to strongly revise their model in the second stage. Since

the second-stage benchmark assigns zero weight to X, while the first-stage benchmark implies that X is predictive of the outcome, any additional weight placed on X by the *X-first* group reflects stickiness.

We find strong evidence supporting our hypothesis that initial exposure to a variable leads to stickiness in beliefs when subjects need to revise their models. After both treatment groups have been exposed to the same full dataset in the second stage of the experiment, subjects in the *X-first* group on average believe in a stronger marginal effect of X on the outcome and are more likely to perceive a *ceteris paribus* effect of X on the outcome. On the extensive margin, *X-first* subjects are 10 percentage points (pp) more likely to agree that variable X conditionally affects the probability of a successful outcome. On the intensive margin, subjects in the *X-first* group estimate the marginal effect of X on success to be, on average, 2.6 pp—about 50%—higher than their counterparts in the *Z-first* group. In contrast, both groups have similar beliefs regarding the marginal impact of Z on either margin. These treatment differences can be characterized by second-stage beliefs remaining partially influenced by the first-stage benchmark. Evidence from choice data between projects with different variable value combinations is qualitatively consistent with results on conditional beliefs, although these estimates are noisier and only partially statistically significant.

Having established 'sticky models' in the *Baseline* experiment, we investigate whether stickiness varies across subjects' approaches to the statistical problem. Using responses to an open-ended question about how they formed beliefs in the second stage, we classify subjects into commonly-observed reasoning approaches. While reasoning types strongly predict subjects' final models, we find little evidence that they differ systematically in the degree of stickiness, suggesting that sticky models reflect a general feature of model revision rather than being tied to a specific approach to extracting models from data.

To further examine the conditions under which sticky models arise, we assess whether path dependence depends on specific features of the data-generating process (DGP). We test this in two follow-up experiments that vary key properties of the DGP while retaining the same experimental design. The *SymmCorr* experiment makes the two predictors symmetric by equalizing their unconditional and conditional marginal effects while preserving the correlation structure of the baseline DGP. Both treatment groups therefore face statistically identical learning environments, isolating whether the asymmetric roles of predictors are necessary for stickiness.

The *SymmUncorr* experiment goes one step further by additionally removing the correlation between predictors and imposing constant marginal effects across stages. In this setting, the marginal effect learned about the initially observed predictor in the first stage is not affected by omitted variable bias and hence does not need to be 'un-

learned,' as it already equals the conditional effect of that predictor in the second stage. An alternative account of stickiness as a 'failure to unlearn' would therefore predict no treatment differences in *SymmUncorr*.

Contrary to the view that stickiness arises only for a narrow class of DGPs, we find strong evidence for sticky models in both additional experiments. This rules out a broad class of explanations that require asymmetry or correlated predictors as features of the data-generating process. In the *SymmCorr* experiment, subjects' beliefs about the marginal effect are about 3.4pp, or 20%, higher for the variable they observed in the first stage compared to the variable encountered only in the second stage. The corresponding difference in beliefs about the marginal effects of predictors in the *SymmUncorr* experiment is 2.7pp, or 18%. The structural specification again confirms that these treatment differences arise because second-stage models remain influenced by the initial model. Similar to the *Baseline* experiment, we find qualitatively consistent but noisier treatment effects in project choices and valuations.

But what explains model stickiness? Our framework predicts that subjects default to their first-stage model unless they invest enough cognitive effort to revise it. Reliance on first-stage benchmarks should therefore be strongest among subjects who form distinctive first-stage beliefs but invest relatively little effort in revising them in the second stage. To test this prediction, we use response time as a proxy for cognitive effort and measure the relative effort devoted to revision by the share of total response time spent in the second stage.

The framework's prediction is strongly supported in all three experiments. Performing a median split of subjects by relative effort, we show that stickiness is concentrated among subjects who spend a below-median share of total time and hence relatively low effort in the second stage. Among low-effort subjects, we find significant treatment differences in beliefs about the marginal effects of predictors. Across experiments, this corresponds to second-stage beliefs placing an additional weight of 10–16pp on the first-stage benchmark, rather than updating fully and relying solely on the second-stage benchmark. In contrast, high-effort subjects dedicating an above-median share of total time to the second stage exhibit no detectable stickiness: treatment differences are statistically indistinguishable from zero, and their beliefs place no detectable additional weight on the first-stage benchmark.

Taken together, our findings establish sticky models as a general feature of dynamic model learning across data-generating processes, tied to how much cognitive effort subjects allocate to forming versus revising their models. If anything, our experimental design likely yields a lower bound: we enforce attention to the expanding data, make all relevant data continuously available, and structure the revision task through elicitations that guide subjects toward the correct contingencies. That sticky models emerge

robustly even in this transparent environment suggests substantially greater path dependence in less guided real-world settings.

Related Literature: To the best of our knowledge, this is the first paper to empirically identify path dependence in model revision when subjects encounter new explanatory variables. In doing so, this project contributes to several strands of the literature.

First, this project contributes to a nascent empirical literature on how people form models based on data without explicit knowledge of the data-generating process. Fréchette et al. (2025) use a closely related setting with two binary inputs and one output to study how people learn stochastic relationships from datasets, identifying two prominent types of errors in model learning: failures to properly condition on correlations and failures to use correlations optimally. Similarly, Kendall and Oprea (2024) study how people learn models from data but focus on deterministic rules of varying complexity. While these papers share our paradigm of people extracting models from data, they hold the set of observable variables fixed throughout. We instead introduce an exogenous expansion of the variable set across stages, which lets us measure how subjects revise their models in response to a well-defined change in the model space.

A related set of papers instead fixes the data and studies the effects of providing people with one or multiple models. Charles and Kendall (2024) experimentally study how externally provided causal models influence data interpretation, supporting the core predictions of Eliaz and Spiegler (2020), while Ambuehl and Thysen (2025) study how individuals choose among competing causal interpretations. Barron and Fries (2025) test the model-based persuasion framework of Schwartzstein and Sunderam (2021), showing that models that fit the data well are persuasive. A recent experimental literature further examines how people weight models when facing model uncertainty, documenting that people tend to commit to a single model rather than averaging across alternatives, with consequences for disagreement, overprecision, and belief volatility (Aina and Schneider, 2025; Augenblick et al., 2025; Fan and Fries, 2026; Musolff and Zimmermann, 2025). Whereas these papers provide subjects with models, we have them form their own models from data and measure whether they adjust both the variables they attend to (extensive margin) and the effects they attribute to them (intensive margin) when new data become available.

Regarding the dynamics of mental models, the most closely related paper is Esponda et al. (2024), who experimentally show that misspecified models can persist despite regular feedback. We share their interest in the persistence of misspecified models but differ in three key respects. First, we study mental models that capture the statistical relationships between multiple variables at both intensive and extensive margins, and subjects must extract these models from data. Second, we force an exogenous model revision through a randomized expansion of the model space, which allows us to cleanly

measure model revisions independently of prior mistakes. Third, we provide novel evidence on the cognitive mechanisms underlying failures in model revision.

Several field studies document patterns consistent with insufficient model revision. Hanna et al. (2014) show that seaweed farmers neglect pod size—a relevant production input—and that observing data alone is insufficient for them to attend to it. Their setting differs from ours in that farmers could, in principle, attend to all relevant inputs from the beginning, whereas we study model revision when relevant relationships in the data can only be learned later on. Han et al. (2026) additionally show in a large-scale field setting that managers' mental models of optimal pricing differ systematically with cognitive skills and persist despite high stakes, extensive experience and feedback. Liu and Zhang (2026) find that initial narratives shape how people interpret subsequent information, and Macchi (2023) show that first impressions have lasting effects on loan decisions despite detailed financial information being revealed subsequently. Our experiment complements this evidence by testing model stickiness in a controlled environment where we provide all available data and prompt subjects to think about the relevant contingencies.

We also add to an extensive literature on path dependence in belief updating, such as confirmation bias (Rabin and Schrag, 1999), prior-biased updating (see Benjamin 2019 for a review), or anchoring-and-adjustment (Tversky and Kahneman, 1974). Unlike previous work that holds the model space fixed, our study examines a setting in which the set of explanatory variables expands, requiring subjects to revise models of a coarser distribution $Y|X$ (or $Y|Z$) into models of the refined distribution $Y|X,Z$. This expansion increases the complexity of the inference problem, connecting our work to a literature documenting belief biases under complexity (e.g., Enke 2020; Enke and Zimmermann 2019; Enke and Graeber 2023; Martínez-Marquina et al. 2019; Niederle and Vespa 2023). We show that model stickiness emerges as a simplification strategy in this environment: as the variable set expands, subjects remain influenced by their initial, lower-dimensional model.

Stickiness in our data is predicted by relative cognitive effort allocation: subjects who allocate little effort to the revision stage exhibit substantial stickiness, while those who invest more exhibit no stickiness. This pattern connects to a literature on attention as a scarce cognitive resource. Sparsity (Gabaix, 2014) and rational inattention models (Maćkowiak and Wiederholt, 2009; Sims, 2003) characterize agents who optimally allocate limited attention across information sources, with inattention typically inducing attenuation toward a generic default action or prior. Our setting documents an endogenous informational default—the agent's own prior model—akin to Dean et al. (2026) and Graeber and Enke (2026).

Finally, this project relates to a growing theoretical literature on misspecified models

and their implications. Several cognitive mechanisms have been proposed to give rise to model misspecification, most prominently limited attention (e.g., Bordalo et al. 2026; Gabaix 2019; Hanna et al. 2014; Schwartzstein 2014), memory limitations (Bordalo et al., 2023; Gennaioli and Shleifer, 2010), and misperceptions regarding the data-generating process (Fudenberg and Lanzani, 2023). Some recent papers also theoretically study the dynamics of misspecified models: Ortoleva (2012) axiomatically characterizes when agents should abandon their current model in favor of a new one, showing that sufficiently surprising evidence triggers a rational "paradigm shift." Gagnon-Bartsch et al. (2023) study the conditions under which agents become aware of model misspecification, while Lanzani (2025) studies how the concern of being misspecified affects dynamic model choice. Ba (2026) examines which misspecified models persist in the long run and finds that simple models can be more robust than correctly specified complex models. Closely related to our setting, Guarda et al. (2026) characterize when agents working with a limited set of variables should optimally acquire an additional one at a cost, providing a normative benchmark for the decision our subjects face of whether to integrate a new variable into their model. We provide a simple theoretical model to derive our main predictions and empirically inform the theoretical literature by providing evidence on a key friction in model revision.

2 Experimental Design

Our goal is to study *path dependence* in model formation and revision. Because confounding factors make this difficult to identify with observational data, we design a minimally framed experiment that isolates cognitive primitives from alternative explanations such as motivated reasoning or heterogeneity in priors.

2.1 Setting and Timeline

In the experiment, subjects assume the role of entrepreneurs tasked with identifying which projects are more likely to be successful among a set of projects with different variable value combinations. To do so, they learn about the impact of independent variables on a dependent outcome variable using a dataset of 40 rows. The complete data includes two independent variables, Color (Blue or Green) and Card (Diamonds or Clubs), and an outcome variable (Success or Failure). Henceforth, we denote the variables Color and Card as X and Z, respectively, and the project outcome by Y.

Figure 1 Experimental Timeline

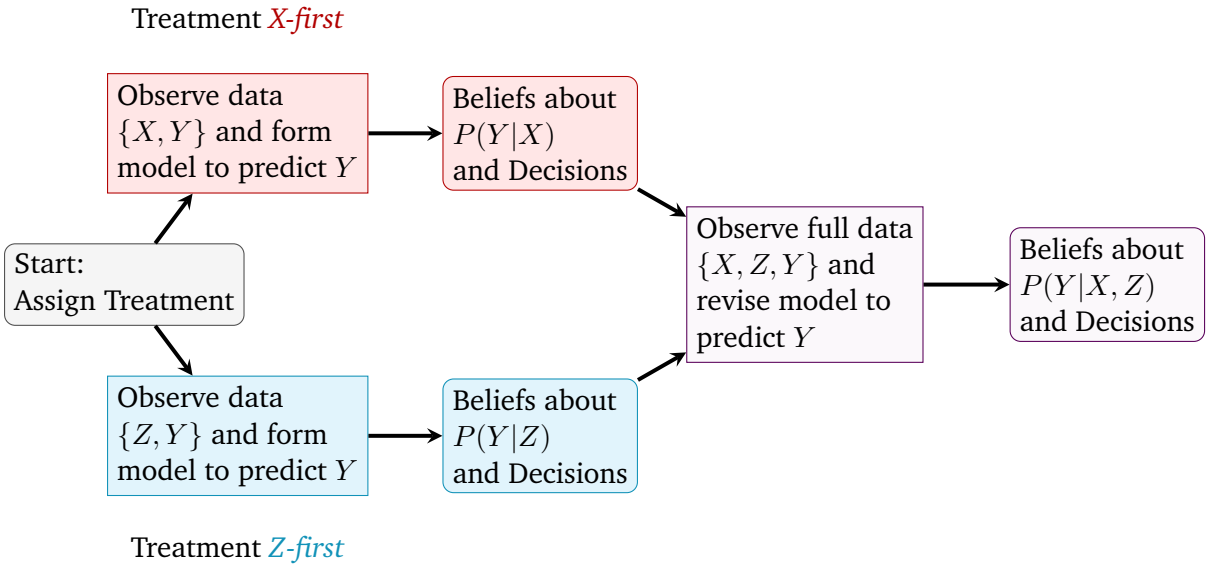


Figure 1 shows the timeline of our experiment. Initially, subjects have to pass a basic attention check before acquainting themselves with the general setting. Only subjects that pass comprehension questions can proceed to the main part. The main part consists of two stages. Therein, subjects are randomly assigned to a treatment group that exogenously varies the predictor in the first stage.

In the first stage, subjects observe past data on only one randomly selected variable (X or Z) and the project's outcome. We refer to the treatment groups generated by the random assignment to the variable observed in the first stage as *X-first* and *Z-first*, respectively. Subjects are told that each project has two predictors but that they are only able to observe one randomly selected variable.[1]

In the second stage, we reveal the formerly missing second variable as an additional column in the same dataset. A comprehension question ensures that subjects understand that the observations from the second-stage dataset are identical to the first stage except for the additional predictor. Further, in each stage, we ensure that subjects pay at least some attention to the data by revealing data points one by one until the table fills up, after which they can proceed to the respective tasks. At the end of each stage, we elicit subjects' models about the relationship between the predictors and the outcome using reported beliefs, binary project choices, and willingness to pay between projects. At the end of the second stage, we elicit an additional measure about the perceived marginal impact of each variable. In the *Baseline* experiment, we further ask subjects to

[1]Announcing the full information structure at the beginning of the experiment reduces heterogeneity in beliefs about the experimental structure by treatment without making it explicit that more information will be available after the first stage. We explain to subjects that they observe a randomly selected variable in the first stage to avoid experimenter demand effects.

describe their approach to the task in an open-ended question.[2]

2.2 Baseline Data-Generating Process

As shown in Figure 2, the data consists of two variables X and Z, and an outcome Y. All variables are binary and the data contains 40 observations that perfectly resemble the following relationships between the predictors and the outcome:

$$
\begin{aligned}
P(Y|X,Z) &= 0.2 + 0.6 \cdot Z \\
P(Y|Z) &= 0.2 + 0.6 \cdot Z \\
P(Y|X) &= 0.35 + 0.3 \cdot X \\
P(Y) &= 0.5
\end{aligned}
$$

N°	Color	Card	Outcome	N°	Color	Card	Outcome	N°	Color	Card	Outcome	N°	Color	Card	Outcome
1	🟢	♣	Failure	11	🟢	♣	Failure	21	🟢	♣	Failure	31	🟢	♣	Failure
2	🔵	♦	Success	12	🔵	♦	Success	22	🟢	♣	Success	32	🔵	♣	Failure
3	🔵	♦	Success	13	🟢	♣	Failure	23	🟢	♣	Failure	33	🔵	♣	Failure
4	🟢	♣	Failure	14	🟢	♦	Failure	24	🔵	♦	Success	34	🔵	♦	Success
5	🟢	♦	Success	15	🔵	♦	Success	25	🔵	♦	Success	35	🔵	♦	Success
6	🔵	♦	Success	16	🔵	♣	Failure	26	🔵	♣	Success	36	🟢	♣	Failure
7	🟢	♣	Failure	17	🔵	♦	Failure	27	🔵	♦	Success	37	🟢	♣	Success
8	🔵	♦	Failure	18	🔵	♦	Failure	28	🟢	♣	Failure	38	🔵	♦	Success
9	🔵	♦	Success	19	🟢	♦	Success	29	🔵	♣	Failure	39	🟢	♦	Success
10	🟢	♣	Failure	20	🟢	♣	Success	30	🟢	♦	Success	40	🟢	♣	Failure

Figure 2 Screenshot of data of past projects as observed in the second stage

Subjects learn that future projects with unrealized outcomes are drawn according to the relationship described by the full data table on past projects, which they have to extract by studying the data. To further minimize the role of subjects' prior beliefs about the DGP, we follow Charles and Kendall (2024) and inform subjects that each row in the dataset corresponds to one thousand projects with identical variable value combinations and outcomes. Using comprehension questions, we ensure that subjects have, in fact, understood these properties of the DGP.

In the first stage, where subjects only encounter X or Z, the rational benchmark is

[2]We do not elicit self-reported reasoning for the symmetric DGPs since several reasoning types (see Section 4.3) imply the same predicted beliefs in these DGPs. We choose to elicit subjects' reasoning last in order not to influence their other choices or beliefs.

thus given by the empirical probability of success conditional on the observed variable, i.e., $P(Y|X)$ for the *X-first* group and $P(Y|Z)$ for the *Z-first* group. In the second stage, the rational benchmark for both groups is the empirical probability of success conditional on both variables, i.e., $P(Y|X,Z)$, since the first-stage dataset is a subset of the second-stage dataset.

In the baseline DGP, Z (Card) is highly predictive of Y, while X (Color) has no predictive power when controlling for Z. Because of its correlation with Z, X is moderately predictive of the outcome when not conditioning on Z. The data structure thus implies a significant change in the rational benchmark across stages for the group that observes X first. In contrast, the same rational benchmark applies in both stages for the *Z-first* group.

As the second-stage benchmark implies that subjects should place zero marginal weight on X, any additional weight placed on X by the *X-first* group relative to the *Z-first* group naturally reflects insufficient revision of an initially formed model, making this DGP a particularly transparent setting for detecting stickiness.

2.3 Additional Data-Generating Processes

To test the generality of model stickiness and to isolate the cognitive mechanisms underlying it, we conducted two additional pre-registered experiments that retain the same experimental design while varying the DGP.

Table 1 Summary of data-generating processes

	Baseline	**SymmCorr**	**SymmUncorr**
Δ_X	0.30	0.50	0.30
Δ_Z	0.60	0.50	0.30
$\Delta_{X\|Z}$	0	0.3125	0.30
$\Delta_{Z\|X}$	0.60	0.3125	0.30

Notes: The table summarizes the data-generating processes used across the different experiments. $\Delta_X \equiv P(Y{=}1 \mid X{=}1) - P(Y{=}1 \mid X{=}0)$ and $\Delta_Z \equiv P(Y{=}1 \mid Z{=}1) - P(Y{=}1 \mid Z{=}0)$ denote unconditional marginal effects of predictors X and Z, respectively. $\Delta_{X|Z} \equiv P(Y{=}1 \mid X{=}1, Z) - P(Y{=}1 \mid X{=}0, Z)$ and $\Delta_{Z|X} \equiv P(Y{=}1 \mid X, Z{=}1) - P(Y{=}1 \mid X, Z{=}0)$ denote marginal effects conditional on the other predictor. Screenshots of the datasets as displayed to subjects are shown in Figure 2 (Baseline), Figure F.1 (*SymmCorr*), and Figure F.2 (*SymmUncorr*).

Table 1 summarizes the three DGPs used across the experiments. Both alternative DGPs are symmetric, in the sense that the empirical benchmarks for the two predictors' marginal effects are identical both in the first stage and after fully conditioning in the second stage. This eliminates systematic differences between treatment groups beyond the predictor subjects initially form a model about.

To further control for potential confounds, we additionally cross-randomize (i) the

mapping between variable labels and dataset columns, and (ii) the order in which beliefs are elicited in the second stage. These cross-randomizations allow us to rule out that results are driven by order effects arising from the sequence in which observations are presented or from the order in which beliefs are elicited, respectively.

2.4 Measurement of Models and Incentives

We measure subjects' mental models—the set of subjective beliefs about the marginal effects of the predictor(s) on the outcome—primarily by eliciting beliefs about conditional success likelihoods for projects with different combinations of predictor values in both stages. As a validation measure, we collect a coarser, qualitative assessment of perceived marginal effects. We further complement our evidence on mental models with data on underlying cognitive processes and resulting decisions (binary choices and willingness-to-pay between projects).

Both conditional belief elicitation and decision tasks (binary and WTP) are incentivized by an additional bonus payment. A random subsample of 10% of subjects is selected for bonus eligibility, and for these subjects, one of their incentivized choices or beliefs is randomly drawn to determine their actual bonus payment.

Mental models (intensive margin): In each stage, subjects state their belief about the success probability of future projects with a particular variable value (first stage) or combination of variable values (second stage) on a slider going from 0% to 100%. By design, this leads to two beliefs being elicited in the first stage and four beliefs elicited in the second stage.[3] We use the following question to elicit subjects' beliefs:

How likely do you think it is that project [description of the project's variable values] will be successful?

These beliefs can be directly translated into beliefs about the marginal effects of variables X and Z by taking the difference in perceived success likelihood between two projects that differ in exactly one variable. As pre-registered, beliefs about the marginal effects of these variables constitute our primary measure of mental models. The elicited beliefs are incentivized using the binarized scoring rule, where subjects can receive a bonus payment of \$10 depending on the accuracy of the stated belief. In each stage, we further ask subjects a non-incentivized slider question asking for their confidence (in %) that all of their stated beliefs were within ± 5 percentage points of

[3]Because the elicited contingencies expand from two beliefs in the first stage to four in the second, numerical anchoring on previously stated beliefs requires some deliberation about how the first and second stage contingencies map to one another.

the true success likelihoods of the respective projects.

Mental models (extensive margin): As a validation measure, we elicit subjects' explicit recognition of marginal effects at the end of the experiment. Specifically, subjects indicate whether they believe each of the following two statements to be true or false:

Statement 1: Assuming that a project's Card remains fixed, changing a project's Color has an effect on the project's success probability.

Statement 2: Assuming that a project's Color remains fixed, changing a project's Card has an effect on the project's success probability.

Decisions: In both stages, we also elicit model-driven decisions, which serve as secondary, indirect measures of mental models. These consist of:

1. a binary choice between two future projects that differ in exactly one predictor's value; and
2. willingness-to-pay (WTP) to remain with their chosen project.

In the binary choice task, subjects select their preferred project, incentivized by a bonus payment of \$10 if the selected project is successful (and \$0 otherwise). Following their choice, subjects indicate their WTP to keep their preferred project over the alternative, using a multiple price list (MPL) with 21 rows (offering fixed payments from \$0 to \$10, in \$0.50 increments, for switching). One MPL row is randomly selected for determining a possible bonus payment.

Model formation process: To complement our measures of mental models, we directly probe subjects' underlying cognitive processes in the *Baseline* experiment. After the second-stage conditional belief elicitation, subjects provide an open-ended description of how they formed their beliefs, using the following prompt:

Please describe how you determined the projects' success likelihoods. You should explicitly state what you paid attention to and which strategy you used to arrive at your response in full sentences.

3 Framework and Research Hypotheses

To guide the analysis and derive hypotheses, we develop a simple framework of dynamic model formation in which a decision-maker (DM) constructs a mental model of how predictors affect project success based on observed data. With binary predictors and outcomes, the DM's mental model is fully characterized by beliefs about success probabilities conditional on observed predictors.

The key feature of the framework is that the model space expands over time. We discuss both model formation in Stage 1 and model revision in Stage 2, formalizing each stage as a distinct statistical inference problem.

3.1 Stage 1: Initial Model

In Stage 1, the DM observes a single predictor $S_1 \in \{0, 1\}$ and forms beliefs about $P(Y = 1 \mid S_1)$, the success probability conditional on S_1. The empirical benchmark for this conditional success probability is denoted by $P^B(Y \mid S_1)$.[4] We model beliefs as a convex combination of the empirical benchmark and a cognitive default $d^{(1)}$:

$$\mu^{(1)}_{S_1}(e_1) = \lambda(e_1) \cdot P^B(Y \mid S_1) + (1 - \lambda(e_1)) \cdot d^{(1)} \tag{1}$$

The function $\lambda(e_1) \in [0, 1]$ captures the weight placed on the empirical benchmark and is weakly increasing in effort e_1. Conversely, the degree of attenuation toward the default is captured by $1 - \lambda(e_1)$. By modeling beliefs as a convex combination of the parameter of interest and a fixed default, we follow a large literature (see Enke 2026 for a review).[5]

The default $d^{(1)}$ captures the belief the DM would hold absent deliberation. We assume that this default is independent of S_1. Under this assumption, beliefs about the marginal effect of S_1 take the form:

$$\mu^{(1)}_{\Delta_{S_1}}(e_1) := \mu^{(1)}_{S_1=1}(e_1) - \mu^{(1)}_{S_1=0}(e_1) = \lambda(e_1) \cdot \Delta^{(1)}_{S_1}$$

where $\Delta^{(1)}_{S_1} = P^B(Y = 1 \mid S_1 = 1) - P^B(Y = 1 \mid S_1 = 0)$ is the empirical benchmark for the marginal effect of S_1. Without loss of generality, we assume that $\Delta^{(1)}_{S_1} \geq 0$.

[4] In our setting, agents observe a large number of signals, so empirical frequencies provide a close approximation to Bayesian posteriors.

[5] While we assume that attenuation decreases with more cognitive effort, we do not take a stance on the exact source of attenuation. In our experimental setting, attenuation may arise from uncertainty about how to map observed frequencies of S_1 and Y into beliefs about marginal effects (e.g. Enke and Graeber, 2023; Ilut and Valchev, 2022), or from noisy perception of the data (e.g. Gabaix, 2019; Khaw et al., 2020).

3.2 Stage 2: Model Revision

In Stage 2, the DM observes an additional predictor S_2 and forms beliefs about $P(Y = 1 \mid S_1, S_2)$, the success probability conditional on both predictors, for which the empirical benchmark is denoted by $P^B(Y \mid S_1, S_2)$.[6] The DM now faces a related but distinct statistical inference problem. We assume that beliefs are formed analogously to Stage 1, except that the default for $P(Y \mid S_1, S_2)$ now depends on the value of the predictor observed in the first stage and is given by the corresponding first-stage belief, i.e., $d^{(2)} = \mu^{(1)}_{S_1}(e_1)$.[7] Beliefs in Stage 2 therefore take the following form:

$$\mu^{(2)}_{S_1,S_2}(e_1, e_2) = \lambda(e_2) \cdot P^B(Y \mid S_1, S_2) + (1 - \lambda(e_2)) \cdot \mu^{(1)}_{S_1}(e_1) \tag{2}$$

The function $\lambda(e_2) \in [0, 1]$ captures the weight placed on the full data observed in Stage 2 relative to first-stage beliefs and is weakly increasing in effort e_2. As effort increases, beliefs converge toward the empirical benchmark $P^B(Y \mid S_1, S_2)$.[8]

Substituting the first-stage belief formation process from equation 1 into equation 2, second-stage beliefs can be expressed as:

$$\mu^{(2)}_{S_1,S_2}(e_1, e_2) = \lambda(e_2) \cdot P^B(Y \mid S_1, S_2) + (1 - \lambda(e_2)) \cdot \lambda(e_1) \cdot P^B(Y \mid S_1) + \nu(e_1, e_2) \cdot d^{(1)} \tag{3}$$

where $\nu(e_1, e_2) = (1 - \lambda(e_2)) \cdot (1 - \lambda(e_1))$. Second-stage beliefs therefore constitute a weighted average of the second-stage empirical benchmark, the first-stage empirical benchmark, and the first-stage cognitive default.[9]

[6] Since the information received in Stage 1 is nested within the information received in Stage 2, Bayesian updating is path independent: regardless of which predictor is observed in Stage 1, a Bayesian decision maker arrives at the same posterior once the full Stage 2 information is observed. This posterior is again closely approximated by the empirical frequencies.

[7] A DM seeking to reduce cognitive effort may adopt this default (possibly consciously) as a simplification strategy, as in theories of attribute substitution (Kahneman and Frederick, 2002). Alternatively, first-stage beliefs about S_1 may serve as a default because they reflect a genuine prior, or because they act as a numerical anchor (Tversky and Kahneman, 1974). We remain agnostic about the precise psychological origin of the cognitive default.

[8] The framework therefore abstracts from heterogeneity in how individuals translate observed data into beliefs, assuming that with sufficient cognitive effort they converge to the same solution. We discuss this assumption in more detail in Appendix D.1 and empirically examine the role of reasoning in Section 4.3.

[9] A framework with sequential Bayesian updating and underreaction to information due to cognitive noise would yield qualitatively similar predictions. The key assumption that such a Bayesian framework requires is that at all stages, including during Stage 1, the DM has a well-defined belief over the joint distribution $P(Y, S_1, S_2)$. After observing only (S_1, Y) in Stage 1, the Bayesian posterior over $P(Y = 1 \mid S_1, S_2)$ conditional on this partial information is independent of S_2, with posterior means that lie between the prior mean of 0.5 and the posterior mean of $P(Y = 1 \mid S_1)$, hence resembling the cognitive default in our framework. Note that updating beliefs in this way is extremely demanding and the DM may not even have a well-defined prior over the joint distribution before entering Stage 2. Our formulation therefore provides a simpler representation that is more closely aligned with subjects' reported approaches to the task.

3.2.1 Beliefs about the first-stage variable

Following equation 3, beliefs about the marginal effect of S_1, conditional on S_2, are given by:

$$\mu^{(2)}_{\Delta_{S_1|S_2}}(e_1, e_2) = \lambda(e_2) \cdot \Delta^{(2)}_{S_1|S_2} + (1 - \lambda(e_2)) \cdot \lambda(e_1) \cdot \Delta^{(1)}_{S_1}$$

where $\Delta^{(2)}_{S_1|S_2} = P^B(Y = 1 \mid S_1 = 1, S_2) - P^B(Y = 1 \mid S_1 = 0, S_2)$ is the empirical benchmark for the marginal effect of S_1 conditional on S_2. Beliefs about the marginal effect of S_1 conditional on S_2 are therefore anchored to the corresponding first-stage beliefs in the marginal effect.

3.2.2 Beliefs about the second-stage variable

It follows from equation 3 that second-stage beliefs that share the same value of S_1 have the same underlying default. Beliefs about the marginal effect of S_2, conditional on S_1, therefore take the form:

$$\mu^{(2)}_{\Delta_{S_2|S_1}}(e_2) = \lambda(e_2) \cdot \Delta^{(2)}_{S_2|S_1}$$

where $\Delta^{(2)}_{S_2|S_1} = P^B(Y = 1 \mid S_2 = 1, S_1) - P^B(Y = 1 \mid S_2 = 0, S_1)$ is the empirical benchmark for the marginal effect of S_2 conditional on S_1. In contrast to S_1, beliefs about the marginal effect of S_2 resemble the form of first-stage beliefs, in that both default to zero in the absence of deliberation.

3.2.3 Differences in beliefs across treatment groups

Let S_1 and S_2 denote the first- and second-stage variables in one treatment group, and $\tilde{S}_1$ and $\tilde{S}_2$ those in the other treatment group. By design, the first-stage variable in one treatment corresponds to the second-stage variable in the other, i.e., $S_1 = \tilde{S}_2$ and $S_2 = \tilde{S}_1$.

Assume that effort in Stage 1, denoted by $\bar{e}_1$, and effort in Stage 2, denoted by $\bar{e}_2$, are identical across individuals and treatment groups.[10] Then, the difference in second-stage beliefs across treatment groups is given by:

$$\mu^{(2)}_{\Delta_{S_1|S_2}}(\bar{e}_1, \bar{e}_2) - \mu^{(2)}_{\Delta_{\tilde{S}_2|\tilde{S}_1}}(\bar{e}_1, \bar{e}_2) = (1 - \lambda(\bar{e}_2)) \cdot \lambda(\bar{e}_1) \cdot \Delta^{(1)}_{S_1} \tag{4}$$

[10] For simplicity, we assume identical effort levels across individuals and treatment groups. In practice, effort may differ across groups if, for example, first-stage predictors vary in predictive strength and therefore affect incentives to revise beliefs in the second stage. We discuss the implications of heterogeneous effort across groups in Appendix D.2.

This difference arises because DMs enter Stage 2 with different defaults induced by their first-stage beliefs. It is, ceteris paribus, increasing in the unconditional marginal effect $\Delta_{S_1}^{(1)}$ and in first-stage effort $\bar{e}_1$, and decreasing in second-stage effort $\bar{e}_2$.

3.3 Sticky Models

We refer to models as *sticky* when people insufficiently revise their initial models as they observe additional information, which corresponds to $\lambda(e_2) < 1$. Following equation 3, if $\lambda(e_1) > 0$, such that first-stage beliefs place positive weight on the information received in Stage 1, stickiness implies that second-stage beliefs remain partially anchored to the corresponding first-stage empirical benchmark. Equation 4 then implies that, for data-generating processes for which $\Delta_{S_1}^{(1)} > 0$, beliefs about the marginal effect of a predictor differ across treatment groups. In particular, individuals who observed a predictor already in Stage 1 assign a larger marginal effect to that predictor than individuals who first observed the same predictor in Stage 2.

Hypothesis 1 (Sticky models): Models are sticky, leading second-stage models to remain partially anchored to the corresponding first-stage empirical benchmark. In particular, for data-generating processes for which $\Delta_{S_1}^{(1)} > 0$, individuals who observed a predictor already in Stage 1 assign a larger marginal effect to that predictor than individuals who first observed the same predictor in Stage 2.

Hypothesis 1 formalizes the central claim that individuals' mental models are 'sticky.' In this conceptualization, stickiness is a general phenomenon of path dependence in model formation that arises across data-generating processes for which the first-stage variable is unconditionally predictive, that is, $\Delta_{S_1}^{(1)} > 0$. An alternative hypothesis is that differences in second-stage models arise only under specific conditions, because particular properties of the data-generating process are required for sticky models to emerge.

To test the generality of hypothesis 1, we study sticky models across three data-generating processes for which $\Delta_{S_1}^{(1)} > 0$, and that vary *correlation* and *symmetry* of predictors — two candidate properties that underlie sticky models in alternative conceptualizations.[11]

We test for sticky models by comparing second-stage beliefs about the marginal effects of predictors across treatment groups in each data-generating process. To confirm that any observed differences are driven by second-stage beliefs remaining anchored to first-

[11] Correlation between variables would drive stickiness if it reflects a failure to unlearn the marginal effects of the first stage variable. Symmetry would drive sticky models if differential predictiveness in first-stage models or the perceived size of the relative marginal effects in Stage 2 were to play an outsized role in model revision.

stage beliefs, we estimate the weight that second-stage beliefs place on the different empirical benchmarks following equation 3. The exact empirical specification depends on whether the first-stage and second-stage benchmarks (partially) coincide or can be sharply distinguished.

3.4 Cognitive effort

The framework further implies that the extent to which second-stage beliefs remain anchored to the first-stage empirical benchmark depends on effort exerted in both Stage 1 and Stage 2. Specifically, effort in Stage 1 determines the degree of attenuation, $\lambda(e_1)$, in first-stage beliefs: higher effort leads first-stage beliefs to place greater weight on the information received in Stage 1, thereby increasing differences in defaults across treatment groups and amplifying differences in second-stage beliefs. By contrast, effort in Stage 2 determines the degree of attenuation, $\lambda(e_2)$, in second-stage beliefs: higher effort leads second-stage beliefs to place greater weight on the information received in Stage 2 and less weight on first-stage beliefs, thereby reducing differences in second-stage beliefs across treatment groups. Consequently, effort exerted in Stage 2 relative to Stage 1 is a key determinant of the extent to which second-stage beliefs remain anchored to the first-stage empirical benchmark and therefore of belief differences across treatment groups.

Hypothesis 2 (Cognitive effort and stickiness): Greater effort exerted in Stage 2 relative to Stage 1 reduces the extent to which second-stage beliefs remain anchored to the first-stage empirical benchmark, thereby reducing belief differences across treatment groups.

4 Baseline Results

In this section, we present the results of our *Baseline* experiment. We begin with reporting summary statistics. We then test whether subjects' second-stage models depend on which predictor they observed first, using both reduced-form and structural specifications. Finally, we examine how reasoning types relate to stickiness.

4.1 Summary Statistics

We conducted the *Baseline* experiment in May 2024 on the online platform Prolific, which is generally known for having a high-quality pool of subjects (Peer et al., 2022). As pre-registered, we collected 800 complete responses and exclude subjects who are

among the 1% fastest and the 1% slowest as measured by the total time spent on our experiment. This screening device excludes 16 subjects, which brings our total sample size to $n = 784$. Subjects spent on average 26 minutes (median: 23 minutes) on the survey. Before participating in the main part of the survey, subjects are carefully introduced to the setting. We ensure their understanding with a set of comprehension questions.[12] Subjects were paid $4 for completing the experiment and received a bonus payment of $0.59 on average.

As shown in Appendix Table A.1, demographics are well-balanced across the *X-first* and *Z-first* treatment groups, with no statistically significant differences between them. Accordingly, we do not include demographic controls in our main regression specifications. Compared with the U.S. population, however, our sample tends to be younger and more highly educated, which is a common phenomenon in online samples Snowberg and Yariv (2021).

Many subjects are generally able to extract the conditional success probabilities from the dataset. Appendix Figures B.1 and B.2 plot the raw belief distributions against the rational benchmarks for the first and second stages, respectively. In every cell, the benchmark corresponds to the modal response. In the first stage, between 42–57% of subjects report beliefs within 5pp of the benchmark across the two beliefs, while in the second stage 29–40% of subjects fall within 5pp of the benchmark across the four beliefs.

4.2 Path Dependence in Model Formation

We start by presenting reduced-form evidence on how second-stage models and decisions vary across treatments, and then turn to a structural approach to assess how closely subjects' beliefs align with the empirical benchmarks.

4.2.1 Reduced-form Evidence

The core feature of subjects' prediction models is the role they attribute to each variable in affecting project outcomes. As pre-registered, we therefore examine whether the predictor observed in the first stage shapes subjects' final beliefs about the marginal impact of each variable on project success. Stickiness implies a stronger reliance on the initial model, predicting that the *X-first* group assigns an at least weakly greater marginal impact to X than the *Z-first* group, and vice versa for Z.

Table 2 examines how second-stage beliefs about the marginal impact of each predic-

[12] 72% of all Prolific users attempting the comprehension check pass it in the first attempt, 14% in the second attempt. The remaining 14% who fail comprehension questions twice are screened out and cannot participate in our main experiment.

tor vary by treatment. We test for path dependence at two margins: the *extensive margin* (Columns 1–2) captures whether subjects believe a variable has any ceteris paribus effect on the outcome; the *intensive margin* (Columns 3–8) captures quantitative beliefs about marginal effects. Columns 3 and 4 provide the primary test by pooling beliefs in the marginal effect of a predictor across values of the other variable.[13] Columns 5–8 report the disaggregated specification used in our pre-registered hypothesis.

Table 2 Treatment effect on second-stage models

	Extensive margin Share agreeing (in %)		**Intensive margin** $\Delta P(Y = \text{Success} \mid \cdots)$		**Intensive margin - disaggregated** $\Delta P(Y = \text{Success} \mid \cdots)$			
	X matters	Z matters	ΔX	ΔZ	ΔX, Z = 0	ΔX, Z = 1	ΔZ, X = 0	ΔZ, X = 1
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
X-first	9.962***	1.143	2.600**	2.127	1.458	3.741**	0.985	3.268*
	(3.130)	(2.323)	(1.176)	(1.771)	(1.497)	(1.669)	(2.171)	(1.955)
Constant	68.718***	87.436***	4.687***	30.010***	−0.610	9.985***	24.713***	35.308***
	(2.351)	(1.680)	(0.903)	(1.253)	(1.134)	(1.185)	(1.541)	(1.352)
Observations	784	784	1,568	1,568	784	784	784	784
R^2	0.013	0.000	0.003	0.001	0.001	0.006	0.000	0.004

Notes: This table presents treatment effects from an exogenous manipulation of the first-stage independent variable on subjects' mental models of project success in the second stage. *X-first* is an indicator equal to 1 if the subject observed X in the first stage. Columns 1 and 2 report the share of subjects agreeing that the predictors X and Z, respectively, have a *ceteris paribus* effect on the outcome. Columns 3 and 4 report beliefs about the marginal effects of X and Z, respectively, pooling across the two possible values of the other variable. For each subject, there are two observations, corresponding to the two values of the other variable. Columns 5-8 report beliefs about the marginal effects of X and Z separately for each value of the other predictor. Clustered standard errors (Columns 1-4) and robust standard errors (Columns 5-8) are in parentheses. * denotes significance at 10 pct., ** at 5 pct., and *** at 1 pct. level.

The main finding is that subjects' beliefs exhibit path dependence at both margins. Those who initially observed X are more likely to believe that a ceteris paribus change in X affects the outcome and attribute a higher marginal effect to X than those who saw Z first. Beliefs about Z, by contrast, do not differ significantly across groups.[14]

At the extensive margin, we observe that about 69% of the subjects in *Z-first* believe that X has an effect on the outcome, whereas 10pp more subjects in *X-first* hold that belief ($p = 0.002$). Across both treatment groups, nearly 90% believe that Z affects the probability of success, indicating a consensus on Z's role across treatments with

[13] Note that in our DGP, the empirical marginal effects do not depend on the value of the other variable such that $\Delta X = (\Delta X, Z = 0) = (\Delta X, Z = 1)$ and $\Delta Z = (\Delta Z, X = 0) = (\Delta Z, X = 1)$. Thus, pooling across both variable values of the variable that remains constant does not obfuscate the benchmarks.

[14] The simple framework predicts differences in beliefs for both X and Z. There are several reasons why differences may be more difficult to detect for Z. First, if the stronger unconditional effect of Z reduces incentives to learn in Stage 2, DMs in the *Z-first* treatment may exert less effort. As discussed in Section D.2, this would attenuate differences in beliefs about Z. Consistent with this interpretation, subjects in the *X-first* treatment spend slightly more time on the second stage (see Table A.17). Second, the relationship between Z and the outcome may be particularly salient because the true marginal effect is large. If beliefs about the effect of Z are already close to a ceiling, exposure to Z in the first stage may have limited additional effects on second-stage beliefs.

no significant differences between groups ($p = 0.623$). When combining these beliefs into extensive models, that is whether only X, only Z, both, or neither are believed to matter, we find that the vast majority (88%) believe either that both variables matter or that only Z does. Compared to *Z-first*, 9pp fewer *X-first* subjects believe that only Z matters, while 10pp more believe that both X and Z matter (see Table A.2). At the intensive margin, both groups predict a significantly higher success likelihood for projects with a high value of X. The X-first treatment group estimates X's impact on success to be 2.6pp higher ($p = 0.027$) than the Z-first group. In relative terms, the X-first group attributes a more than 50% higher marginal effect of X compared to the Z-first group. For Z, both groups infer a marginal effect of roughly 30pp, and we observe no statistically significant differences (see Column 4). When disaggregating the pooled beliefs in the marginal effects (Columns 5-8), we find that the treatment difference for X is driven particularly by heterogeneous assessments of its role when Z takes a high value ($Z = 1$): subjects exposed to X first update their success probabilities by 3.7pp more in this case ($p = 0.025$, Column 6). The estimate for Z when $X = 1$ (Column 8) is marginally significant ($p = 0.095$) in the opposite direction of our hypothesis, but this result is not robust when using the pooled belief in the marginal impact of Z, a measure which better controls for noise-driven differences in a single belief.[15]

We confirm that adding a second variable in the second stage increases the problem's complexity, as reflected in subjects' reported confidence in their beliefs (Enke and Graeber, 2023). As shown in Appendix Table A.3, average confidence in the accuracy of subjects' prediction models decreases from about 68% in the first stage to 60% in the second stage. Despite adopting different final models, the average confidence is very similar in the second stage for both treatment groups ($p = 0.439$).

To rule out that our results are driven by subjects being confused or possibly misunderstanding the experimental instructions, we replicate our findings when restricting our sample on the 84% of subjects who pass the comprehension questions in their first attempt. As reported in Appendix Table A.4, our results remain virtually unchanged when excluding subjects that initially make a mistake in the comprehension quiz.

The results confirm our pre-registered hypothesis: subjects' beliefs exhibit path dependence at both margins, with initial models persisting despite the need for fully revising a variable's marginal effect once observing the second predictor.

[15] In our pre-registration, our primary pre-registered hypothesis was formulated using the disaggregated beliefs in the marginal impact of both variables, as outlined in Columns 5 to 8. Since the average belief in the marginal impact provides a more direct and robust test of our main hypothesis, we consider it more appropriate for assessing it.

4.2.2 Spillovers on Decision-making

We now examine whether the path dependence observed in subjects' beliefs extends to their decision-making, which we pre-registered as a secondary analysis. In the second stage, subjects make four binary choices and corresponding willingness-to-pay (WTP) elicitations between pairs of projects that differ in a single variable value. Appendix Tables A.5 and A.6 report treatment differences in choices and WTP, respectively.

Subjects' binary choices are sensitive to project variables overall, with the likelihood of choosing the project with $X = 1$ or $Z = 1$ ranging from 50% to over 90% across pairs. On average, subjects respond more strongly to variation in Z than in X, with a 24pp difference in sensitivity that mirrors beliefs. Pooled treatment effects on choices are insignificant, though directionally consistent with beliefs. In the disaggregated specification, *X-first* subjects are 6.6pp more likely than *Z-first* subjects to choose the project with $X = 1$ when $Z = 1$ ($p = 0.032$), again mirroring the corresponding result for beliefs.

WTP exhibits the same qualitative pattern with no significant treatment differences. Z-variations are valued \$3.6 more than X-variations on average, but neither pooled nor disaggregated comparisons differ across treatments. The imperfect belief-WTP correlation ($\rho = 0.48$) and the larger scaled standard errors on WTP are consistent with behavioral attenuation between beliefs and actions (Enke et al., 2025; Yang, 2025), which leaves us underpowered to detect treatment differences in WTP.

4.2.3 Structural Evidence

Next, we use a structural approach to assess how closely subjects' second-stage beliefs align with empirical benchmarks. In line with our pre-registered goal of studying the dependence of second-stage beliefs on first-stage information, this method offers several advantages over a reduced-form treatment comparison. First, it allows us to test whether the observed differences between treatment groups stem from a stronger reliance on first-stage information. Second, it quantifies how closely subjects' models align with the rational benchmark.

Following Equation 3 from the framework, we estimate the weight second-stage beliefs place on the first-stage and second-stage empirical benchmarks, respectively. In the *Baseline* experiment, the second-stage benchmark coincides with the first-stage benchmark of the *Z-first* group. We therefore estimate the following regression model to identify differential responses to the two benchmarks:

$$\mu_{(x,z),i} = \beta_0 + \beta_1 \text{BMX}_{(x,z),i} + \beta_2 \text{BMZ}_{(x,z),i} + \beta_3(\text{BMX}_{(x,z),i} \times \textit{X-first}_i) + \beta_4(\text{BMZ}_{(x,z),i} \times \textit{X-first}_i) + \varepsilon_{(x,z),i}, \tag{5}$$

where $\mu_{(x,z),i}$ is subject i's belief (in %) about the success probability conditional on $(X, Z) = (x, z)$, and *X-first*$_i$ is an indicator for whether the subject has seen X in the first stage. Benchmark X, denoted by BMX$_{(x,z),i}$, and Benchmark Z, denoted by BMZ$_{(x,z),i}$, correspond to the first-stage benchmarks $P(Y \mid X)$ and $P(Y \mid Z)$, respectively. Benchmark Z additionally serves as the second-stage empirical benchmark. We demean the benchmarks using an uninformative baseline of 50% that corresponds to the unconditional probability $P(Y = Success)$ so that deviations capture the strength of subjects' updating toward each benchmark. As summarized in Table A.7, a rational subject should have a constant β_0 of 50%, fully incorporate Benchmark Z ($\beta_2 = 1$), and ignore Benchmark X ($\beta_1 = 0$). The interaction coefficients β_3 and β_4 indicate whether treatment groups load differentially on the benchmarks. Stickiness can then be observed as a greater reliance on information received in the first stage. This implies that β_3 is weakly greater than 0 and β_4 is weakly smaller than 0, with at least one of them being significantly different from 0.

Table 3 reports the structural estimates. Two main findings emerge. First, both treatment groups deviate from the rational benchmarks, overreacting to Benchmark X and underreacting to Benchmark Z. Second, the treatment difference stems primarily from the *X-first* group loading more heavily on Benchmark X; both groups load similarly on Benchmark Z.

More specifically, although X is not predictive of success when conditioning on Z, both groups react significantly to Benchmark X. The *Z-first* group responds at roughly $\hat{\beta}_1 = 16\%$; the *X-first* group responds an additional $\hat{\beta}_3 = 8.7$pp more ($p = 0.027$). Both groups load similarly on Benchmark Z ($\hat{\beta}_2 \approx 50\%$, $p = 0.230$ for the treatment interaction).

For stickiness to explain these treatment differences, subjects must have meaningfully updated their beliefs in the first stage. Appendix Table A.8 confirms this: first-stage beliefs strongly load on the respective first-stage empirical benchmark, validating that the two groups entered the second stage with meaningfully different models. When excluding subjects who update in the wrong direction in stage 1, that is failing to infer a higher success probability for the objectively better value, we find that stickiness becomes, if anything, slightly larger (Table 3, Column 2). The results are also robust to excluding subjects who failed the comprehension quiz on their first attempt (Column 3), and to a using a log-odds specification (Column 4).

Table 3 Path dependence: Structural approach

	Linear probability			Log odds
	Full sample	S1 directionally correct	CQ no mistake	Full sample
	(1)	(2)	(3)	(4)
Benchmark X	0.156***	0.175***	0.155***	0.156***
	(0.030)	(0.030)	(0.032)	(0.034)
Benchmark X	0.087**	0.124***	0.090**	0.103**
× X-first	(0.039)	(0.041)	(0.042)	(0.046)
Benchmark Z	0.500***	0.554***	0.532***	0.538***
	(0.021)	(0.021)	(0.022)	(0.023)
Benchmark Z	0.035	0.047	0.040	0.031
× X-first	(0.030)	(0.030)	(0.032)	(0.032)
Constant	47.348***	46.160***	46.992***	−0.164***
	(0.447)	(0.459)	(0.462)	(0.025)
Observations	3,136	2,668	2,628	3,093
R^2	0.363	0.441	0.406	0.342

Notes: This table analyzes the weights that second-stage beliefs place on empirical benchmarks using variations of Equation 5. The dependent variable is the reported belief about the success probability for a project with variables $(X = x, Z = z)$. For each subject, there are four observations, corresponding to the four combinations of predictor values. Column 1 reports results from estimating Equation 5. Column 2 restricts the sample to subjects who update in the correct direction in stage 1, that is infer a higher success probability for the objectively better value, and Column 3 to those who passed the comprehension quiz on their first attempt. Column 4 reports results using log-odds transformed variables. For this specification, we drop 43 observations with degenerate beliefs of 0% or 100%. Clustered standard errors are in parentheses. * denotes significance at 10 pct., ** at 5 pct., and *** at 1 pct. level.

Result 1 (Sticky Models): In the baseline DGP, subjects' second-stage models are sticky: they remain partially anchored to the first-stage empirical benchmark. Relative to subjects who observed the conditionally predictive variable Z first, those who observed the conditionally unpredictive X first

(a.) are more likely to perceive X as having a ceteris paribus effect on the outcome,

(b.) attribute a higher marginal effect to X.

4.3 Reasoning types

Individuals differ in the strategies they use to form beliefs from observed data, and these strategies may shape how they revise models when new information arrives. This raises the question whether stickiness is tied to flawed statistical reasoning or is instead a more general feature of model revision: subjects who rely on heuristics when forming beliefs might also fail to properly revise them, concentrating stickiness among those types.

To address this, we classify subjects' reasoning using open-text responses in which subjects described their approach to forming quantitative beliefs in the second stage.

Two independent coders labeled responses, with disagreements resolved by a third coder (see Appendix E for the full coding procedure and manual).[16]

We distinguish three main reasoning types: *Frequentist* reasoning, which yields the correct estimates of conditional success probabilities; *Separate* reasoning, which evaluates X and Z independently and thus neglects correlation; and *Absolute Success* reasoning, which focuses only on successes and thus neglects base rates.[17] Subjects whose reasoning cannot be classified based on their response, or does not fit one of these prominent categories, are labeled *Undetermined*.

Appendix Table A.9 shows that around two-thirds of subjects can be assigned a well-defined type: 34% Frequentist, 12% Separate, and 16% Absolute Success. The modal type is the normatively correct one, yet a substantial share rely on simpler strategies that neglect either correlation or base rates. Reasoning types are balanced across treatment groups ($\chi^2(3) = 3.495$, $p = 0.321$), confirming that treatment assignment did not shape which strategies subjects used.

Each type generates distinct predictions for benchmark loadings: Frequentists should load primarily on Benchmark Z, Separate reasoners more evenly on both benchmarks, and Absolute Success reasoners predominantly on the unconditional frequency of success. Analyzing second-stage beliefs by reasoning type confirms that each type corresponds to a distinct weighting of empirical benchmarks, closely aligning with the conceptual descriptions of the reasoning types and validating that the classification captures genuine differences in how subjects form beliefs (see Appendix Table A.10).

To test how stickiness relates to reasoning types, we estimate Equation 5 separately by reasoning type. Table 4 reports results for well-defined types, both pooled in Column 2 and separately in Columns 4–6, as well as for subjects who were assigned none of the three reasoning types in Column 3.

First, stickiness, if anything, becomes more pronounced when restricting the sample to subjects with a well-defined reasoning type, ruling out that it is driven by severe inattention or failure to engage with the task. Moreover, stickiness emerges qualitatively across all three reasoning types: the average difference in loading on Benchmark X between the *X-first* and *Z-first* group is about 8pp for *Frequentists*, 10pp for *Separate* reasoners, and 15pp for *Absolute Success* reasoners. By contrast, *Undetermined* sub-

[16]The two primary coders agree on 72% of the 784 responses; 18pp or 63% of the remaining disagreement is driven by one coder assigning a specific reasoning type while the other assigns the residual 'undetermined' category.

[17]Our text-based classification relates to the data-driven type elicitation in Fréchette et al. (2025), who use a finite mixture model to classify subjects into deterministic prediction rules. Their "Optimal" type corresponds to our *Frequentist* reasoners, while their "Ignores relevant variable" and "Conditions on irrelevant variable" types parallel aspects of our *Separate* and *Absolute Success* categories. They find that 96% of variance in prediction optimality is between subjects rather than across datasets, which is consistent with the interpretation that the reasoning types we capture are stable personal traits. In our data, reasoning types predict belief levels but are largely orthogonal to stickiness.

Table 4 Path dependence across reasoning types

	Grouped			Specific types		
	All	Well-defined type	Undetermined	Frequentist	Separate	Abs. success
	(1)	(2)	(3)	(4)	(5)	(6)
Benchmark X	0.156***	0.183***	0.117**	0.069**	0.398***	0.259***
	(0.030)	(0.033)	(0.056)	(0.033)	(0.110)	(0.064)
Benchmark X	0.087**	0.094**	0.068	0.077*	0.101	0.154*
× X first	(0.039)	(0.045)	(0.071)	(0.047)	(0.139)	(0.087)
Benchmark Z	0.500***	0.622***	0.322***	0.798***	0.426***	0.421***
	(0.021)	(0.025)	(0.032)	(0.027)	(0.063)	(0.038)
Benchmark Z	0.035	0.054	−0.024	0.032	0.097	−0.001
× X first	(0.030)	(0.033)	(0.047)	(0.036)	(0.078)	(0.055)
Constant	47.348***	45.899***	49.637***	47.143***	47.875***	41.740***
	(0.447)	(0.525)	(0.784)	(0.583)	(1.073)	(1.348)
Observations	3,136	1,920	1,216	1,060	368	492
R^2	0.363	0.523	0.156	0.716	0.405	0.295

Notes: This table reports results from estimating Equation 5 separately for different reasoning types. The dependent variable is the reported belief about the success probability for a project with variables $(X = x, Z = z)$. For each subject, there are four observations, corresponding to the four combinations of predictor values. Column 1 reports results for the full sample. Columns 2 and 3 report results separately for subjects with well-defined and undetermined types, respectively. Columns 4-6 report results for the three specific well-defined types: *Frequentist*, *Separate*, and *Absolute Success*. Clustered standard errors are in parentheses. * denotes significance at 10 pct., ** at 5 pct., and *** at 1 pct. level.

jects display qualitatively lower stickiness than any of the well-defined reasoning types, consistent with stickiness not being driven by subjects who failed to form a coherent strategy.[18]

Overall, these results suggest that stickiness emerges across different reasoning types, pointing to it being a general phenomenon rather than one tied to a specific approach of belief updating. These results are robust to AI-based coding via GPT-5.5 (see Appendix E.1).

5 Stickiness Across Data-Generating Processes

In the *Baseline* experiment, we showed that individuals form sticky models regardless of reasoning type. An open question is whether specific features of the baseline DGP are necessary for path dependence to arise. We test this in two follow-up experiments, *SymmCorr* and *SymmUncorr*, which progressively remove candidate features and thereby help isolate the underlying mechanisms of sticky models.

[18] Since splitting the sample by specific reasoning type with heterogeneous sample sizes greatly reduces statistical power to detect heterogeneous effects across types, these results should be interpreted cautiously.

First, in the baseline DGP the two predictors play asymmetric roles. They differ in their unconditional and conditional marginal effects, creating differences in the learning environment across groups. Second, the predictors are correlated, implying that marginal effects change across stages and making incorrect conditioning consequential for updating. The *SymmCorr* experiment shuts down the first feature, by preserving the correlation structure while making the predictors symmetric. The *SymmUncorr* experiment goes one step further by additionally removing the correlation between predictors and imposing constant marginal effects across stages.

We conducted both experiments on Prolific in March 2026, applying the same screening rules, comprehension checks, and incentive structure as in *Baseline*.[19] After excluding the 1% fastest and 1% slowest subjects, the final samples consist of $n = 601$ subjects in *SymmCorr* and $n = 600$ in *SymmUncorr*. Subjects spent an average of 23 minutes in both experiments and received $4 for completion plus an average bonus of $0.61. Demographics are broadly comparable to the baseline sample and balanced across the *X-first* and *Z-first* treatment groups within each experiment (see Appendix Table A.1).

Similar to the *Baseline* experiment, we find that a substantial share of subjects competently extracted the conditional success probabilities. The benchmark is again the modal response in every cell, and 52–57% of subjects fall within 5pp of it across the two first-stage beliefs, while 34–42% fall within 5pp across the four second-stage beliefs in the two experiments (see Appendix Figures B.3–B.6).

Since predictors are symmetric within each experiment, we pool treatment groups and test for path dependence by comparing perceived marginal effects depending on whether a predictor was observed in the first stage.[20] To further assess the extent to which second-stage beliefs remain anchored to the first-stage model, we regress second-stage beliefs on both the first-stage and second-stage benchmarks, which, unlike in *Baseline*, no longer coincide. Specifically, we estimate:

$$\mu_{(x,z),i} = \beta_0 + \beta_1 \text{BM-S1}_{(x,z),i} + \beta_2 \text{BM-S2}_{(x,z),i} + \varepsilon_{(x,z),i}, \tag{6}$$

where $\text{BM-S1}_{(x,z),i}$ is $P(Y \mid X = x)$ or $P(Y \mid Z = z)$ depending on which predictor was observed first, and $\text{BM-S2}_{(x,z),i} = P(Y \mid (X, Z) = (x, z))$ for both groups.

5.1 Stickiness Beyond Asymmetric Predictors

In the *Baseline* experiment, Z is the true predictor, while X is predictive only through its correlation with Z. Differences in the predictive strength of first-stage models or in

[19] We replaced the original attention check with a video-based attention screener.

[20] Results in a format parallel to Table 2 are reported in Appendix Table A.11.

the salience of the correlation between predictors and the outcome could therefore be necessary for differences in second-stage models to arise across groups.

The *SymmCorr* DGP isolates the role of asymmetry by giving the two predictors identical unconditional and conditional marginal effects while preserving their correlation. Both treatment groups now face statistically identical first stages and identical revisions in the second stage. The only remaining difference is which predictor they observed first. The relationship between variables is given by:

$$\begin{aligned} P(Y|X,Z) &= 0.1875 + 0.3125 \cdot X + 0.3125 \cdot Z \\ P(Y|V) &= 0.25 + 0.5 \cdot V, \quad V \in \{X, Z\} \\ P(Y) &= 0.5 \end{aligned}$$

In both treatment groups, subjects must revise downward the marginal effect of their first-stage predictor and learn that it equals the marginal effect of the other predictor.

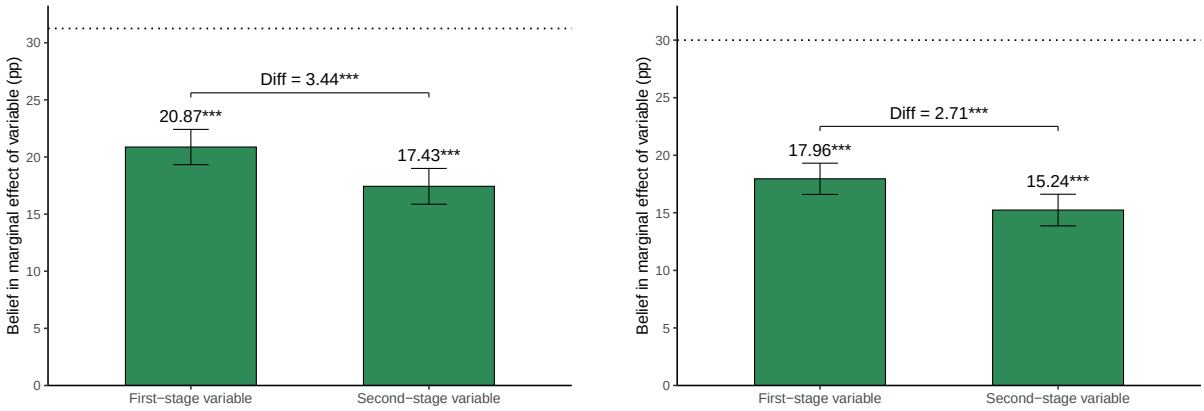
Panel A of Figure 3 shows the reduced-form results. Subjects' average belief in the marginal effect of the first-stage predictor is 3.4pp—or about 20%—higher than for the second-stage predictor ($p < 0.001$), a substantial deviation from the equal-weight benchmark implied by the symmetric DGP. We also find significant differences at the extensive margin. Subjects are 5pp more likely to agree that their first-stage predictor matters for success than to agree that their second-stage predictor does ($p = 0.004$; see Column 1 of Table A.12). These differences establish path dependence in model formation absent any asymmetry between predictors.

As a first validation, the treatment successfully induced distinct first-stage models, as subjects' first-stage beliefs load by 72% on the empirical benchmark (see Column 1 of Table A.13, Panel A). The structural specification further confirms that the treatment differences in second-stage models take the form of stickiness. Subjects' second-stage beliefs place a weight of about 56% on the second-stage empirical benchmark and still place a weight of about 7% on their first-stage benchmark (see Column 1 of Table A.14), indicating that second-stage models remain partially anchored to first-stage models. Asymmetry between predictors is therefore not a necessary condition for sticky models to arise.

Further supporting these findings, excluding subjects who failed to infer a higher success probability for the empirically better-performing value in the first stage, if anything, increases the treatment effects on second-stage models. The same pattern emerges when excluding subjects who initially made a mistake on the comprehension quiz (Appendix Table A.14).

Spillovers to decisions and valuations show the same qualitative pattern but are not statistically significant in this DGP (Panel A in Appendix Table A.15 and Table A.16,

Figure 3 Treatment effect on second-stage models — Symmetric DGPs



(a) SymmCorr DGP **(b)** SymmUncorr DGP

Notes: This figure plots treatment effects from an exogenous manipulation of the first-stage independent variable on subjects' mental models of project success in the second stage for the SymmCorr DGP (Panel A) and SymmUncorr DGP (Panel B). More specifically, it shows average beliefs about the marginal effect of predictors depending on whether the predictor was observed in the first stage ("First-stage variable") or observed only in the second stage ("Second-stage variable"). The empirical benchmark for the marginal effects is denoted by the dotted line. Corresponding regression results are reported in Table A.12. Significance levels displayed above each bar indicate whether the estimated belief about the marginal effect differs from zero. The bracketed difference reports the difference between beliefs about the marginal effects of first-stage and second-stage variables, together with the corresponding significance level. Error bars represent 95% confidence intervals based on clustered standard errors. *, **, and *** denote significance at the 10%, 5%, and 1% levels, respectively.

respectively), consistent with the noisier transmission from beliefs to decisions documented in *Baseline*.

5.2 Stickiness Beyond Correlated Predictors

Having ruled out asymmetry of predictors as a necessary condition, we next ask whether the correlation of predictors is necessary for sticky models to arise. In the DGPs for both *Baseline* and *SymmCorr*, predictors are correlated. Consequently, for at least one treatment group, the unconditional marginal effect of the variable subjects learn in the first stage differs from the conditional marginal effect of that same variable they should apply in the second stage.

This suggests a possible mechanism for stickiness in these environments: subjects may fail to 'unlearn' the marginal effect of the initially observed predictor, anchoring their second-stage beliefs about its effect on the first-stage benchmark. At the same time, subjects might accurately learn the marginal effect of the second-stage predictor for which no such anchor exists. The *SymmUncorr* DGP shuts down this 'failure to unlearn' mechanism by holding marginal effects of both predictors constant across stages.

The relationship between variables is given by:

$$\begin{aligned} P(Y|X,Z) &= 0.2 + 0.3 \cdot X + 0.3 \cdot Z \\ P(Y|V) &= 0.35 + 0.3 \cdot V, \quad V \in \{X, Z\} \\ P(Y) &= 0.5 \end{aligned}$$

Panel B of Figure 3 reports the reduced-form results. We again find significant differences in beliefs about the marginal effects of predictors. Subjects' average belief about the marginal effect of the first-stage predictor is 2.7pp—or about 18%—higher than that of the second-stage predictor ($p = 0.001$). At the extensive margin, subjects are 4.7pp more likely to agree that the first-stage predictor matters for success than to agree that the second-stage predictor does ($p = 0.005$; see Column 3 of Table A.12). These differences demonstrate path dependence in model formation in a setting without correlation between predictors.

Characterizing beliefs by the weight they place on empirical benchmarks, we again find substantial updating in the first stage, with subjects placing 76% weight on the benchmark (see Column 1 of Table A.13, Panel B). Subjects' second-stage beliefs then place around 51% weight on the second-stage empirical benchmark while still assigning about 9% weight to their first-stage benchmark (see Column 4 of Table A.14). This again confirms that the observed path dependence takes the form of stickiness.

These results are again robust to excluding subjects who failed to update in the correct direction in the first stage, as well as those who initially failed the comprehension quiz (Appendix Table A.14).

These belief differences also extend to decisions: subjects are significantly more likely to choose projects favored by their first-seen predictor, and report higher willingness-to-pay for them (Appendix Table A.15 and Table A.16, Panel B).

Stickiness therefore persists even when predictors are uncorrelated and marginal effects do not change across stages, ruling out failure-to-unlearn as a necessary condition. More broadly, reasoning biases tied to the distinction between unconditional and conditional marginal effects—such as correlation neglect (Enke and Zimmermann, 2019)—are inconsequential in *SymmUncorr*, yet stickiness remains.

Result 2 (Sticky Models Across DGPs): Sticky models, in the sense that subjects' final models remain partially anchored to the corresponding first-stage empirical benchmark, arise across data-generating processes in which the first-stage variable is unconditionally predictive. In particular, stickiness is robust to environments

(a.) in which predictors play symmetric roles (*SymmCorr*);

(b.) in which predictors are additionally uncorrelated and marginal effects remain con-

stant across stages (*SymmUncorr*).

Taken together, the two follow-up experiments rule out a broad class of explanations that trace stickiness back to features of the data-generating process, along with the reasoning biases tied to those features.

6 The Role of Cognitive Effort

In the previous sections, we demonstrated the generality of stickiness, supporting the framework outlined in Section 3. The results suggest that individuals enter the second stage with different cognitive defaults shaped by their first-stage models, independent of their approach to belief formation and the structure of the DGP.

What remains is the question under which conditions first-stage models continue to influence second-stage beliefs across these settings. The framework in Section 3 identifies cognitive effort as an important predictor. Subjects who invest less effort in revising their models in the second stage are expected to rely more heavily on first-stage beliefs as defaults. At the same time, first-stage beliefs are expected to place greater weight on the information received in Stage 1 among subjects who exert more effort in the first stage. We therefore hypothesize that stickiness, measured as reliance on first-stage benchmarks, is more pronounced among subjects who allocate relatively less effort to the second stage compared to the first stage.

We test this prediction across all three DGPs. Following a large literature, we use response-time data as a proxy for deliberation (e.g., Caplin et al. 2020; Rubinstein 2016; Wilcox 1993). We define *relative effort* as the share of total response time allocated to the second stage.[21] We then perform a median split based on relative effort and test the reliance of second-stage models on first-stage benchmarks separately above and below the median.

Summary statistics on time allocation are reported in Appendix Table A.17. In all three experiments, subjects spend on average 55–56% of the total time in the second stage. There are generally no significant differences in time allocation across treatment groups within our experiments, indicating that the treatment does not differentially affect effort allocation across stages.[22]

[21] For the *Baseline* experiment, we pre-registered a measure of cognitive effort based on whether subjects clicked a button to revisit the data table. Due to a technical issue, however, click data were not recorded on the pages where subjects reported their willingness to pay. We therefore rely on response-time data for the *Baseline* experiment as well, which are available for all pages. An analysis using click data on the belief-elicitation page only is reported in the Appendix (Table A.18).

[22] The one partial exception is that individuals in the *Baseline* experiment who start with the weaker predictor X spend, on average, 15 seconds more in the second stage ($p = 0.091$). Consistent with models of attention that trade off costs and benefits of information acquisition (e.g., Sims 2003; Caplin and Dean 2015; Gabaix 2014; Guarda et al. 2026), this may reflect stronger incentives for the *X-first* group to exert

6.1 Baseline DGP

To test the effect of relative effort on the reliance of second-stage models on first-stage benchmarks for the baseline DGP, we estimate equation 5 separately for individuals with relatively low and high shares of time spent in the second stage. Column 1 of Table 5 reports results for the full sample, Column 2 for individuals with a relatively high share of time spent in the second stage, and Column 3 for those with a relatively low share.

Table 5 Path dependence by cognitive effort

	All	High rel. S2 effort	Low rel. S2 effort
	(1)	(2)	(3)
Benchmark X	0.156***	0.163***	0.150***
	(0.030)	(0.043)	(0.042)
Benchmark X	0.087**	0.019	0.163***
× X-first	(0.039)	(0.052)	(0.059)
Benchmark Z	0.500***	0.626***	0.392***
	(0.021)	(0.030)	(0.027)
Benchmark Z	0.035	0.024	0.010
× X-first	(0.030)	(0.041)	(0.039)
Constant	47.348***	46.668***	48.028***
	(0.447)	(0.640)	(0.623)
Observations	3,136	1,568	1,568
R^2	0.363	0.476	0.260

Notes: This table reports results from estimating Equation 5 separately for subjects with above-median relative time spent in the second stage (Column 2) and subjects with below-median relative time spent in the second stage (Column 3). Column 1 reports results for the full sample. Relative time spent in the second stage is measured as the share of time spent in the second stage relative to the total time spent in stages 1 and 2 combined. The dependent variable is the reported belief about the success probability for a project with variables $(X = x, Z = z)$. For each subject, there are four observations, corresponding to the four combinations of predictor values. Clustered standard errors are in parentheses. * denotes significance at 10 pct., ** at 5 pct., and *** at 1 pct. level.

Two key findings emerge that confirm the prediction. First, model stickiness is driven primarily by individuals who allocate relatively less cognitive effort to the second stage. Among *X-first* subjects with lower relative effort, adherence to Benchmark X is 16.3pp higher than for their *Z-first* counterparts ($p = 0.006$), while adherence to Benchmark Z is similar across treatments. Correspondingly, *X-first* subjects report beliefs in the marginal effect of X that are 4.9pp higher than those of *Z-first* subjects, while there is no treatment difference in beliefs about the marginal effect of Z (see Appendix Table A.19). In contrast, among high-effort subjects there are no significant treatment differences in adherence to either benchmark and consequently no significant treatment differences in beliefs about the marginal effects of either predictor.

Second, subjects who spend relatively more time in stage two report beliefs that are

effort in the second stage, as they enter it with a weaker model.

closer to, but still fall short of, the rational benchmark. High-effort subjects follow Benchmark Z to about 63%–65%, roughly 23pp more than low-effort subjects ($p < 0.001$). In addition, reliance on the conditionally uninformative Benchmark X remains unchanged in the *Z-first* group and declines in the *X-first* group. Subjects also move closer to the rational benchmark at the extensive margin, according to which only Z has a ceteris paribus effect on the outcome. In particular, among *X-first* subjects, high-effort individuals are 4pp less likely to believe that only X has a ceteris paribus effect on the outcome, and 7pp more likely to believe that only Z does. In contrast, stickiness at the extensive margin does not similarly hurt *Z-first* subjects, as they start out with the true predictor. Even among low-effort subjects, 23% hold the rational model, and this share does not significantly increase with effort (see Appendix Table A.20).

Although effort allocation differs across reasoning types as introduced in Section 4.3, the central role of relative cognitive effort in predicting stickiness remains robust to controlling for reasoning type (see Appendix Table A.21 and Table A.22).

6.2 Symmetric DGPs

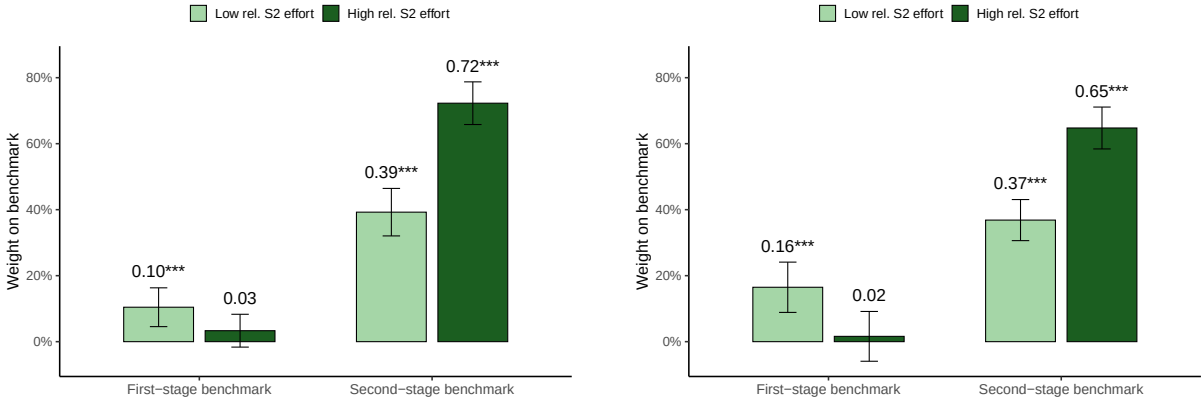
To test the effect of relative effort on the reliance of second-stage models on first-stage benchmarks for the symmetric DGPs, we estimate Equation 6 separately for individuals with relatively low and high shares of time spent in the second stage. Figure 4 displays the estimated weights placed on the first-stage and second-stage benchmarks, separately by relative effort, for the SymmCorr DGP (Panel A) and the SymmUncorr DGP (Panel B). Detailed regression results are reported in Table A.23.

The patterns mirror those for the baseline DGP, with the same two findings emerging in both symmetric settings. Only subjects who exert relatively low effort in the second stage exhibit stickiness. Among these individuals, second-stage models load on the second-stage benchmark by about 39% in the SymmCorr DGP and 37% in the SymmUncorr DGP, and on the first-stage benchmark by about 10% and 16%, respectively (all $p < 0.001$). This implies that individuals assign a marginal effect to the first-stage variable that exceeds that of the second-stage variable by around 5pp—or about 43% and 45%, respectively—despite the benchmark implying equal marginal effects (see Column 3 and 6 of Table A.24).

Effort again moves beliefs closer to the rational benchmark: high-effort subjects follow the second-stage benchmark to 72% and 65%, respectively, and do not meaningfully rely on the first-stage benchmark. These beliefs are therefore closer to the rational benchmark, both in being less attenuated overall and in assigning more similar weights to the two variables.

The same pattern emerges across all three experiments: stickiness is concentrated

Figure 4 Path dependence by cognitive effort — Symmetric DGPs



(a) SymmCorr DGP **(b)** SymmUncorr DGP

Notes: This figure plots the estimated weights that second-stage beliefs place on the first-stage benchmark and the second-stage benchmark separately for subjects with below-median relative time spent in the second stage ("Low rel. S2 effort") and above-median relative time spent in the second stage ("High rel. S2 effort"), for the SymmCorr DGP (Panel A) and SymmUncorr DGP (Panel B). Relative time spent in the second stage is measured as the share of time spent in the second stage relative to the total time spent in stages 1 and 2 combined. Detailed results from estimating Equation 6 for the respective subgroups are reported in Table A.23. Significance levels displayed above each bar indicate whether the estimated weight placed on the respective benchmark differs from zero. Error bars represent 95% confidence intervals based on clustered standard errors. *, **, and *** denote significance at the 10%, 5%, and 1% levels, respectively.

among subjects who allocate below-median relative effort to the second stage. Only these individuals significantly rely on the first-stage benchmark. Among those who invest relatively more effort, stickiness disappears and beliefs move significantly closer to the second-stage benchmark.

As a robustness check, we use a median split based on total time spent in stage 2 rather than relative time shares. We find that this measure also predicts stickiness across all three DGPs. Additionally, subjects with lower confidence in their second-stage beliefs exhibit stronger stickiness across DGPs (see Appendix Table A.25 and Table A.26).

Result 3 (Cognitive Effort Modulates Stickiness). Stickiness is modulated by the share of cognitive effort allocated to the second stage. Across all three experimental settings, the share of cognitive effort exerted in stage two affects the extent to which second-stage models rely on first-stage versus second-stage empirical benchmarks. In particular,

(a.) subjects who spend a lower share of time in stage two exhibit significantly greater stickiness;

(b.) subjects who spend a higher share of time in stage two form models that are significantly closer to the rational benchmark.

7 Conclusion

This paper addresses a central question in understanding how economic agents learn in dynamic environments: How do individuals adjust their mental models when they encounter new dimensions of information? Across three pre-registered experiments, we document that models are 'sticky,' as subjects fail to sufficiently revise underspecified models even when being provided with the complete data on all relevant variables.

Path dependence in model formation emerges robustly across distinct data-generating processes, which is consistent with individuals relying on previously learned models as defaults when updating becomes more cognitively demanding. Further supporting this interpretation, we find that stickiness is concentrated among subjects who allocate relatively little cognitive effort to the revision stage. It also arises across all well-defined reasoning types and is thus not tied to any specific inferential failure.

In the taxonomy of Handel and Schwartzstein (2018), our design isolates the *integration* of information from its acquisition since all relevant data is presented to subjects at the revision stage. The friction we identify—a generic cost of combining new dimensions of information with an existing model that is modulated by effort—therefore operates at this integration step, which informs possible policy interventions aiming to debias stickiness.

Our findings highlight stickiness as an important source of heterogeneity in mental models. Even when individuals observe the same data, prior learning shapes how new information is incorporated. Subjects often recognize newly relevant predictors while continuing to rely excessively on variables that were emphasized in earlier model learning stages, even when those variables become conditionally irrelevant. This provides a natural explanation for persistent misconceptions observed in various economic settings.

Five design factors suggest these estimates are a lower bound on stickiness outside the lab: (i) we ensure minimum attention to all potentially relevant variables, (ii) we provide all the data to infer the relationships throughout the revision stage, (iii) we limit the role of preference-based model revision by employing minimal framing, (iv) we are transparent about the existence of the second variable from the beginning and (v) we use elicitations that already guide people to think in the correct contingencies when revising their models. In naturally-occurring settings, where these scaffolds are absent and other biases plausibly compound the cognitive friction, we expect stickiness to be more severe.

Several open questions point towards promising avenues for further research. First, many cases of dynamic model formation involve new variables for which very few observations initially exist. It would thus be interesting to explore how individuals handle

the trade-off between adopting new variables in a model that better fits the data and the uncertainty arising from limited data. Furthermore, many environments are often more complex, involving a larger number of predictors. This raises the question of how the number of predictors affects the adoption of new variables and the revision of existing models. Greater complexity may intensify reliance on prior models and thereby reinforce stickiness, but it may also eventually induce more substantial model revision or even paradigm shifts. More generally, future work could investigate how stickiness evolves over time: whether individuals gradually converge toward correctly specified models as more data becomes available, or whether early models continue to shape learning and thereby reinforce path dependence.

References

Aina, Chiara and Florian Schneider, "Weighting Competing Models," 2025. CESifo Working Paper No. 11757.

Ambuehl, Sandro and Heidi C. Thysen, "Choosing Between Causal Interpretations: An Experimental Study," 2025. CESifo Working Paper No. 11103.

Augenblick, Ned, Matthew Backus, Andrew T. Little, and Don A. Moore, "Assumptions, Disagreement, and Overprecision: Theory and Evidence," 2025. Working Paper.

Ba, Cuimin, "Robust Misspecified Models," *American Economic Review*, 2026, *116* (4), 1340–1379.

Barron, Kai and Tilman Fries, "Narrative Persuasion," 2025. Working paper.

Benjamin, Daniel J., "Errors in Probabilistic Reasoning and Judgment Biases," in Douglas Bernheim, Stefano DellaVigna, and David Laibson, eds., *Handbook of Behavioral Economics*, North Holland, Amsterdam, 2019, pp. 69–186.

Bohren, J. Aislinn, Peter Hull, and Alex Imas, "Systemic Discrimination: Theory and Measurement," *The Quarterly Journal of Economics*, 2025, *140* (3), 1743–1799.

Bordalo, Pedro, John J Conlon, Nicola Gennaioli, Spencer Y Kwon, and Andrei Shleifer, "Memory and Probability*," *The Quarterly Journal of Economics*, January 2023, *138* (1), 265–311.

_ , John J. Conlon, Nicola Gennaioli, Spencer Y. Kwon, and Andrei Shleifer, "How People Use Statistics," *The Review of Economic Studies*, 2026, *93* (1), 250–285.

Caplin, Andrew and Mark Dean, "Revealed Preference, Rational Inattention, and Costly Information Acquisition," *American Economic Review*, 2015, *105* (7), 2183–2203.

_ , Dániel Csaba, John Leahy, and Oded Nov, "Rational Inattention, Competitive Supply, and Psychometrics*," *The Quarterly Journal of Economics*, August 2020, *135* (3), 1681–1724.

Charles, Constantin and Chad Kendall, "Causal Narratives," 2024. NBER Working Paper No. 30346.

Dean, Mark, Benjamin Enke, Thomas Graeber, and Pietro Ortoleva, "Reference Points as Information," *Working Paper*, 2026.

Eliaz, Kfir and Ran Spiegler, "A Model of Competing Narratives," *American Economic Review*, 2020, *110* (12), 3786–3816.

Enke, Benjamin, "What You See Is All There Is," *The Quarterly Journal of Economics*, 2020, *135* (3), 1363–1398.

_ , "The Cognitive Turn in Behavioral Economics," 2026. Working Paper.

_ **and Florian Zimmermann**, "Correlation Neglect in Belief Formation," *The Review of Economic Studies*, 2019, *86* (1), 313–332.

_ **and Thomas Graeber**, "Cognitive Uncertainty," *The Quarterly Journal of Economics*, 2023, *138* (4), 2021–2067.

_ , _ , **Ryan Oprea, and Jeffrey Yang**, "Behavioral Attenuation," 2025. NBER Working Paper No. 32973.

Esponda, Ignacio, Emanuel Vespa, and Sevgi Yuksel, "Mental Models and Learning: The Case of Base-Rate Neglect," *American Economic Review*, 2024, *114* (3), 752–782.

Fan, Tony Q. and Tilman Fries, "Narratives, Belief Movements, and Economic Fluctuations," 2026. Working Paper.

Fréchette, Guillaume, Emanuel Vespa, and Sevgi Yuksel, "Extracting Models From Data Sets: An Experiment," 2025. Working Paper.

Fudenberg, Drew and Giacomo Lanzani, "Which misspecifications persist?," *Theoretical Economics*, 2023, *18* (3), 1271–1315. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.3982/TE5298.

Gabaix, Xavier, "A Sparsity-Based Model of Bounded Rationality," *Quarterly Journal of Economics*, 2014, *129*, 1661–1710.

_ , "Chapter 4 - Behavioral inattention," in B. Douglas Bernheim, Stefano DellaVigna, and David Laibson, eds., *Handbook of Behavioral Economics: Applications and Foundations 1*, Vol. 2 of *Handbook of Behavioral Economics - Foundations and Applications 2*, North-Holland, January 2019, pp. 261–343.

Gagnon-Bartsch, Tristan, Matthew Rabin, and Joshua Schwartzstein, "Channeled Attention and Stable Errors," 2023. Working Paper.

Gennaioli, Nicola and Andrei Shleifer, "What Comes to Mind," *The Quarterly Journal of Economics*, 2010, *125* (4), 1399–1433.

Graeber, Thomas and Benjamin Enke, "Comparisons," *Working Paper*, 2026.

Guarda, Sebastián, José Luis Montiel Olea, and Pietro Ortoleva, "Endogenous Misspecification," 2026. Working Paper.

Han, Yi, David Huffman, and Yiming Liu, "Minds, Models and Markets: How Managerial Cognition Affects Pricing Strategies," 2026. Working Paper.

Handel, Benjamin and Joshua Schwartzstein, "Frictions or Mental Gaps: What's Behind the Information We (Don't) Use and When Do We Care?," *Journal of Economic Perspectives*, February 2018, *32* (1), 155–178.

Hanna, Rema, Sendhil Mullainathan, and Joshua Schwartzstein, "Learning Through Noticing: Theory and Evidence from a Field Experiment *," *The Quarterly Journal of Economics*, 2014, *129* (3), 1311–1353.

Hong, Harrison, Jeremy C. Stein, and Jialin Yu, "Simple Forecasts and Paradigm Shifts," *The Journal of Finance*, 2007, *62* (3), 1207–1242. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1540-6261.2007.01234.x.

Ilut, Cosmin and Rosen Valchev, "Economic Agents as Imperfect Problem Solvers," *The Quarterly Journal of Economics*, 2022, *138* (1), 313–362.

Kahneman, Daniel and Shane Frederick, "Representativeness Revisited: Attribute Substitution in Intuitive Judgment," in Thomas Gilovich, Dale Griffin, and Kahneman, eds., *Heuristics and Biases: The Psychology of Intuitive Judgment*, Cambridge University Press, 2002, p. 49–81.

Kendall, Chad and Ryan Oprea, "On the complexity of forming mental models," *Quantitative Economics*, 2024, *15* (1), 175–211.

Khaw, Mel Win, Ziang Li, and Michael Woodford, "Cognitive Imprecision and Small-Stakes Risk Aversion," *The Review of Economic Studies*, 08 2020, *88* (4), 1979–2013.

Kojima, Takeshi, Shixiang Shane Gu, Machel Reid, Yutaka Matsuo, and Yusuke Iwasawa, "Large Language Models are Zero-Shot Reasoners," in "Advances in Neural Information Processing Systems," Vol. 35 2022.

Lanzani, Giacomo, "Dynamic Concern for Misspecification," *Econometrica*, 2025, *93* (4), 1333–1370.

Liu, Manwei and Sili Zhang, "Counteracting Narratives: Evidence from an Online Experiment," *The Economic Journal*, January 2026, *136* (673), 125–162.

Macchi, Elisa, "Worth Your Weight: Experimental Evidence on the Benefits of Obesity in Low-Income Countries," *American Economic Review*, 2023, *113* (9), 2287–2322.

Maćkowiak, Bartosz and Mirko Wiederholt, "Optimal Sticky Prices under Rational Inattention," *American Economic Review*, 2009, *99* (3), 769–803.

Martínez-Marquina, Alejandro, Muriel Niederle, and Emanuel Vespa, "Failures in Contingent Reasoning: The Role of Uncertainty," *American Economic Review*, 2019, *109* (10), 3437–3474.

Musolff, Robin and Florian Zimmermann, "Model Uncertainty," 2025. CESifo Working Paper No. 12041.

Niederle, Muriel and Emanuel Vespa, "Cognitive Limitations: Failures of Contingent Thinking," *Annual Review of Economics*, 2023, *15* (1), 307–328. _eprint: https://doi.org/10.1146/annurev-economics-091622-124733.

Ortoleva, Pietro, "Modeling the Change of Paradigm: Non-Bayesian Reactions to Unexpected News," *American Economic Review*, 2012, *102* (6), 2410–2436.

Pager, Devah, "The Mark of a Criminal Record," *American Journal of Sociology*, 2003, *108* (5), 937–975. Publisher: The University of Chicago Press.

Peer, Eyal, David Rothschild, Andrew Gordon, Zak Evernden, and Ekaterina Damer, "Data quality of platforms and panels for online behavioral research," *Behavior Research Methods*, August 2022, *54* (4), 1643–1662.

Rabin, Matthew and Joel L. Schrag, "First Impressions Matter: A Model of Confirmatory Bias," *The Quarterly Journal of Economics*, 1999, *114* (1), 37–82.

Rubinstein, Ariel, "A Typology of Players: Between Instinctive and Contemplative," *The Quarterly Journal of Economics*, May 2016, *131* (2), 859–890.

Schwartzstein, Joshua, "Selective Attention and Learning," *Journal of the European Economic Association*, 2014, *12* (6), 1423–1452.

_ **and Adi Sunderam**, "Using Models to Persuade," *American Economic Review*, 2021, *111* (1), 276–323.

Sims, Christopher, "Implications of Rational Inattention," *Journal of Monetary Economics*, 2003, *50*, 665–690.

Snowberg, Erik and Leeat Yariv, "Testing the Waters: Behavior across Participant Pools," *American Economic Review*, 2021, *111* (2), 687–719.

Tversky, Amos and Daniel Kahneman, "Judgment under Uncertainty: Heuristics and Biases: Biases in judgments reveal some heuristics of thinking under uncertainty.," *science*, 1974, *185* (4157), 1124–1131.

Wilcox, Nathaniel T., "Lottery Choice: Incentives, Complexity and Decision Time," *The Economic Journal*, November 1993, *103* (421), 1397–1417.

Yang, Jeffrey, "On the Decision-Relevance of Subjective Beliefs," 2025. Working Paper.

For Online Publication Only:

Appendix
Sticky Models

Paul Grass, Philipp Schirmer, Malin Siemers

Summary of the Online Appendix

Section A provides additional tables.

Section B provides additional figures.

Section C contains details on our preregistration.

Section D discusses extensions to the theoretical framework.

Section E provides the coding of the open-ended responses and tables using AI codes.

Section F includes the experimental instructions.

A Additional Tables

Table A.1 Summary statistics and balancing — All DGPs

		Baseline			*SymmCorr*			*SymmUncorr*		
Variable	**ACS** (2022)	**X-first**	**Z-first**	**p-value**	**X-first**	**Z-first**	**p-value**	**X-first**	**Z-first**	**p-value**
Gender										
Female	50%	49%	51%	0.617	49%	48%	0.837	51%	47%	0.371
Age										
18–34	29%	42%	44%	0.629	44%	44%	0.963	43%	48%	0.246
35–54	32%	45%	42%	0.378	45%	41%	0.391	43%	38%	0.225
55+	38%	13%	14%	0.565	11%	15%	0.232	14%	14%	0.960
Household net income										
Below $50k	34%	35%	32%	0.380	32%	29%	0.410	30%	35%	0.170
$50k–$100k	29%	38%	43%	0.116	36%	42%	0.120	41%	34%	0.058
Above $100k	37%	27%	25%	0.417	32%	29%	0.410	29%	31%	0.547
Education										
Bachelor's degree or more	33%	58%	60%	0.593	55%	55%	0.906	59%	51%	0.069
Region										
Northeast	17%	25%	21%	0.126	19%	21%	0.529	24%	22%	0.664
Midwest	21%	20%	21%	0.803	22%	24%	0.493	18%	21%	0.275
South	39%	36%	37%	0.855	40%	38%	0.474	42%	37%	0.216
West	24%	18%	21%	0.287	19%	17%	0.615	17%	20%	0.361
F-Stat (SUR)				0.658			0.981			0.206
Sample size	1,980,550	394	390	784	300	301	601	298	302	600

Notes: This table presents summary statistics for the demographics of the *Baseline*, *SymmCorr* and *SymmUncorr* experiments. It compares them to benchmark characteristics for the US adult population based on data from the American Community Survey 2022. Column 4, 7, and 10 report the p-value of t-tests, testing for equality between both treatment groups for a given experiment, as well as the statistic of a 'seemingly unrelated regressions' (SUR) F-test.

Table A.2 Treatment effect on extensive models

	Type of model (extensive margin)			
	Only X matters	Only Z matters	X and Z matter	Both don't matter
	(1)	(2)	(3)	(4)
X-first	−0.001	−0.089***	0.100***	−0.011
	(0.016)	(0.028)	(0.033)	(0.017)
Constant	0.056***	0.244***	0.631***	0.069***
	(0.012)	(0.022)	(0.024)	(0.013)
Observations	784	784	784	784
R^2	0.000	0.012	0.012	0.000

Notes: This table presents treatment effects from an exogenous manipulation of the first-stage independent variable on subjects' mental models at the extensive margin in the second stage. *X-first* is an indicator equal to 1 if the subject observed X in the first stage. The table reports the share of subjects holding each of the four possible extensive models: only X matters (Column 1), only Z matters (Column 2), both variables matter (Column 3), or neither variable matters (Column 4) for project success. A variable is considered to matter if subjects agree that the predictors X and Z, respectively, have a *ceteris paribus* effect on the outcome. Robust standard errors are in parentheses. * denotes significance at 10 pct., ** at 5 pct., and *** at 1 pct. level.

Table A.3 Summary statistics on confidence across stages — All DGPs

Panel A: Baseline DGP

Variable	All	X-first	Z-first	p-value (KS)
Confidence S1	68%	67%	69%	0.261
Confidence S2	60%	61%	60%	0.439
Above median confidence S1	56%	55%	58%	
Above median confidence S2	51%	52%	50%	
Sample size	784	394	390	

Panel B: *SymmCorr*

Variable	All	X-first	Z-first	p-value (KS)
Confidence S1	71%	71%	71%	0.997
Confidence S2	64%	64%	64%	0.669
Above median confidence S1	52%	51%	53%	
Above median confidence S2	51%	51%	51%	
Sample size	601	300	301	

Panel C: *SymmUncorr*

Variable	All	X-first	Z-first	p-value (KS)
Confidence S1	68%	68%	67%	0.681
Confidence S2	60%	61%	60%	0.929
Above median confidence S1	51%	54%	48%	
Above median confidence S2	56%	57%	55%	
Sample size	600	298	302	

Notes: This table reports descriptive statistics on subjects' self-reported confidence in their stated beliefs in the first (Row 1) and second stage (Row 2) of the experiment by assigned treatment group in the *Baseline* (Panel A), *SymmCorr* (Panel B), and *SymmUncorr* (Panel C) experiments. Confidence is measured as the perceived likelihood that subjects' assessments are within ± 5 percentage points of the correct success probabilities. Column 1 reports statistics for the full sample, while Columns 2 and 3 report statistics separately for subjects assigned to the *X-first* and *Z-first* treatment groups, respectively. Row 1 and 2 report mean values. Row 3 reports the share of subjects with above-median confidence in the first stage. Row 4 reports the share of subjects with above-median confidence in the second stage. Column 4 reports the p-value of a Kolmogorov-Smirnov-test for the equality of distribution across treatment groups.

Table A.4 Treatment effect on second-stage models: Only subjects without comprehension quiz mistakes

	Extensive margin Share agreeing (in %)		**Intensive margin** $\Delta P(Y = \text{Success} \mid \cdots)$		**Intensive margin - disaggregated** $\Delta P(Y = \text{Success} \mid \cdots)$			
	X matters	Z matters	ΔX	ΔZ	ΔX, Z = 0	ΔX, Z = 1	ΔZ, X = 0	ΔZ, X = 1
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
X-first	10.501***	2.182	2.714**	2.427	1.429	4.000**	1.141	3.712*
	(3.420)	(2.420)	(1.272)	(1.909)	(1.591)	(1.778)	(2.363)	(2.040)
Constant	68.249***	88.131***	4.653***	31.941***	−0.629	9.935***	26.659***	37.223***
	(2.540)	(1.765)	(0.969)	(1.320)	(1.229)	(1.274)	(1.655)	(1.414)
Observations	657	657	1,314	1,314	657	657	657	657
R^2	0.014	0.001	0.004	0.002	0.001	0.008	0.000	0.005

Notes: This table reproduces Table 2, restricting the sample to those subjects that make no mistakes in the comprehension quiz, passing in their first attempt. It presents treatment effects from an exogenous manipulation of the first-stage independent variable on subjects' mental models of project success in the second stage. *X-first* is an indicator equal to 1 if the subject observed X in the first stage. Columns 1 and 2 report the share of subjects agreeing that the predictors X and Z, respectively, have a *ceteris paribus* effect on the outcome. Columns 3 and 4 report beliefs about the marginal effects of X and Z, respectively, pooling across the two possible values of the other variable. For each subject, there are two observations, corresponding to the two values of the other variable. Columns 5–8 report beliefs about the marginal effects of X and Z separately for each value of the other predictor. Clustered standard errors (Columns 1–4) and robust standard errors (Columns 5–8) are in parentheses. * denotes significance at 10 pct., ** at 5 pct., and *** at 1 pct. level.

Table A.5 Treatment effect on project choice

	Share choosing project with higher features (in percent)					
	Pooled		**Disaggregated**			
	ΔX	ΔZ	ΔX, Z = 0	ΔX, Z = 1	ΔZ, X = 0	ΔZ, X = 1
	(1)	(2)	(3)	(4)	(5)	(6)
X-first	2.420	1.162	−1.792	6.632**	1.213	1.112
	(2.461)	(2.000)	(3.575)	(3.084)	(2.799)	(2.040)
Constant	61.667***	85.513***	51.538***	71.795***	80.513***	90.513***
	(1.758)	(1.453)	(2.534)	(2.282)	(2.008)	(1.486)
Observations	1,568	1,568	784	784	784	784
R^2	0.001	0.000	0.000	0.006	0.000	0.000

Notes: This table presents treatment effects from an exogenous manipulation of the first-stage independent variable on project choice in the second stage. The dependent variable is an indicator equal to 1 if the subject chose the project with the higher value of a predictor when comparing two projects that differ only in that predictor. *X-first* is an indicator equal to 1 if the subject observed X in the first stage. Columns 1 and 2 report the share of subjects choosing the project with the higher value of X and Z, respectively, pooling across the two values of the other variable. For each subject, there are two observations, corresponding to the two values of the other variable. Columns 3–6 report choices separately for each value of the other predictor. Clustered standard errors (Columns 1–2) and robust standard errors (Columns 3–6) are in parentheses. * denotes significance at 10 pct., ** at 5 pct., and *** at 1 pct. level.

Table A.6 Treatment effect on project valuation

	Willingness-to-pay (WTP) for project with higher features (in USD)					
	Pooled		**Disaggregated**			
	ΔX	ΔZ	ΔX, Z = 0	ΔX, Z = 1	ΔZ, X = 0	ΔZ, X = 1
	(1)	(2)	(3)	(4)	(5)	(6)
X-first	0.146	0.142	−0.322	0.614	0.045	0.238
	(0.311)	(0.304)	(0.435)	(0.414)	(0.394)	(0.324)
Constant	1.611***	5.206***	−0.018	3.239***	4.348***	6.065***
	(0.224)	(0.219)	(0.308)	(0.298)	(0.286)	(0.230)
Observations	1,568	1,568	784	784	784	784
R^2	0.000	0.000	0.001	0.003	0.000	0.001

Notes: This table presents treatment effects from an exogenous manipulation of the first-stage independent variable on willingness to pay (WTP) in the second stage. The dependent variable is the amount subjects are willing to pay for a project with a higher value of a predictor, holding the value of the other predictor constant. *X-first* is an indicator equal to 1 if the subject observed X in the first stage. Columns 1 and 2 report WTP for projects with a higher value of X and Z, respectively, pooling across the two values of the other variable. For each subject, there are two observations, corresponding to the two values of the other variable. Columns 3–6 report WTP separately for each value of the other predictor. Clustered standard errors (Columns 1–2) and robust standard errors (Columns 3–6) are in parentheses. * denotes significance at 10 pct., ** at 5 pct., and *** at 1 pct. level.

Table A.7 Rational benchmark for coefficient estimates in the second stage

Coefficient	**Rational Benchmark Value**	**Interpretation**
β_0	50	Intercept; baseline belief level under no information
β_1	0	No incorporation of Benchmark X
β_2	1	Full incorporation of Benchmark Z
β_3	0	No differential effect on X incorporation between treatment groups
β_4	0	No differential effect on Z incorporation between treatment groups

Notes: This table reports which estimates from Equation 5 are consistent with beliefs equal to the rational benchmark. The rational benchmark assumes a frequentist approach that perfectly incorporates all relevant information. The beliefs of a Bayesian learner with full-support prior will be very close to the empirical benchmark given by the above table since the full dataset contains a total of 40 rows $*1,000$ identical observations per row $= 40,000$ observations.

Table A.8 First-stage models

	First stage subjective success probability (pooled)	
	X first	Z first
	(1)	(2)
Benchmark X	0.809***	
	(0.043)	
Benchmark Z		0.717***
		(0.025)
Constant	51.852***	50.983***
	(0.468)	(0.425)
Observations	788	780
R^2	0.374	0.616

Notes: This table analyzes model formation in the first stage of the experiment by estimating the extent to which first-stage beliefs rely on the first-stage empirical benchmark. This corresponds to $P_{emp}(Y = 1|X = x)$ for *X-first* subjects and $P_{emp}(Y = 1|Z = z)$ for *Z-first* subjects. Benchmarks are demeaned by their average across possible predictor values. The dependent variable is the reported belief about the success probability for a project with variable $(X = x)$ or $(Z = z)$, respectively. For each subject, there are two observations, corresponding to the two predictor values. Column 1 reports results for the *X-first* group, and Column 2 reports results for the *Z-first* group. Clustered standard errors are in parentheses. * denotes significance at 10 pct., ** at 5 pct., and *** at 1 pct. level.

Table A.9 Distribution of reasoning types by treatment group

Reasoning type	**All**	**X-first**	**Z-first**	**p-value** H_0: *X-first = Z-first*
Frequentist	265 (34%)	142 (36%)	123 (32%)	0.183
Separate	92 (12%)	50 (13%)	42 (11%)	0.404
Absolute Success	123 (16%)	56 (14%)	67 (17%)	0.254
Undetermined	304 (39%)	146 (37%)	158 (41%)	0.321
Sample size	784	394	390	

Notes: This table reports the distribution of reasoning types for the full sample (Column 1) and by assigned treatment group, *X-first* (Column 2) and *Z-first* (Column 3). All columns report the number of subjects, with percentages in parentheses, classified into each reasoning type based on their description of the belief formation process in the second stage. Column 4 reports the p-value from a Kolmogorov Smirnov test of equality of distributions across treatment groups.

Table A.10 Second-stage models across reasoning types

	Subjective success probability (pooled), by reasoning				
	All	Frequentist	Separate	Abs. success	Undetermined
	(1)	(2)	(3)	(4)	(5)
Benchmark X	0.082***	0.041	0.339***	0.114**	0.028
	(0.021)	(0.026)	(0.075)	(0.050)	(0.039)
Benchmark Z	0.401***	0.746***	0.366***	0.206***	0.189***
	(0.019)	(0.027)	(0.050)	(0.040)	(0.030)
BM number of successes	0.235***	0.138***	0.227***	0.429***	0.243***
	(0.021)	(0.030)	(0.061)	(0.056)	(0.038)
Constant	47.348***	47.143***	47.875***	41.740***	49.637***
	(0.447)	(0.583)	(1.071)	(1.347)	(0.784)
Observations	3,136	1,060	368	492	1,216
R^2	0.374	0.719	0.413	0.339	0.171

Notes: This table analyzes second-stage models by reasoning type, by estimating the extent to which second-stage models rely on the empirical benchmarks of the first-stage conditional success probabilities *Benchmark X* ($P_{emp}(Y = 1|X = x)$) and *Benchmark Z* ($P_{emp}(Y = 1|Z = z)$), as well as the empirical benchmark for the (rescaled) number of successes *BM number of successes* ($P_{emp}(X = x, Z = z|Y = 1)$). All benchmarks are demeaned by their average across possible predictor values. The dependent variable is the reported belief about the success probability for a project with variables $(X = x, Z = z)$. For each subject, there are four observations, corresponding to the four combinations of predictor values. Column 1 reports results for the full sample. Column 2–4 report results for the three specific well-defined types: *Frequentist*, *Separate*, and *Absolute Success*. Column 5 reports results for subjects with an undetermined type, comprising all responses that cannot be assigned to any of the other three categories. Clustered standard errors are in parentheses. * denotes significance at 10 pct., ** at 5 pct., and *** at 1 pct. level.

Table A.11 Treatment effect on second-stage models (detailed) — Symmetric DGPs

Panel A: *SymmCorr*

	Extensive margin Share agreeing (in %)		**Intensive margin** $\Delta P(Y = \text{Success} \mid \cdots)$		**Intensive margin - disaggregated** $\Delta P(Y = \text{Success} \mid \cdots)$			
	X matters	Z matters	ΔX	ΔZ	ΔX, Z = 0	ΔX, Z = 1	ΔZ, X = 0	ΔZ, X = 1
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
X-first	6.259*	−3.719	3.362**	−3.507**	2.335	4.388**	−4.533**	−2.481
	(3.204)	(3.100)	(1.560)	(1.602)	(1.845)	(2.156)	(2.000)	(2.078)
Constant	77.741***	84.385***	15.988***	22.390***	11.615***	20.362***	18.017***	26.764***
	(2.402)	(2.096)	(1.178)	(1.191)	(1.349)	(1.550)	(1.436)	(1.490)
Observations	601	601	1,202	1,202	601	601	601	601
R^2	0.006	0.002	0.004	0.005	0.003	0.007	0.009	0.002

Panel B: *SymmUncorr*

	Extensive margin Share agreeing (in %)		**Intensive margin** $\Delta P(Y = \text{Success} \mid \cdots)$		**Intensive margin - disaggregated** $\Delta P(Y = \text{Success} \mid \cdots)$			
	X matters	Z matters	ΔX	ΔZ	ΔX, Z = 0	ΔX, Z = 1	ΔZ, X = 0	ΔZ, X = 1
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
X-first	1.765	−7.547**	4.099***	−1.313	2.898*	5.300***	−2.514	−0.112
	(3.118)	(2.986)	(1.437)	(1.341)	(1.701)	(1.883)	(1.733)	(1.704)
Constant	81.457***	87.748***	13.856***	17.955***	13.632***	14.079***	17.732***	18.179***
	(2.240)	(1.890)	(1.040)	(0.966)	(1.197)	(1.339)	(1.216)	(1.203)
Observations	600	600	1,200	1,200	600	600	600	600
R^2	0.001	0.011	0.009	0.001	0.005	0.013	0.004	0.000

Notes: This table presents treatment effects from an exogenous manipulation of the first-stage independent variable on subjects' mental models of project success in the second stage for the *SymmCorr* DGP (Panel A) and *SymmUncorr* DGP (Panel B). *X-first* is an indicator equal to 1 if the subject observed X in the first stage. Columns 1 and 2 report the share of subjects agreeing that the predictors X and Z, respectively, have a *ceteris paribus* effect on the outcome. Columns 3 and 4 report beliefs about the marginal effects of X and Z, respectively, pooling across the two possible values of the other variable. For each subject, there are two observations, corresponding to the two values of the other variable. Columns 5–8 report beliefs about the marginal effects of X and Z separately for each value of the other predictor. Clustered standard errors (Columns 1–4) and robust standard errors (Columns 5–8) are in parentheses. * denotes significance at 10 pct., ** at 5 pct., and *** at 1 pct. level.

Table A.12 Treatment effect on second-stage models — Symmetric DGPs

	Symmetric-Correlated DGP		Symmetric-Uncorrelated DGP	
	Variable matters (%)	Δ Variable (pp)	Variable matters (%)	Δ Variable (pp)
	(1)	(2)	(3)	(4)
First-stage predictor	4.992***	3.439***	4.667***	2.715***
	(1.703)	(0.983)	(1.641)	(0.824)
Constant	79.201***	17.433***	80.833***	15.240***
	(1.658)	(0.798)	(1.609)	(0.701)
Observations	1,202	2,404	1,200	2,400
R^2	0.004	0.005	0.004	0.004

Notes: This table presents treatment effects from an exogenous manipulation of the first-stage independent variable on subjects' mental models of project success in the second stage for the *SymmCorr* DGP (Column 1 and 2) and *SymmUncorr* DGP (Column 3 and 4). "First-stage predictor" is an indicator equal to one if the predictor was observed in the first stage. The dependent variable in Columns 1 and 3 is whether the subject agrees that the variable has a ceteris paribus effect on the outcome. For each subject, there are two observations, corresponding to the two predictors. The dependent variable in Columns 2 and 4 is the belief about the marginal effect of the variable. For each subject, there are four observations, corresponding to beliefs about the marginal effects of the two predictors at both values of the other predictor. Clustered standard errors are in parentheses. * denotes significance at 10 pct., ** at 5 pct., and *** at 1 pct. level.

Table A.13 First-stage models — Symmetric DGPs

Panel A: *SymmCorr*

	Pooled	X-first	Z-first
	(1)	(2)	(3)
Benchmark S1	0.720***	0.698***	0.743***
	(0.021)	(0.032)	(0.029)
Constant	50.801***	50.218***	51.382***
	(0.374)	(0.497)	(0.558)
p-value, H_0: *X-first = Z-first*	0.144		
Observations	1,202	600	602
R^2	0.558	0.534	0.583

Panel B: *SymmUncorr*

	Pooled	X-first	Z-first
	(1)	(2)	(3)
Benchmark S1	0.758***	0.771***	0.744***
	(0.033)	(0.048)	(0.046)
Constant	51.410***	51.503***	51.318***
	(0.335)	(0.498)	(0.451)
p-value, H_0: *X-first = Z-first*	0.874		
Observations	1,200	596	604
R^2	0.374	0.372	0.376

Notes: This table analyzes model formation in the first stage of the experiment by estimating the extent to which first-stage beliefs rely on the first-stage empirical benchmark for the *SymmCorr* DGP (Panel A) and *SymmUncorr* DGP (Panel B). This corresponds to $P_{emp}(Y = 1|X = x)$ for *X-first* subjects and $P_{emp}(Y = 1|Z = z)$ for *Z-first* subjects. Benchmarks are demeaned by their average across possible predictor values. The dependent variable is the reported belief about the success probability for a project with variable $(X = x)$ or $(Z = z)$, respectively. For each subject, there are two observations, corresponding to the two predictor values. Under symmetry the benchmark is identical across treatments, so Column 1 pools *X-first* and *Z-first* subjects. Columns 2 and 3 fit the same regression separately for each treatment group. The bottom row reports the p-value of a joint Wald test of H_0: *X-first* and *Z-first* have identical intercept and slope on the benchmark. Clustered standard errors are in parentheses. * denotes significance at 10 pct., ** at 5 pct., and *** at 1 pct. level.

Table A.14 Path dependence: Structural approach — Symmetric DGPs

Panel A: *SymmCorr*

	All	S1 dir. correct	CQ no mistake
	(1)	(2)	(3)
Benchmark S1	0.069***	0.090***	0.076***
	(0.020)	(0.022)	(0.021)
Benchmark S2	0.558***	0.607***	0.554***
	(0.026)	(0.027)	(0.028)
Constant	47.75***	47.02***	47.75***
	(0.47)	(0.49)	(0.53)
Observations	2,404	2,076	1,956
R^2	0.323	0.383	0.327

Panel B: *SymmUncorr*

	All	S1 dir. correct	CQ no mistake
	(1)	(2)	(3)
Benchmark S1	0.090***	0.110***	0.103***
	(0.027)	(0.031)	(0.031)
Benchmark S2	0.508***	0.577***	0.518***
	(0.023)	(0.026)	(0.026)
Constant	48.25***	47.51***	48.27***
	(0.49)	(0.51)	(0.54)
Observations	2,400	1,920	1,964
R^2	0.293	0.380	0.310

Notes: This table reports estimates of Equation 6 for the *SymmCorr* (Panel A) and *SymmUncorr* (Panel B) experiments across different samples. The dependent variable is the reported belief about the success probability for a project with variables $(X = x, Z = z)$. For each subject, there are four observations, corresponding to the four combinations of predictor values. Benchmark S1 is the first-stage empirical benchmark, which corresponds to $P_{emp}(Y \mid X = x)$ for *X-first* subjects and $P_{emp}(Y \mid Z = z)$ for *Z-first* subjects. Benchmark S2 is the second-stage empirical benchmark $P_{emp}(Y \mid (X, Z) = (x, z))$. Both benchmarks are demeaned by the unconditional mean of 50. Column 1 uses the full sample. Column 2 restricts the sample to subjects who update in the correct direction in stage 1, that is infer a higher success probability for the objectively better value, and Column 3 to those who passed the comprehension quiz on their first attempt. Clustered standard errors are in parentheses. * denotes significance at 10 pct., ** at 5 pct., and *** at 1 pct. level.

Table A.15 Treatment effect on project choice — Symmetric DGPs

Panel A: *SymmCorr*

	Pooled	ΔX	ΔZ	ΔX, Z=0	ΔX, Z=1	ΔZ, X=0	ΔZ, X=1
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
First-stage predictor varies	1.082						
	(1.303)						
X-first		1.927	−0.233	0.567	3.287	−2.431	1.965
		(2.460)	(2.323)	(3.731)	(2.681)	(3.759)	(2.411)
Constant	78.952***	78.073***	80.066***	70.100***	86.047***	70.764***	89.369***
	(1.220)	(1.792)	(1.631)	(2.643)	(2.001)	(2.626)	(1.780)
Observations	2,404	1,202	1,202	601	601	601	601
R^2	0.000	0.001	0.000	0.000	0.003	0.001	0.001

Panel B: *SymmUncorr*

	Pooled	ΔX	ΔZ	ΔX, Z=0	ΔX, Z=1	ΔZ, X=0	ΔZ, X=1
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
First-stage predictor varies	4.083***						
	(1.452)						
X-first		5.769**	−2.395	6.471**	5.067	1.098	−5.887*
		(2.494)	(2.514)	(2.878)	(3.263)	(3.118)	(3.029)
Constant	80.750***	79.801***	84.106***	82.119***	77.483***	81.788***	86.424***
	(1.345)	(1.924)	(1.670)	(2.209)	(2.408)	(2.225)	(1.974)
Observations	2,400	1,200	1,200	600	600	600	600
R^2	0.003	0.006	0.001	0.008	0.004	0.000	0.006

Notes: This table presents treatment effects from an exogenous manipulation of the first-stage independent variable on project choice in the second stage for the *SymmCorr* DGP (Panel A) and *SymmUncorr* DGP (Panel B). The dependent variable is an indicator equal to 1 if the subject chose the project with the higher value of a predictor when comparing two projects that differ only in that predictor. "First-stage predictor varies" is an indicator equal to one if the choice is between projects that vary the predictor observed in the first stage. *X-first* is an indicator equal to 1 if the subject observed X in the first stage. Column 1 pools across all project comparisons. For each subject, there are four observations, corresponding to the four comparisons between projects that differ in exactly one predictor. Columns 2 and 3 report the share of subjects choosing the project with the higher value of X and Z, respectively, pooling across the two values of the other variable. For each subject, there are two observations, corresponding to the two values of the other variable. Columns 4–7 report choices separately for each value of the other predictor. Clustered standard errors (Columns 1–3) and robust standard errors (Columns 4–7) are in parentheses. * denotes significance at 10 pct., ** at 5 pct., and *** at 1 pct. level.

Table A.16 Treatment effect on project valuation — Symmetric DGPs

Panel A: *SymmCorr*

	Pooled	ΔX	ΔZ	ΔX, Z=0	ΔX, Z=1	ΔZ, X=0	ΔZ, X=1
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
First-stage predictor varies	0.113						
	(0.183)						
X-first		0.106	−0.119	−0.128	0.340	−0.566	0.328
		(0.350)	(0.327)	(0.505)	(0.407)	(0.497)	(0.379)
Constant	4.257***	4.104***	4.531***	2.724***	5.483***	3.075***	5.987***
	(0.172)	(0.260)	(0.236)	(0.363)	(0.304)	(0.355)	(0.286)
Observations	2,404	1,202	1,202	601	601	601	601
R^2	0.000	0.000	0.000	0.000	0.001	0.002	0.001

Panel B: *SymmUncorr*

	Pooled	ΔX	ΔZ	ΔX, Z=0	ΔX, Z=1	ΔZ, X=0	ΔZ, X=1
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
First-stage predictor varies	0.411**						
	(0.184)						
X-first		0.915***	0.092	1.022***	0.809*	0.497	−0.313
		(0.345)	(0.351)	(0.386)	(0.437)	(0.423)	(0.420)
Constant	4.014***	3.821***	4.118***	3.887***	3.755***	3.796***	4.439***
	(0.180)	(0.249)	(0.237)	(0.278)	(0.308)	(0.300)	(0.276)
Observations	2,400	1,200	1,200	600	600	600	600
R^2	0.002	0.008	0.000	0.012	0.006	0.002	0.001

Notes: This table presents treatment effects from an exogenous manipulation of the first-stage independent variable on willingness to pay (WTP) in the second stage for the *SymmCorr* DGP (Panel A) and *SymmUncorr* DGP (Panel B). The dependent variable is the amount subjects are willing to pay for a project with a higher value of a predictor, holding the value of the other predictor constant. "First-stage predictor varies" is an indicator equal to one if the choice is between projects that vary the predictor observed in the first stage. *X-first* is an indicator equal to 1 if the subject observed X in the first stage. Column 1 pools across all project comparisons. For each subject, there are four observations, corresponding to the four comparisons between projects that differ in exactly one predictor. Columns 2 and 3 report WTP for projects with a higher value of X and Z, respectively, pooling across the two values of the other variable. For each subject, there are two observations, corresponding to the two values of the other variable. Columns 4–7 report WTP separately for each value of the other predictor. Clustered standard errors (Columns 1–3) and robust standard errors (Columns 4–7) are in parentheses. * denotes significance at 10 pct., ** at 5 pct., and *** at 1 pct. level.

Table A.17 Summary statistics on time spent across stages — All DGPs

Panel A: Baseline DGP

Variable	All	X-first	Z-first	p-value (KS)
Time spent S1	05:44 (04:44)	05:44 (04:44)	05:45 (04:43)	0.519
Time spent S2	07:22 (06:10)	07:30 (06:35)	07:15 (05:59)	0.091
Relative time S2/(S1+S2)	0.55	0.56	0.54	0.147
Above med. rel. time S2	50%	54%	46%	0.201
Sample size	784	394	390	

Panel B: *SymmCorr*

Variable	All	X-first	Z-first	p-value (KS)
Time spent S1	05:31 (04:32)	05:26 (04:35)	05:35 (04:32)	0.382
Time spent S2	07:39 (06:11)	07:41 (06:41)	07:37 (05:49)	0.231
Relative time S2/(S1+S2)	0.56	0.56	0.56	0.485
Above med. rel. time S2	50%	52%	48%	0.922
Sample size	601	300	301	

Panel C: *SymmUncorr*

Variable	All	X-first	Z-first	p-value (KS)
Time spent S1	05:39 (04:42)	05:31 (04:41)	05:46 (04:44)	0.736
Time spent S2	07:17 (06:15)	07:16 (06:16)	07:18 (06:08)	0.698
Relative time S2/(S1+S2)	0.55	0.55	0.55	0.278
Above med. rel. time S2	50%	52%	48%	0.970
Sample size	600	298	302	

Notes: This table reports descriptive statistics on the time, measured in minutes and seconds, that subjects allocated to the first (Row 1) and second stage (Row 2) of the experiment by assigned treatment group in the *Baseline* (Panel A), *SymmCorr* (Panel B), and *SymmUncorr* (Panel C) experiments. Column 1 reports statistics for the full sample, while Columns 2 and 3 report statistics separately for subjects assigned to the *X-first* and *Z-first* treatment groups, respectively. Row 1 and 2 report mean values, with median values in parentheses. Row 3 reports the average share of total time spent in the second stage. Row 4 reports the percentage of subjects who spent an above-median share of total time in the second stage. Column 4 reports the p-value of a Kolmogorov-Smirnov-test for the equality of distribution across treatment groups.

Table A.18 Path dependence by click data on belief elicitation page

	All	Extra Clicks Beliefs S2	No Extra Clicks Beliefs S2
	(1)	(2)	(3)
Benchmark X	0.156***	0.130***	0.170***
	(0.030)	(0.043)	(0.040)
Benchmark X	0.087**	0.099*	0.082
× X first	(0.039)	(0.054)	(0.054)
Benchmark Z	0.500***	0.680***	0.404***
	(0.021)	(0.031)	(0.025)
Benchmark Z	0.035	−0.012	0.047
× X first	(0.030)	(0.044)	(0.037)
Constant	47.348***	43.607***	49.545***
	(0.447)	(0.738)	(0.538)
Observations	3,136	1,160	1,976
R^2	0.363	0.499	0.294

Notes: This table reports results from estimating Equation 5 separately for subjects who revisited the data table on the belief-elicitation page in Stage 2 (Column 2) and subjects who did not revisit the data table (Column 3). Column 1 reports results for the full sample. The dependent variable is the reported belief about the success probability for a project with variables $(X = x, Z = z)$. For each subject, there are four observations, corresponding to the four combinations of predictor values. Clustered standard errors are in parentheses. * denotes significance at 10 pct., ** at 5 pct., and *** at 1 pct. level.

Table A.19 Treatment effect on second-stage models by cognitive effort

Panel A: Low relative S2 time (below median)

	Extensive margin Share agreeing (in %)		**Intensive margin** $\Delta P(Y = \text{Success} \mid \cdots)$		**Intensive margin - disaggregated** $\Delta P(Y = \text{Success} \mid \cdots)$			
	X matters	Z matters	ΔX	ΔZ	ΔX, Z = 0	ΔX, Z = 1	ΔZ, X = 0	ΔZ, X = 1
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
X-first	11.795***	0.879	4.901***	0.600	3.884*	5.919**	−0.417	1.618
	(4.305)	(3.649)	(1.781)	(2.353)	(2.292)	(2.575)	(2.910)	(2.856)
Constant	69.524***	84.286***	4.502***	23.545***	−1.648	10.652***	17.395***	29.695***
	(3.185)	(2.518)	(1.270)	(1.622)	(1.596)	(1.786)	(2.087)	(1.848)
Observations	392	392	784	784	392	392	392	392
R^2	0.018	0.000	0.009	0.000	0.007	0.013	0.000	0.001

Panel B: High relative S2 time (at or above median)

	Extensive margin Share agreeing (in %)		**Intensive margin** $\Delta P(Y = \text{Success} \mid \cdots)$		**Intensive margin - disaggregated** $\Delta P(Y = \text{Success} \mid \cdots)$			
	X matters	Z matters	ΔX	ΔZ	ΔX, Z = 0	ΔX, Z = 1	ΔZ, X = 0	ΔZ, X = 1
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
X-first	8.637*	0.398	0.567	1.445	−0.944	2.077	−0.066	2.956
	(4.554)	(2.865)	(1.557)	(2.453)	(1.966)	(2.108)	(2.998)	(2.553)
Constant	67.778***	91.111***	4.903***	37.553***	0.600	9.206***	33.250***	41.856***
	(3.492)	(2.127)	(1.279)	(1.795)	(1.603)	(1.503)	(2.121)	(1.873)
Observations	392	392	784	784	392	392	392	392
R^2	0.009	0.000	0.000	0.001	0.001	0.002	0.000	0.003

Notes: This table replicates Table 2 separately for subjects with below-median relative time spent in the second stage (Panel A) and subjects with above-median relative time spent in the second stage (Panel B). Relative time spent in the second stage is measured as the share of time spent in the second stage relative to the total time spent in stages 1 and 2 combined. *X-first* is an indicator equal to one if the subject observed X in the first stage. Columns 1 and 2 report the share of subjects agreeing that the predictor variables X and Z, respectively, have a *ceteris paribus* effect on the outcome. Columns 3 and 4 report beliefs about the marginal effects of X and Z, respectively, pooling across the two possible values of the other variable. Columns 5–8 report the disaggregated marginal effects. Clustered standard errors (Columns 1–4) and robust standard errors (Columns 5–8) are in parentheses. * denotes significance at 10 pct., ** at 5 pct., and *** at 1 pct. level.

Table A.20 Effect of relative effort on extensive models

	Z-first				X-first			
	Only X	Only Z	X and Z	None	Only X	Only Z	X and Z	None
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
High rel.	−0.053**	0.033	0.036	−0.015	−0.039*	0.073**	−0.010	−0.024
S2 effort	(0.023)	(0.044)	(0.049)	(0.026)	(0.024)	(0.036)	(0.045)	(0.024)
Constant	0.081***	0.229***	0.614***	0.076***	0.077***	0.115***	0.736***	0.071***
	(0.019)	(0.029)	(0.034)	(0.018)	(0.020)	(0.024)	(0.033)	(0.019)
Observations	390	390	390	390	394	394	394	394
R^2	0.013	0.001	0.001	0.001	0.007	0.010	0.000	0.003

Notes: This table presents the effects of spending more relative time in the second stage on subjects' mental models at the extensive margin, separately for the *Z-first* group (Columns 1–4) and the *X-first* group (Columns 5–8). *High rel. S2 effort* is an indicator equal to 1 if a subject spent above-median relative time in the second stage, measured as the share of time spent in the second stage relative to the total time spent across stages 1 and 2. The table reports the share of subjects holding each of the four possible extensive models: only X matters (Columns 1 and 5), only Z matters (Columns 2 and 6), both variables matter (Columns 3 and 7), or neither variable matters (Columns 4 and 8) for project success. A variable is considered to matter if subjects agree that the predictors X and Z, respectively, have a *ceteris paribus* effect on the outcome. Robust standard errors are in parentheses. * denotes significance at 10 pct., ** at 5 pct., and *** at 1 pct. level.

Table A.21 Summary statistics on time spent across stages by reasoning type

Variable	**Frequentist**	**Separate**	**Abs. success**	**Undetermined**
Time spent S1	06:08	04:40	05:15	05:55
	(05:12)	(04:03)	(04:07)	(04:44)
Time spent S2	09:45	05:28	06:40	06:10
	(08:38)	(04:17)	(04:57)	(04:54)
Relative time S2/(S1+S2)	0.61	0.52	0.54	0.51
Above med. rel. time S2	73%	35%	46%	36%
Sample size	265	92	123	304

Notes: This table reports descriptive statistics on the time, measured in minutes and seconds, that subjects allocated to the first (Row 1) and second stage (Row 2) of the *Baseline* experiment by reasoning type. Columns 1–3 report results for the three specific well-defined types: *Frequentist*, *Separate*, and *Absolute Success*. Column 4 reports results for subjects with an undetermined type, comprising all responses that cannot be assigned to any of the other three categories. Row 1 and 2 report mean values, with median values in parentheses. Row 3 reports the average share of total time spent in the second stage. Row 4 reports the percentage of subjects who spent an above-median share of total time in the second stage.

Table A.22 Path dependence by cognitive effort controlling for reasoning type

	Subjective success probability (pooled), by effort		
	Baseline	Controlling for reasoning	Controlling for reasoning × treatment
	(1)	(2)	(3)
High effort	−1.360	−1.044	−1.044
	(0.893)	(1.006)	(1.007)
Benchmark X	0.150***	0.130***	0.127***
	(0.042)	(0.042)	(0.042)
Benchmark X	0.163***	0.165***	0.174***
× X-first	(0.059)	(0.059)	(0.058)
Benchmark Z	0.392***	0.467***	0.463***
	(0.027)	(0.025)	(0.025)
Benchmark Z	0.010	−0.010	−0.007
× X-first	(0.039)	(0.036)	(0.037)
High effort	0.013	0.055	0.062
× Benchmark X	(0.060)	(0.062)	(0.067)
High effort	−0.144*	−0.151*	−0.168**
× Benchmark X × X-first	(0.079)	(0.077)	(0.084)
High effort	0.233***	0.095**	0.101**
× Benchmark Z	(0.040)	(0.038)	(0.040)
High effort	0.014	0.031	0.025
× Benchmark Z × X-first	(0.057)	(0.050)	(0.057)
Constant	48.028***	50.014***	50.014***
	(0.622)	(0.844)	(0.844)
Controls	–	Reasoning	Reasoning × X-first
Observations	3,136	3,136	3,136
R^2	0.383	0.444	0.445

Notes: This table analyzes the model formation in the second stage of the experiment by subjects' reasoning type and attention allocation. It reports estimates of Equation 5 interacting all covariates with *high effort*, an indicator equal to 1 if a subject spent above-median relative time in the second stage, measured as the share of time spent in stage 2 relative to the total time spent across stages 1 and 2. Column 1 reports results without controlling for reasoning type. Column 2 additionally includes interactions between reasoning type and the benchmarks. Column 3 further adds interactions between reasoning type, the benchmarks, and treatment. The dependent variable is the reported belief about the success probability for a project with variables $(X = x, Z = z)$. For each subject, there are four observations, corresponding to the four combinations of predictor values. Clustered standard errors are in parentheses. * denotes significance at 10 pct., ** at 5 pct., and *** at 1 pct. level.

Table A.23 Path dependence by cognitive effort — Symmetric DGPs

	SymmCorr			*SymmUncorr*		
	All	High rel. S2 effort	Low rel. S2 effort	All	High rel. S2 effort	Low rel. S2 effort
	(1)	(2)	(3)	(4)	(5)	(6)
Benchmark S1	0.069***	0.033	0.104***	0.090***	0.016	0.165***
	(0.020)	(0.025)	(0.030)	(0.027)	(0.038)	(0.039)
Benchmark S2	0.558***	0.723***	0.392***	0.508***	0.648***	0.368***
	(0.026)	(0.033)	(0.037)	(0.023)	(0.032)	(0.032)
Constant	47.750***	46.140***	49.366***	48.249***	47.255***	49.243***
	(0.473)	(0.623)	(0.702)	(0.493)	(0.670)	(0.720)
Observations	2,404	1,204	1,200	2,400	1,200	1,200
R^2	0.323	0.456	0.211	0.293	0.392	0.210
Adjusted R^2	0.322	0.455	0.210	0.293	0.391	0.209
Note:						*p<0.1; **p<0.05; ***p<0.01

Notes: This table reports estimates of Equation 6 for the *SymmCorr* (Columns 1–3) and *SymmUncorr* (Columns 4–6) experiments. Columns 1 and 4 use the full sample, Columns 2 and 5 restrict the sample to subjects with above-median relative time spent in the second stage, and Columns 3 and 6 restrict the sample to subjects with below-median relative time spent in the second stage. Relative time spent in the second stage is measured as the share of time spent in the second stage relative to the total time spent in stages 1 and 2 combined. The dependent variable is the reported belief about the success probability for a project with variables $(X = x, Z = z)$. For each subject, there are four observations, corresponding to the four combinations of predictor values. Clustered standard errors are in parentheses. * denotes significance at 10 pct., ** at 5 pct., and *** at 1 pct. level.

Table A.24 Treatment effect on second-stage models by cognitive effort — Symmetric DGPs

	Symmetric-Correlated DGP			**Symmetric-Uncorrelated DGP**		
	All	High rel. S2 effort	Low rel. S2 effort	All	High rel. S2 effort	Low rel. S2 effort
	(1)	(2)	(3)	(4)	(5)	(6)
First-stage predictor	3.439***	1.664	5.220***	2.715***	0.483	4.947***
	(0.983)	(1.268)	(1.499)	(0.824)	(1.154)	(1.166)
Constant	17.433***	22.585***	12.265***	15.240***	19.427***	11.053***
	(0.798)	(1.031)	(1.147)	(0.701)	(0.971)	(0.952)
Observations	2,404	1,204	1,200	2,400	1,200	1,200
R^2	0.005	0.001	0.010	0.004	0.000	0.012

Notes: This table extends Table A.12 by splitting the sample by relative effort in the second stage for the *SymmCorr* DGP (Columns 1–3) and the *SymmUncorr* DGP (Columns 4–6). Columns 1 and 4 report the full sample; Columns 2 and 5 restrict the sample to subjects with above-median relative time spent in the second stage; Columns 3 and 6 restrict the sample to those with below-median relative time. Relative time spent in the second stage is measured as the share of time spent in the second stage relative to the total time spent in stages 1 and 2 combined. "First-stage predictor" is an indicator equal to one if the predictor was observed in the first stage. The dependent variable is the belief about the marginal effect of the variable; for each subject there are four observations, corresponding to beliefs about the marginal effects of the two predictors at both values of the other predictor. Clustered standard errors are in parentheses. * denotes significance at 10 pct., ** at 5 pct., and *** at 1 pct. level.

Table A.25 Path dependence by alternative effort measure and confidence — Baseline DGP

		Total time: stage 2		Confidence: stage 2	
	Full sample	below median	above median	below median	above median
	(1)	(2)	(3)	(4)	(5)
Benchmark X	0.156***	0.161***	0.151***	0.110***	0.203***
	(0.030)	(0.046)	(0.038)	(0.039)	(0.046)
Benchmark X	0.087**	0.135**	0.045	0.129**	0.044
× X-first	(0.039)	(0.062)	(0.048)	(0.055)	(0.057)
Benchmark Z	0.500***	0.405***	0.607***	0.421***	0.580***
	(0.021)	(0.029)	(0.029)	(0.027)	(0.031)
Benchmark Z	0.035	−0.011	0.056	−0.005	0.067
× X-first	(0.030)	(0.040)	(0.040)	(0.039)	(0.042)
Constant	47.348***	48.290***	46.406***	45.209***	49.402***
	(0.447)	(0.588)	(0.670)	(0.603)	(0.641)
Observations	3,136	1,568	1,568	1,536	1,600
R^2	0.363	0.272	0.460	0.296	0.433

Notes: This table reports estimates of Equation 5 for the baseline DGP separately by different sample splits, analyzing the weights that second-stage beliefs place on the empirical benchmarks. The dependent variable is the reported belief about the success probability for a project with variables $(X = x, Z = z)$; for each subject there are four observations, corresponding to the four combinations of predictor values. Column 1 uses the full sample and reproduces the headline specification (Column 1 of Table 3). Columns 2–3 split the sample at the participant-level median of total time spent in stage 2, while Columns 4–5 split the sample at the participant-level median of stated confidence in stage 2. Clustered standard errors are in parentheses. * denotes significance at 10 pct., ** at 5 pct., and *** at 1 pct. level.

Table A.26 Path dependence by alternative effort measure and confidence — Symmetric DGPs

Panel A: *SymmCorr*

		Total time: stage 2		Confidence: stage 2	
	All	below median	above median	below median	above median
	(1)	(2)	(3)	(4)	(5)
Benchmark S1	0.069***	0.132***	0.005	0.088***	0.050*
	(0.020)	(0.031)	(0.024)	(0.026)	(0.030)
Benchmark S2	0.558***	0.377***	0.738***	0.419***	0.692***
	(0.026)	(0.036)	(0.033)	(0.034)	(0.037)
Constant	47.75***	48.79***	46.72***	45.61***	49.82***
	(0.47)	(0.69)	(0.65)	(0.63)	(0.68)
Observations	2,404	1,200	1,204	1,180	1,224
R^2	0.323	0.225	0.438	0.259	0.390

Panel B: *SymmUncorr*

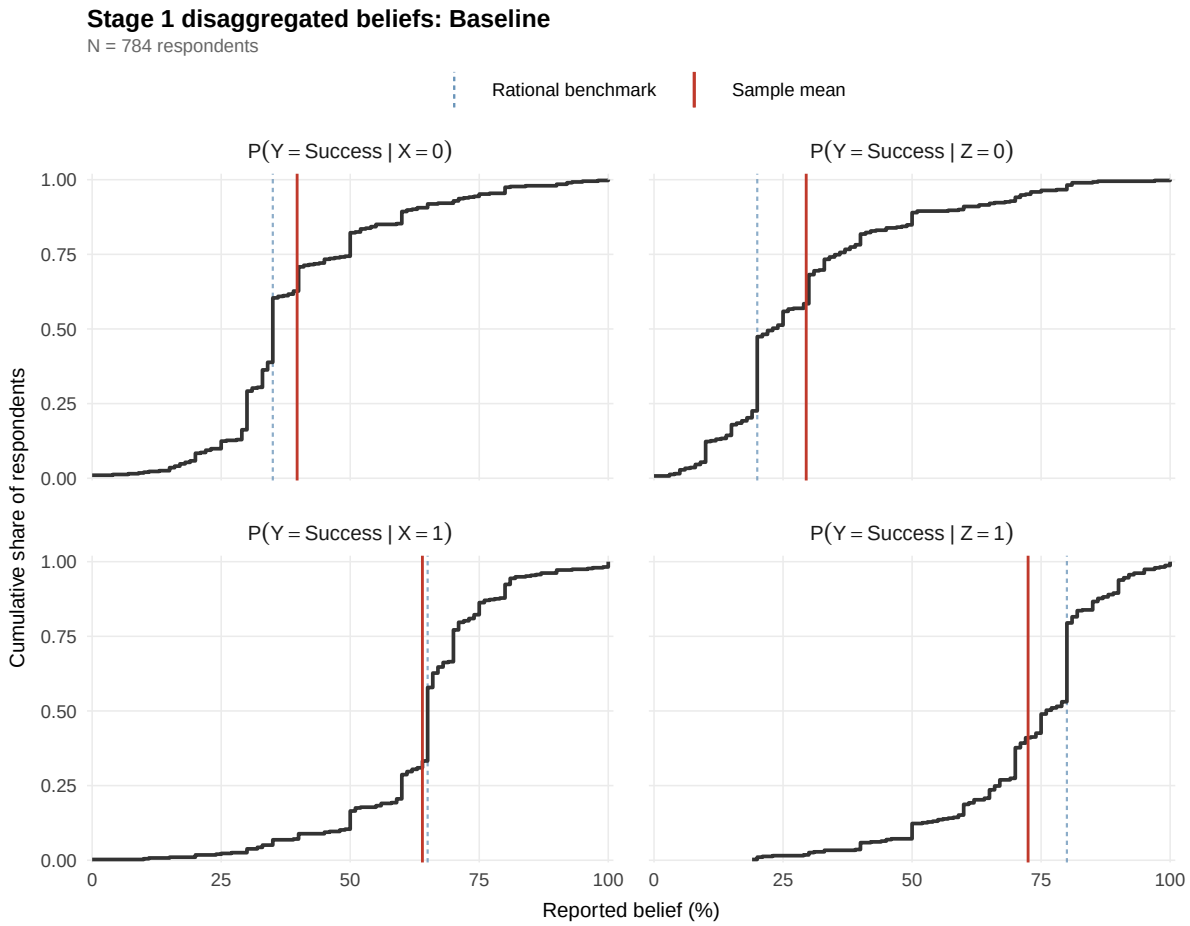
		Total time: stage 2		Confidence: stage 2	
	All	below median	above median	below median	above median
	(1)	(2)	(3)	(4)	(5)
Benchmark S1	0.090***	0.173***	0.008	0.104**	0.080**
	(0.027)	(0.040)	(0.037)	(0.045)	(0.034)
Benchmark S2	0.508***	0.343***	0.673***	0.370***	0.615***
	(0.023)	(0.032)	(0.031)	(0.033)	(0.031)
Constant	48.25***	48.83***	47.67***	45.95***	50.03***
	(0.49)	(0.71)	(0.69)	(0.68)	(0.68)
Observations	2,400	1,200	1,200	1,048	1,352
R^2	0.293	0.196	0.411	0.211	0.364

Notes: This table reports estimates of Equation 6 for the *SymmCorr* (Panel A) and *SymmUncorr* (Panel B) experiments separately by different sample splits. The dependent variable is the reported belief about the success probability for a project with variables $(X = x, Z = z)$. For each subject, there are four observations, corresponding to the four combinations of predictor values. Benchmark S1 is the first-stage empirical benchmark, which corresponds to $P_{emp}(Y \mid X = x)$ for *X-first* subjects and $P_{emp}(Y \mid Z = z)$ for *Z-first* subjects. Benchmark S2 is the second-stage empirical benchmark $P_{emp}(Y \mid (X, Z) = (x, z))$. Both benchmarks are demeaned by the unconditional mean of 50. Column 1 uses the full sample and reproduces the headline specification (Column 1 of Table A.14). Columns 2–3 split the sample at the participant-level median of total time spent in stage 2, while Columns 4–5 split the sample at the participant-level median of stated confidence in stage 2. Clustered standard errors are in parentheses. * denotes significance at 10 pct., ** at 5 pct., and *** at 1 pct. level.

B Additional Figures

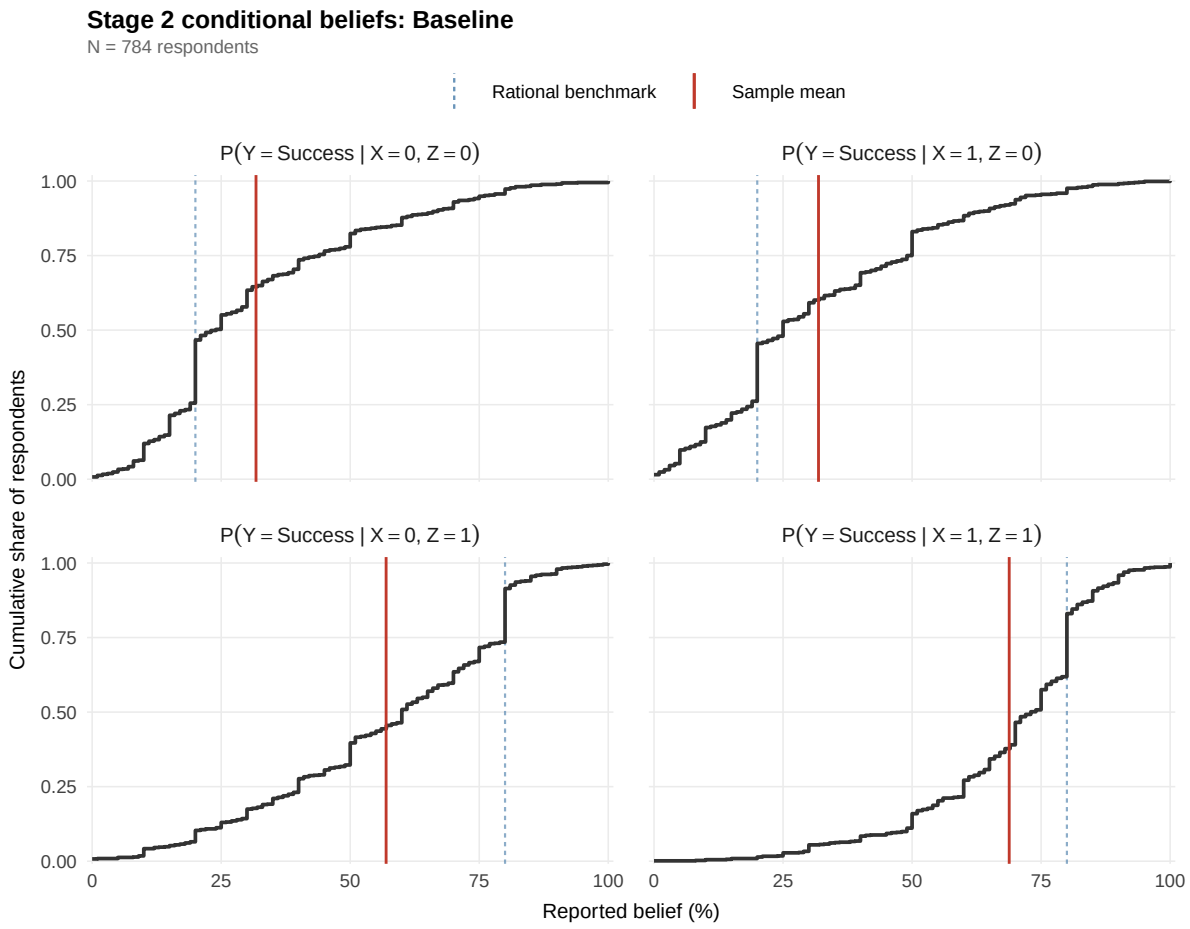
This appendix reports the distributions of subjects' raw elicited beliefs alongside the rational/frequentist benchmarks, separately for the first and second stage and for each of the three DGPs. Figures use the prepared analysis samples (i.e., post comprehension quiz, post speed-filter). In each panel the light dashed line marks the rational benchmark and the solid line marks the sample mean.

Figure B.1 Distribution of first-stage beliefs — *Baseline*



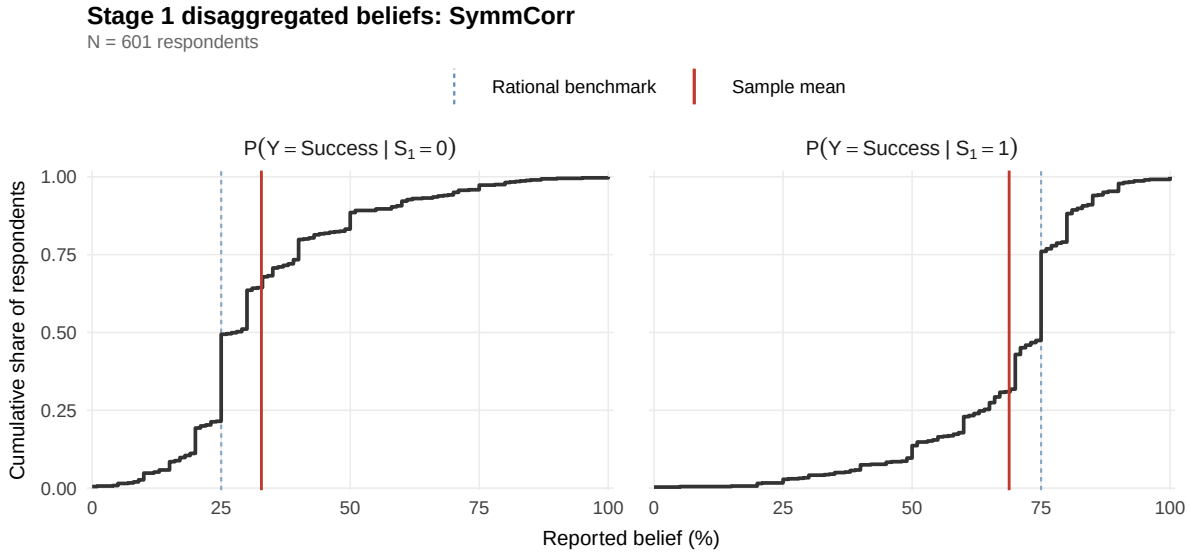
Notes: This figure shows empirical CDFs of subjects' raw first-stage beliefs $P(Y = \textit{Success} \mid S_1)$ in the *Baseline* experiment, where $S_1 = X$ for the *X-first* group and $S_1 = Z$ for the *Z-first* group. The figure is split into four panels because in *Baseline* the rational benchmark depends on the treatment arm (*X-first*: $35\%/65\%$; *Z-first*: $20\%/80\%$). Rows correspond to the belief about $P(Y \mid S_1 = 0)$ (top) and $P(Y \mid S_1 = 1)$ (bottom). Columns correspond to the two treatment arms. Within each panel, the dashed line marks the rational benchmark and the solid line marks the sample mean.

Figure B.2 Distribution of second-stage conditional beliefs — *Baseline*



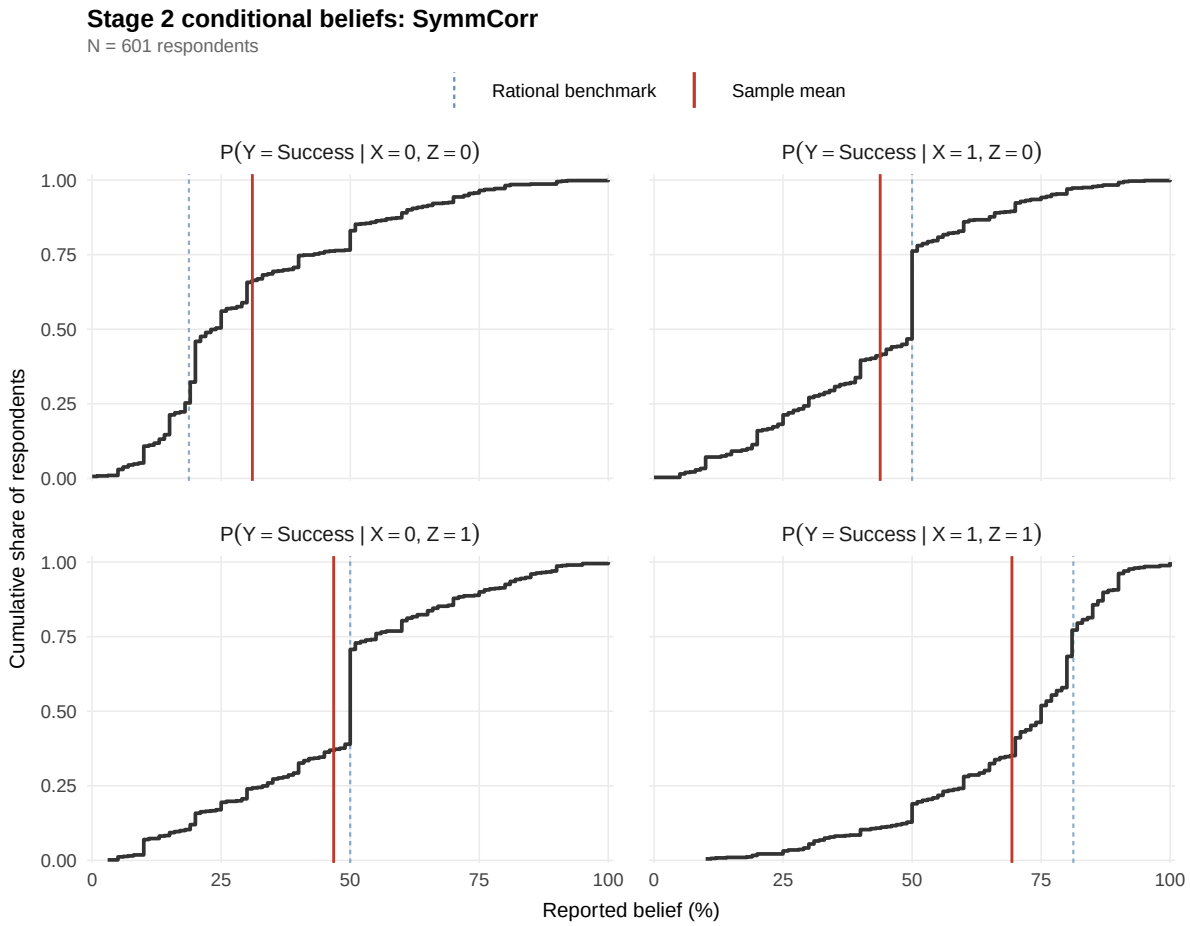
Notes: This figure shows empirical CDFs of subjects' raw second-stage conditional beliefs $P(Y = \textit{Success} \mid X, Z)$ in the *Baseline* experiment, for each of the four possible values of (X, Z). Beliefs are pooled across treatment arms because the rational second-stage benchmark does not depend on the assigned treatment ($20\%/20\%/80\%/80\%$ for both treatment arms). Within each panel, the dashed line marks the rational benchmark and the solid line marks the sample mean.

Figure B.3 Distribution of first-stage beliefs — *SymmCorr*



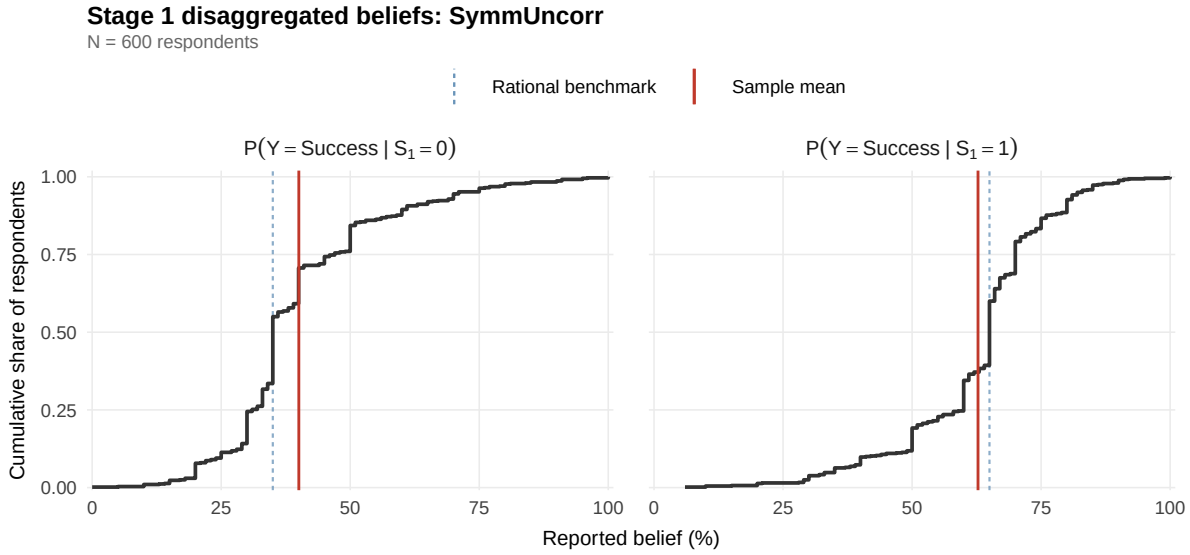
Notes: This figure shows empirical CDFs of subjects' raw first-stage beliefs $P(Y = \textit{Success} \mid S_1)$ in the *SymmCorr* experiment, pooled across treatment arms because the rational first-stage benchmark does not depend on the assigned treatment ($25\%/75\%$ for both treatment arms). Within each panel, the dashed line marks the rational benchmark and the solid line marks the sample mean.

Figure B.4 Distribution of second-stage conditional beliefs — *SymmCorr*



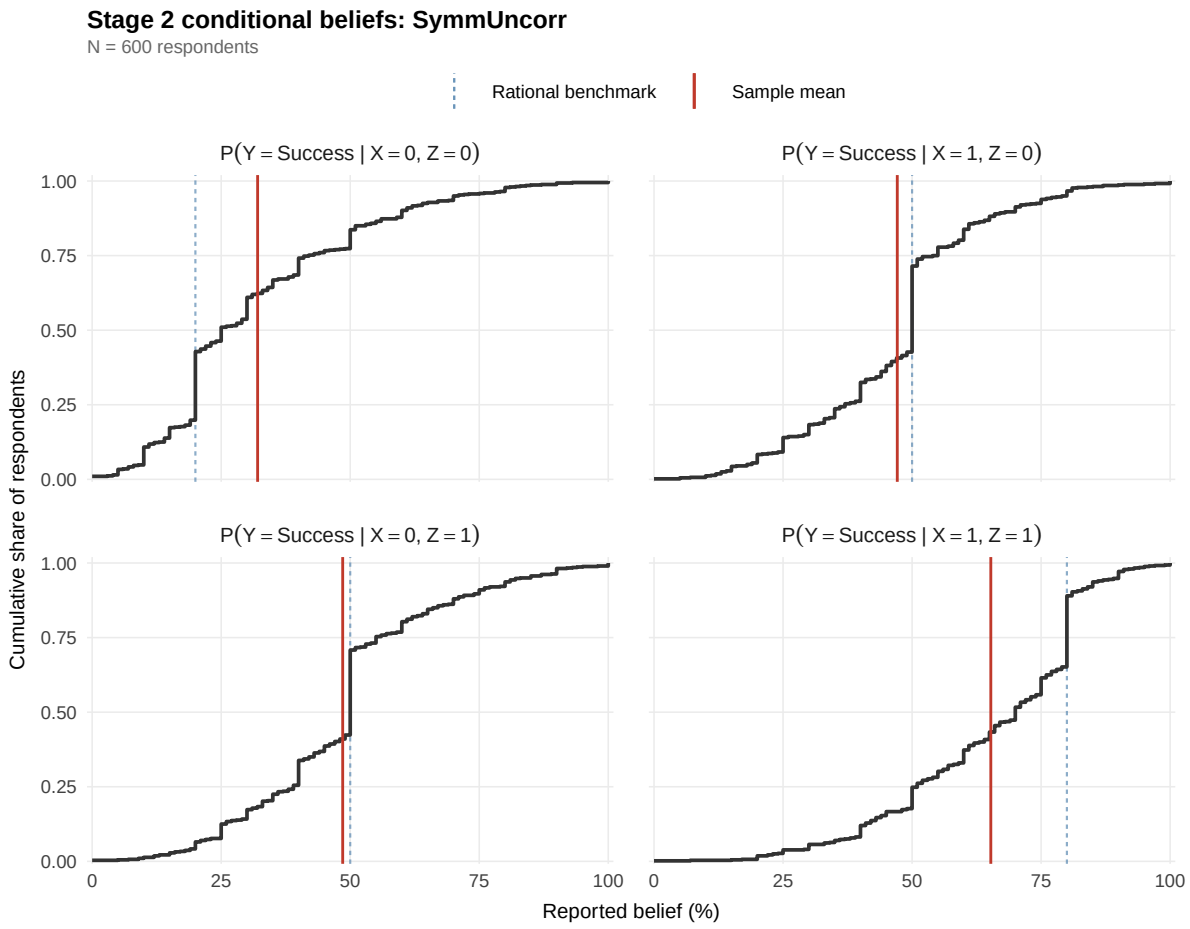
Notes: This figure shows empirical CDFs of subjects' raw second-stage conditional beliefs $P(Y = \textit{Success} \mid X, Z)$ in the *SymmCorr* experiment, for each of the four possible values of (X, Z). Beliefs are pooled across treatment arms because the rational second-stage benchmark does not depend on the assigned treatment ($18.75\%/50\%/50\%/81.25\%$ for both treatment arms). Within each panel, the dashed line marks the rational benchmark and the solid line marks the sample mean. The corner-cell benchmarks (18.75% / 81.25%) are non-integer, so the modal integer response (e.g., 20) sits within $\leq$ 2pp of the benchmark.

Figure B.5 Distribution of first-stage beliefs — *SymmUncorr*



Notes: This figure shows empirical CDFs of subjects' raw first-stage beliefs $P(Y = \textit{Success} \mid S_1)$ in the *SymmUncorr* experiment, pooled across treatment arms because the rational first-stage benchmark does not depend on the assigned treatment ($35\%/65\%$ for both treatment arms). Within each panel, the dashed line marks the rational benchmark and the solid line marks the sample mean.

Figure B.6 Distribution of second-stage conditional beliefs — *SymmUncorr*



Notes: This figure shows empirical CDFs of subjects' raw second-stage conditional beliefs $P(Y = \textit{Success} \mid X, Z)$ in the *SymmUncorr* experiment, for each of the four possible values of (X, Z). Beliefs are pooled across treatment arms because the rational second-stage benchmark does not depend on the assigned treatment ($20\%/50\%/50\%/80\%$ for both treatment arms). Within each panel, the dashed line marks the rational benchmark and the solid line marks the sample mean.

C Research Transparency

Preregistration We preregistered our surveys and experiments at AsPredicted.org, registration numbers #174937 #281,544 and #281,804. The preregistration includes details on the survey design, survey instructions, sampling process, planned sample size, exclusion criteria, and research questions. The pre-registrations for the *Baseline* experiment, *SymmCorr*, and *SymmUncorr* can be accessed at `https://aspredicted.org/xqfv-9s2d.pdf`, `https://aspredicted.org/cu4ub3.pdf`, and `https://aspredicted.org/5dj8x6.pdf`, respectively.

Conflicting interests We declare that we have no conflicting interests.

D Framework Extensions

This appendix discusses two extensions to our theoretical framework introduced in Section 3.

D.1 Heterogeneity in Learning Models from Datasets

The framework has abstracted from heterogeneity in how subjects translate observed data into beliefs about conditional success probabilities. We have assumed that individuals correctly represent the underlying statistical problem and converge to the correct solution when exerting sufficient cognitive effort. In practice, however, subjects may rely on distinct, potentially sub-optimal approaches for mapping observed data into beliefs about success probabilities.

We refer to a subject's approach for extracting conditional success probabilities from a fixed set of observations as their *reasoning type*. Formally, a reasoning type can be characterized by the estimator $\hat{P}(Y \mid X, Z)$ used to form beliefs. A subject's reasoning type therefore affects the second-stage beliefs they form from a given set of observations.

Ex ante, the relationship between stickiness and reasoning type is unclear. Stickiness may be orthogonal to an individual's reasoning type. Alternatively, stickiness may arise only for specific reasoning types, or the two may be correlated, either because both are associated with a certain degree of naivety or because reasoning types are correlated with how subjects allocate cognitive effort across stages. In Section 4.3, we consider common reasoning types and examine their relationship to stickiness.

D.2 Different Effort Across Groups

Treatment groups may exert different effort in Stage 2. One potential reason is that the strength of the first-stage predictor affects incentives to learn: if a predictor has a stronger marginal effect, DMs may rely more on their initial beliefs and exert less effort in Stage 2.

Assume both groups exert effort $\bar{e}_1$ in Stage 1, but may differ in effort exerted in Stage 2, denoted by $\bar{e}_2^X$ and $\bar{e}_2^Z$ for the *X-first* and *Z-first* treatments, respectively. The resulting difference in second-stage beliefs about the marginal effect of X is given by:

$$\left(\lambda(\bar{e}_2^X) - \lambda(\bar{e}_2^Z)\right) \cdot \Delta^{(2)}_{X|Z} + \left(1 - \lambda(\bar{e}_2^X)\right) \cdot \lambda(\bar{e}_1) \cdot \Delta^{(1)}_X \tag{7}$$

and analogously, the difference in beliefs about the marginal effect of Z is given by:

$$\left(\lambda(\bar{e}_2^Z) - \lambda(\bar{e}_2^X)\right) \cdot \Delta^{(2)}_{Z|X} + \left(1 - \lambda(\bar{e}_2^Z)\right) \cdot \lambda(\bar{e}_1) \cdot \Delta^{(1)}_{Z} \tag{8}$$

When effort differs across groups in Stage 2, the overall effect on belief differences is generally ambiguous and depends on the empirical benchmarks for both unconditional and conditional marginal effects, as well as on the effort exerted by each group.

While the general effect of heterogeneous effort is ambiguous, sharper predictions emerge once we consider a specific DGP. In our *Baseline* DGP, the unconditional marginal effect of Z is stronger than that of X. Suppose therefore that DMs in the *Z-first* treatment exert less effort in Stage 2 than DMs in the *X-first* treatment, i.e., $\bar{e}_2^Z < \bar{e}_2^X$.

Since the conditional marginal effect of X is zero, differences in second-stage beliefs about X depend only on the effort exerted by the *X-first* group. As long as DMs in the *X-first* treatment do not exert sufficient effort to fully converge to the empirical benchmark, we continue to expect higher perceived marginal effects of X in the *X-first* treatment.

For Z, the unconditional and conditional marginal effects coincide. Differences in beliefs about the marginal effect of Z are therefore attenuated as the *X-first* group exerts more effort relative to the *Z-first* group.

E Empirical Approach

Based on theoretical considerations and pilot data, we first identified the most prominent types of reasoning and created a coding manual detailing how to assign the mutually exclusive labels. Two research assistants then coded the text responses independently and unaware of the research hypotheses. A response was assigned a reasoning type if both two coders agreed. When the two coders disagreed, we served as tie breakers by assigning a label. If this label matched either of the two coders' labels, that reasoning type was assigned. Otherwise, the response was labeled *Undetermined*. Table E.1 summarizes the coding scheme. To check the robustness of our coding approach, we employed OpenAI's GPT-5.5 API in a few-shot reasoning approach (Kojima et al. 2022), instructing it to assign specific reasoning types conservatively and the label *Undetermined* when responses were ambiguous. We provide an overview of robustness analyses using AI-generated reasoning labels in Appendix E.1.

Table E.1 Overview of categories of the coding scheme

Category	Explanation	Examples
Frequentist	Subjects who determine the success likelihood by correctly grouping the projects based on their joint Color (X) and Card (Z) combination. For each combination, they determine the likelihood of success by dividing the number of successes compared to the total occurrences of projects with the same combination of X and Z.	"I counted the number of relevant successes and failures for each set of features and then derived a probability of success based on number of successes divided by the total number of experiments for that set of features."
		"To determine the success likelihoods I considered the number of successful examples versus the number of failures of the same example to stimate the percent of success. Those with a higher success to failure ratio therefore had a higher percentage likelihood and were favored."
Separate	Subjects who determine the success likelihood of a project by assessing the variables X and Z separately and then aggregate the unconditional effects of both variables to derive the likelihood of success. By simply aggregating the unconditional effects, they fail to account for the correlation between X and Z.	"I looked at the table and made a rough estimate of how likely each symbol was associated with success. When it came to combining symbols I combined those odds. So a low chance of success symbol combined with another low chance symbol would have a lower chance than either separately. A high chance combined with a low chance would be somewhere in the middle."
		"I first counted how likely each individual metric was to succeed individually: Blue Circle = 65% Green Circle = 35% Club = 20% Diamond = 80% Then I based my predictions on these. For the single metric predictions, I simply estimated around what they would be individually. For the the multi-metric (e.g. Blue Diamond) I averaged the two metrics and picked the project that was most likely to succeed based on past outcomes."
Absolute success	Subjects who compare the absolute number of successes with for variable combinations. By focusing only on the number of successes instead of the relative success likelihood, they fail to account that some types of projects occur more frequently in the sample.	"I looked at the trends in the suit or the color. For instance I looked at the diamonds with green to see how many of them were successful to determine how likely it was to be successful. I then compared it to the other option (diamond blue, clubs blue, clubs green, etc.) to determine which one would be more successful and which one I would have a better chance with."
		"I counted how many successes there were on each project and used that information to choose."
Undetermined	Responses that are not clearly classifiable in the categories above. This may be because the responses do not specify any strategy, because they are ambiguous and in principle consistent with several of the above strategies, or because they suggest that the specified beliefs were random or without explicit consideration of the data.	"I took into consideration the color, card and outcome"
		"I was indifferent to which project I choose. I like green so I picked green. I like diamonds so I picked diamonds."

Notes: This table provides an overview of the different categories in our coding scheme, an explanation for each category, and example extracts from the open-text responses.

E.1 Robustness analyses using AI-labeled reasoning labels

Table E.2 Summary statistics on time spent across stages by reasoning type (AI codes)

Variable	All	Frequentist	Separate	Abs. success	Undetermined
Time spent S1	05:44 (04:44)	06:17 (05:16)	05:06 (04:11)	05:07 (03:57)	05:55 (04:46)
Time spent S2	07:22 (06:10)	09:29 (08:28)	05:49 (04:37)	06:33 (04:44)	06:27 (05:06)
Relative time S2/(S1+S2)	0.55	0.6	0.52	0.55	0.52
Above med. rel. time S2	50%	67%	39%	49%	39%
Sample size	784	284	227	75	198

Notes: This table reports descriptive statistics on the time allocated by subjects to the first and second stage of the experiment in minutes and seconds by the assigned reasoning type group (labeled using GPT-5.5). Rows 1 and 2 report the mean time spent in stage 1 and stage 2, respectively, with the median in parentheses. Row 3 reports the average share allocated to the second stage relative to time spent in both stages. Row 4 reports the share of subjects that spent above-median relative time in the second stage.

Table E.3 Second-stage models across reasoning types (AI Codes)

	Subjective success probability (pooled), by reasoning				
	All	Frequentist	Separate	Abs. success	Undetermined
	(1)	(2)	(3)	(4)	(5)
Benchmark X	0.082***	0.053*	0.209***	0.071	−0.015
	(0.021)	(0.027)	(0.049)	(0.045)	(0.045)
Benchmark Z	0.401***	0.664***	0.387***	0.193***	0.118***
	(0.019)	(0.029)	(0.032)	(0.043)	(0.038)
BM number of successes	0.235***	0.170***	0.236***	0.369***	0.275***
	(0.021)	(0.033)	(0.036)	(0.067)	(0.049)
Constant	47.348***	46.929***	47.416***	39.040***	51.019***
	(0.447)	(0.609)	(0.651)	(1.929)	(1.046)
Observations	3,136	1,136	908	300	792
R^2	0.374	0.633	0.413	0.268	0.128

Notes: This table analyzes second-stage models by reasoning type (labeled using GPT-5.5), by estimating the extent to which second-stage models rely on the empirical benchmarks of the first-stage conditional success probabilities *Benchmark X* ($P_{emp}(Y = 1|X = x)$) and *Benchmark Z* ($P_{emp}(Y = 1|Z = z)$), as well as the empirical benchmark for the (rescaled) number of successes *BM number of successes* ($P_{emp}(X = x, Z = z|Y = 1)$). All benchmarks are demeaned by their average across possible predictor values. The dependent variable is the reported belief about the success probability for a project with variables $(X = x, Z = z)$. For each subject, there are four observations, corresponding to the four combinations of predictor values. Column 1 reports results for the full sample. Column 2–4 report results for the three specific well-defined types: *Frequentist*, *Separate*, and *Absolute Success*. Column 5 reports results for subjects with an undetermined type, comprising all responses that cannot be assigned to any of the other three categories. Clustered standard errors are in parentheses. * denotes significance at 10 pct., ** at 5 pct., and *** at 1 pct. level.

Table E.4 Path dependence across reasoning types (AI codes)

	Grouped			Specific types		
	All	Well-defined type	Undetermined	Frequentist	Separate	Abs. success
	(1)	(2)	(3)	(4)	(5)	(6)
Benchmark X	0.156***	0.166***	0.127**	0.102***	0.250***	0.144**
	(0.030)	(0.034)	(0.065)	(0.035)	(0.071)	(0.067)
Benchmark X	0.087**	0.118***	−0.010	0.069	0.156*	0.225**
× X first	(0.039)	(0.045)	(0.080)	(0.049)	(0.090)	(0.097)
Benchmark Z	0.500***	0.586***	0.255***	0.736***	0.491***	0.334***
	(0.021)	(0.023)	(0.039)	(0.028)	(0.037)	(0.048)
Benchmark Z	0.035	0.041	0.001	0.024	0.028	0.087
× X first	(0.030)	(0.031)	(0.057)	(0.041)	(0.050)	(0.070)
Constant	47.348***	46.108***	51.019***	46.929***	47.416***	39.040***
	(0.447)	(0.472)	(1.047)	(0.609)	(0.651)	(1.932)
Observations	3,136	2,344	792	1,136	908	300
R^2	0.363	0.475	0.107	0.627	0.402	0.241

Notes: This table reports results from estimating Equation 5 separately for different reasoning types (labeled using GPT-5.5). The dependent variable is the reported belief about the success probability for a project with variables $(X = x, Z = z)$. For each subject, there are four observations, corresponding to the four combinations of predictor values. Column 1 reports results for the full sample. Columns 2 and 3 report results separately for subjects with well-defined and undetermined types, respectively. Columns 4–6 report results for the three specific well-defined types: *Frequentist*, *Separate*, and *Absolute Success*. Clustered standard errors are in parentheses. * denotes significance at 10 pct., ** at 5 pct., and *** at 1 pct. level.

Table E.5 Path dependence by cognitive effort, controlling for reasoning type (AI Codes)

	Subjective success probability (pooled), by effort		
	Baseline	Controlling for reasoning	Controlling for reasoning × treatment
	(1)	(2)	(3)
High effort	−1.360	−0.920	−0.920
	(0.893)	(0.944)	(0.945)
Benchmark X	0.150***	0.140***	0.138***
	(0.042)	(0.042)	(0.042)
Benchmark X	0.163***	0.161***	0.165***
× X-first	(0.059)	(0.059)	(0.058)
Benchmark Z	0.392***	0.440***	0.439***
	(0.027)	(0.026)	(0.026)
Benchmark Z	0.010	−0.009	−0.009
× X-first	(0.039)	(0.037)	(0.038)
High effort	0.013	0.032	0.036
× Benchmark X	(0.060)	(0.062)	(0.065)
High effort	−0.144*	−0.140*	−0.146*
× Benchmark X × X-first	(0.079)	(0.078)	(0.081)
High effort	0.233***	0.141***	0.141***
× Benchmark Z	(0.040)	(0.037)	(0.039)
High effort	0.014	0.046	0.045
× Benchmark Z × X-first	(0.057)	(0.051)	(0.054)
Constant	48.028***	51.377***	51.377***
	(0.622)	(1.088)	(1.089)
Controls	–	Reasoning	Reasoning × X-first
Observations	3,136	3,136	3,136
R^2	0.383	0.439	0.440

Notes: This table analyzes the model formation in the second stage of the experiment by subjects' reasoning type (labeled using GPT-5.5) and attention allocation. It reports estimates of Equation 5 interacting all covariates with *high effort*, an indicator equal to 1 if a subject spent above-median relative time in the second stage, measured as the share of time spent in stage 2 relative to the total time spent across stages 1 and 2. Column 1 reports results without controlling for reasoning type. Column 2 additionally includes interactions between reasoning type and the benchmarks. Column 3 further adds interactions between reasoning type, the benchmarks, and treatment. The dependent variable is the reported belief about the success probability for a project with variables $(X = x, Z = z)$. For each subject, there are four observations, corresponding to the four combinations of predictor values. Clustered standard errors are in parentheses. * denotes significance at 10 pct., ** at 5 pct., and *** at 1 pct. level.

F Experimental Instructions

At the beginning of the survey, several eligibility checks were implemented. Subjects were only allowed to proceed if they accessed the survey via a computer device. Additionally, they were required to pass an attention check, provide informed consent, and successfully complete a CAPTCHA verification. Only after meeting all these criteria could subjects continue to the main survey. Upon completion of the survey, subjects reported demographic information and were given the opportunity to provide feedback.

Screenshots of the main survey are presented below from the perspective of a subject randomly assigned to the *Z-first* group, i.e., a subject who is presented with Card first. The instructions are identical across experimental conditions. The only difference is that subjects assigned to the *X-first* group are presented with Color in the first stage and with Card in the second stage. Accordingly, for subjects in the *X-first* group, the first-stage data table displays Color, and the corresponding instructions refer to Color rather than Card. In the second stage, Card is revealed as an additional feature. The only remaining difference between groups at this stage is that the ordering of project values reflects the sequence in which features were presented.

The screenshots shown below are taken from the *Baseline* experiments. The *SymmCorr* and *SymmUncorr* experiments differ primarily in their data-generating processes. Screenshots of the second-stage datasets for these experiments are provided in Section F.1.

Welcome

Thank you for participating in this study about your reasoning! This study consists of two parts and will take approximately **24 minutes** to complete. You will earn a **reward of $4** for completing the study in its entirety. To complete the study and earn the full reward, you have to read all instructions carefully, correctly answer the comprehension questions and pay attention during the entire study.

One out of every ten participants is eligible for an additional **bonus of up to $20**!

Instructions

In this survey, you will take on the role of an entrepreneur**.** Your task will be to evaluate and select potential projects to undertake based on past data**.** The data you will observe includes **the features and outcomes of past projects**.

Example data:
In the table below, you can observe example past data. Each entry in this example has the feature **Weather** (Sun ☼ or Cloud 🌩) and an **outcome** (**Success** or Failure).

N°	Weather	Outcome
E1	☼	**Success**
E2	☼	**Success**
E3	🌩	Failure
E4	🌩	**Success**
E5	☼	Failure
E6	🌩	Failure

Data structure:

- In the table columns, you can find the project **identifier,** its **features** and the **outcome**.
- Each feature can only take on two possible values (e.g. ☼ and 🌩), and the outcome is Success or Failure.
- **Each row** contains information about **1000 identical past projects**.

Role of features:

- A project's **success likelihood** is **determined by its features**.
- Changing the value of a feature (e.g. from ☼ to 🌩) can affect the success likelihood of a project.
- The success likelihood is the same across all projects with identical features.

Learning from past projects:

- The data on past projects is the only information you should use to determine the impact of features on project outcomes.
- The order of rows in the table does not matter.

Your Bonus:

- If eligible, one randomly selected task will determine your bonus payment of **up to $20**.
- The answers you provide influence the bonus payment you receive.

Comprehension questions

According to the example data, which statement is correct?

- ◯ For the example data, approximately **67%** of the projects **with symbol ⬟** are successful.
- ◯ For the example data, **100%** of the projects **with symbol ⬟** are successful.
- ◯ For the example data, approximately **33%** of the projects **with symbol ⬟** are successful.

Please select the correct statement.

- ◯ I **cannot use** the information about past projects to learn about the impact of features on project outcomes.
- ◯ I **can use** the information about past projects to learn about the impact of features on project outcomes.

Please select the correct statement.

- ◯ In the past data, each row represents **one thousand identical** past projects.
- ◯ In the past data, each row represents **a single** past project.
- ◯ In the past data, each row represents several past projects but **I cannot tell the exact number**.

Please select the correct statement.

- ◯ I cannot obtain a bonus payment in this study.
- ◯ The decisions I make throughout this study affect my bonus payment. This study has real stakes!
- ◯ The decisions I make throughout this study don't matter for my bonus payment.

Stage 1 [Example: *Z-first*]

Data

Each project listed in the table has **two features, Card and Color,** and an **outcome** (**Success** or Failure)**.** However, you can only observe information about one randomly determined feature. The feature you observe is **Card (Clubs ♣ or Diamonds ♦)**.

To **reveal the information** about past projects, please click on the button below and think about how the **feature Card** might **affect the outcome**.

You don't have to memorize the table of data as you **can access it at any later point by clicking the "Revisit past projects" button.**

Reveal past projects

N°	Card	Outcome	N°	Card	Outcome	N°	Card	Outcome	N°	Card	Outcome
1	♣	Failure	11	♣	Failure	21	♣	Failure	31	♣	Failure
2	♦	Success	12	♦	Success	22	♣	Success	32	♣	Failure
3	♦	Success	13	♣	Failure	23	♣	Failure	33	♣	Failure
4	♣	Failure	14	♦	Failure	24	♦	Success	34	♦	Success
5	♦	Success	15	♦	Success	25	♦	Success	35	♦	Success
6	♦	Success	16	♣	Failure	26	♣	Success	36	♣	Failure
7	♣	Failure	17	♦	Failure	27	♦	Success	37	♣	Success
8	♦	Failure	18	♦	Failure	28	♣	Failure	38	♦	Success
9	♦	Success	19	♦	Success	29	♣	Failure	39	♦	Success
10	♣	Failure	20	♣	Success	30	♦	Success	40	♣	Failure

Introduction decisions

Your next decisions revolve around **undertaking one of two potential projects**. The project you undertake will pay you **$10 if it is a Success** and $0 if it is a Failure.

Your next two tasks are:

1. You choose **the project you prefer to undertake.**
2. You indicate **how much you prefer** the chosen project.

Note for your bonus:
Both tasks are equally likely to determine your bonus.

Project choice

Revisit past projects

In the table below you can find **two potential future projects,** F1(♣) and F2(♦). Each project **pays $10 if it becomes a Success** and $0 if it becomes a Failure.

N°	Card	Outcome
F1	♣	?
F2	♦	?

▸ *Click here to learn more about the bonus*

Please select the project you prefer to undertake.

Note: If you are indifferent you can select either of the two projects.

○ F1(♣)

○ F2(♦)

WTP [Example: Previous choice = F2]

You just answered that your preferred project is F2(♦). Next, we are interested in **how much you prefer** project F2(♦) compared to project F1(♣), when the project you undertake will pay you **$10 if it is a Success** and **$0 if it is a Failure**.

Each row below is a distinct choice between either **project F2(♦) or project F1(♣) along with an increasing amount of money.** The amount shown in each row is an **additional payment regardless of the outcome** of the project.

For each row you will need to select which of the two options you prefer. Each choice is equally likely to be drawn to be relevant for your bonus payment.

Instructions:

- **Click on the row** with the **minimum amount** for which you would **switch to your less preferred project F1(♣).**
- The computer then automatically completes your choices, **highlighting the options you prefer.**
- The **more you prefer project F2(♦),** the **higher should be the row number** you select.
- If **you are indifferent** between either project, it is your best strategy to select the **first row.**

▸ *Click here to learn more about the bonus*

	Project F2(♦)	Project F1(♣)
1	F2(♦)	F1(♣) + $0
2	F2(♦)	F1(♣) + $0.5
3	F2(♦)	F1(♣) + $1
4	F2(♦)	F1(♣) + $1.5
5	F2(♦)	F1(♣) + $2
6	F2(♦)	F1(♣) + $2.5
7	F2(♦)	F1(♣) + $3
8	F2(♦)	F1(♣) + $3.5
9	F2(♦)	F1(♣) + $4
10	F2(♦)	F1(♣) + $4.5
11	F2(♦)	F1(♣) + $5
12	F2(♦)	F1(♣) + $5.5
13	F2(♦)	F1(♣) + $6
14	F2(♦)	F1(♣) + $6.5
15	F2(♦)	F1(♣) + $7
16	F2(♦)	F1(♣) + $7.5
17	F2(♦)	F1(♣) + $8
18	F2(♦)	F1(♣) + $8.5
19	F2(♦)	F1(♣) + $9
20	F2(♦)	F1(♣) + $9.5
21	F2(♦)	F1(♣) + $10

Submit Your Choices

Model elicitation (intensive margin)

Your next task is to **assess the likelihood of success** for the two potential future projects, F3(♣) and F4(♦).

N°	Card	Outcome
F3	♣	?
F4	♦	?

▸ *Click here to learn more about the bonus*

How likely do you think it is that **project F3(♣)** will be **successful**?

Never successful — Always successful

0% 10% 20% 30% 40% 50% 60% 70% 80% 90% 100%

How likely do you think it is that **project F4(♦)** will be **successful**?

Never successful — Always successful

0% 10% 20% 30% 40% 50% 60% 70% 80% 90% 100%

How certain are you that all your above-stated **project assessments** are within +/- 5 percentage points of the true success likelihoods?

Not at all certain — Extremely certain

0% 10% 20% 30% 40% 50% 60% 70% 80% 90% 100%

Stage 2 [Example: *Z-first*]

Introduction part 2

You now entered the second part of this study!

On the next screen you will be shown a table with information about the **same 40 past projects** as in the first part but you now observe additional information about the second **feature Color (Green 🟢 or Blue 🔵)** which you were unable to observe in the first part.

Before proceeding to the next page, please select the option that best describes the data you will see next.

- ○ The data I will see is unrelated to the data I saw in the first part.
- ○ The data I will see contains information on potential projects that exhibit the same relationship between features and outcome as in the first part.
- ○ The data I will see contains information on the same projects as before, but now with an additional feature that I was unable to see before.

Data

To **reveal** the previously unavailable **feature Color (Green 🟢 or Blue 🔵)** in past projects, please click on the button below and think about how the **features** might **affect the outcome**.

You don't have to memorize the table of data as you **can access it at any later point by clicking the "Revisit past projects" button.**

Reveal past projects

N°	Card	Color	Outcome
1	♣	🟢	Failure
2	♦	🔵	Success
3	♦	🔵	Success
4	♣	🟢	Failure
5	♦	🟢	Success
6	♦	🔵	Success
7	♣	🟢	Failure
8	♦	🔵	Failure
9	♦	🔵	Success
10	♣	🟢	Failure

N°	Card	Color	Outcome
11	♣	🟢	Failure
12	♦	🔵	Success
13	♣	🟢	Failure
14	♦	🟢	Failure
15	♦	🔵	Success
16	♣	🔵	Failure
17	♦	🔵	Failure
18	♦	🔵	Failure
19	♦	🟢	Success
20	♣	🟢	Success

N°	Card	Color	Outcome
21	♣	🟢	Failure
22	♣	🟢	Success
23	♣	🟢	Failure
24	♦	🔵	Success
25	♦	🔵	Success
26	♣	🔵	Success
27	♦	🔵	Success
28	♣	🟢	Failure
29	♣	🔵	Failure
30	♦	🟢	Success

N°	Card	Color	Outcome
31	♣	🟢	Failure
32	♣	🔵	Failure
33	♣	🔵	Failure
34	♦	🔵	Success
35	♦	🔵	Success
36	♣	🟢	Failure
37	♣	🟢	Success
38	♦	🔵	Success
39	♦	🟢	Success
40	♣	🟢	Failure

Introduction decisions

Similar to the first part, your next decisions revolve around **undertaking potential projects.** In total you will make decisions for **four pairs of potential projects**.

For each pair of projects, your tasks are the following:

1. You choose **the project you prefer to undertake.**
2. You indicate **how much you prefer** the chosen project.

Note for your bonus:
Both tasks are equally likely to determine your bonus.

Project choice [Example: P1(♦,🔵) vs. P2(♣,🔵)]

Revisit past projects

In the table below you can find **two potential future projects,** P1(♦,🔵) and P2(♣,🔵). Each project **pays $10 if it becomes a Success** and $0 if it becomes a Failure.

N°	Card	Color	Outcome
P1	♦	🔵	?
P2	♣	🔵	?

▸ *Click here to learn more about the bonus*

Please select the project you prefer to undertake.

Note: If you are indifferent you can select either of the two projects.

○ P1(♦,🔵)

○ P2(♣,🔵)

WTP [Example: Previous choice = P1]

You just answered that your preferred project is P1(♦,🔵). Next, we are interested in **how much you prefer** project P1(♦,🔵) compared to project P2(♣,🔵), when the project you undertake will pay you **$10 if it is a Success** and **$0 if it is a Failure**.

Each row below is a distinct choice between either **project P1(♦,🔵) or project P2(♣,🔵) along with an increasing amount of money.** The amount shown in each row is an **additional payment regardless of the outcome** of the project.

For each row you will need to select which of the two options you prefer. Each choice is equally likely to be drawn to be relevant for your bonus payment.

Instructions:

- **Click on the row** with the **minimum amount** for which you would **switch to your less preferred project P2(♣,🔵).**
- The computer then automatically completes your choices, **highlighting the options you prefer.**
- The **more you prefer project P1(♦,🔵),** the **higher should be the row number** you select.
- If **you are indifferent** between either project, it is your best strategy to select the **first row.**

▸ *Click here to learn more about the bonus*

	Project P1(♦,🔵)	Project P2(♣,🔵)
1	P1(♦,🔵)	P2(♣,🔵) + $0
2	P1(♦,🔵)	P2(♣,🔵) + $0.5
3	P1(♦,🔵)	P2(♣,🔵) + $1
4	P1(♦,🔵)	P2(♣,🔵) + $1.5
5	P1(♦,🔵)	P2(♣,🔵) + $2
6	P1(♦,🔵)	P2(♣,🔵) + $2.5
7	P1(♦,🔵)	P2(♣,🔵) + $3
8	P1(♦,🔵)	P2(♣,🔵) + $3.5
9	P1(♦,🔵)	P2(♣,🔵) + $4
10	P1(♦,🔵)	P2(♣,🔵) + $4.5
11	P1(♦,🔵)	P2(♣,🔵) + $5
12	P1(♦,🔵)	P2(♣,🔵) + $5.5
13	P1(♦,🔵)	P2(♣,🔵) + $6
14	P1(♦,🔵)	P2(♣,🔵) + $6.5
15	P1(♦,🔵)	P2(♣,🔵) + $7
16	P1(♦,🔵)	P2(♣,🔵) + $7.5
17	P1(♦,🔵)	P2(♣,🔵) + $8
18	P1(♦,🔵)	P2(♣,🔵) + $8.5
19	P1(♦,🔵)	P2(♣,🔵) + $9
20	P1(♦,🔵)	P2(♣,🔵) + $9.5
21	P1(♦,🔵)	P2(♣,🔵) + $10

Submit Your Choices

Model elicitation (intensive margin)

Revisit past projects

Your next task is to **assess the likelihood of success** for the four potential future projects P9(♣,🟢), P10(♣,🔵), P11(♦,🟢) and P12(♦,🔵).

N°	Card	Color	Outcome
P9	♣	🟢	?
P10	♣	🔵	?
P11	♦	🟢	?
P12	♦	🔵	?

▸ *Click here to learn more about the bonus*

How likely do you think it is that **P9(♣,🟢) will be successful**?

Never successful — Always successful

0% 10% 20% 30% 40% 50% 60% 70% 80% 90% 100%

How likely do you think it is that **P10(♣,🔵) will be successful**?

Never successful — Always successful

0% 10% 20% 30% 40% 50% 60% 70% 80% 90% 100%

How likely do you think it is that **P11(♦,🟢) will be successful**?

Never successful — Always successful

0% 10% 20% 30% 40% 50% 60% 70% 80% 90% 100%

How likely do you think it is that **P12(♦,🔵) will be successful**?

Never successful — Always successful

0% 10% 20% 30% 40% 50% 60% 70% 80% 90% 100%

How certain are you that all your above-stated **project assessments** are within +/- 5 percentage points of the true success likelihoods?

Not at all certain — Extremely certain

0% 10% 20% 30% 40% 50% 60% 70% 80% 90% 100%

Reasoning

Revisit past projects

On the last pages, you made decisions based on your perceived success likelihood of different projects.

Please describe how you determined the projects' success likelihoods.
You should explicitly state what you **paid attention to** and which **strategy you used** to arrive at your response in full sentences.

Model elicitation (extensive margin)

You have almost finished this study! Please assess the statements below about how the features affect the outcome.

Assuming that a project's Card remains fixed, changing a project's Color (Green 🟢 or Blue 🔵) has an effect on the project's success probability.

- ◯ True
- ◯ False

How certain are you about your above answer?

Not at all certain — Extremely certain

0% 10% 20% 30% 40% 50% 60% 70% 80% 90% 100%

Assuming that a project's Color remains fixed, changing a project's Card (Clubs ♣ or Diamonds ♦) has an effect on the project's success probability.

- ◯ True
- ◯ False

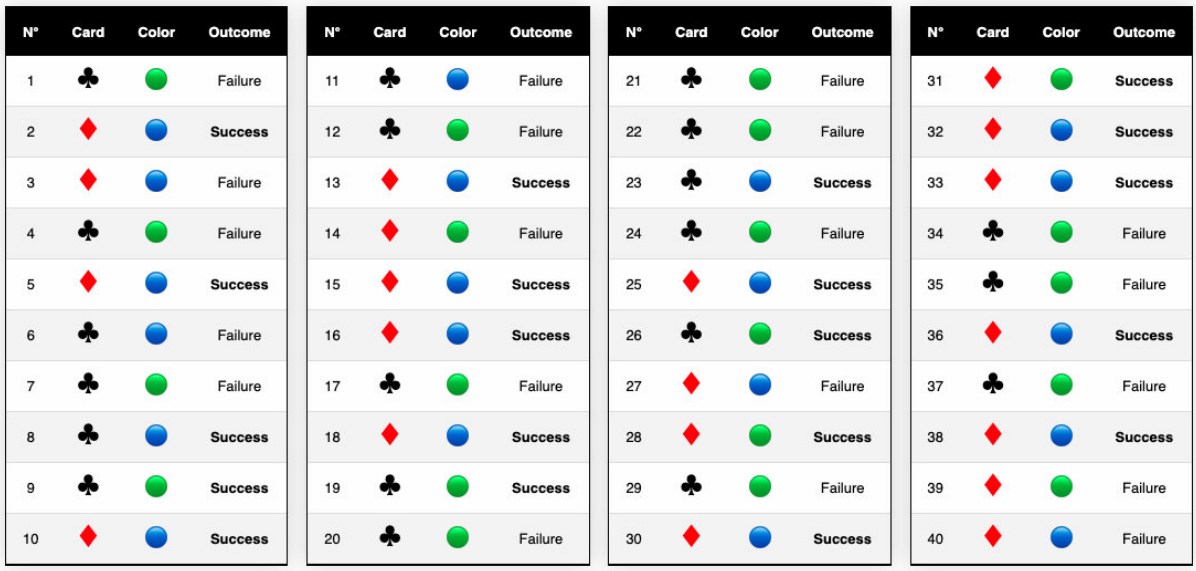
How certain are you about your above answer?

Not at all certain — Extremely certain

0% 10% 20% 30% 40% 50% 60% 70% 80% 90% 100%

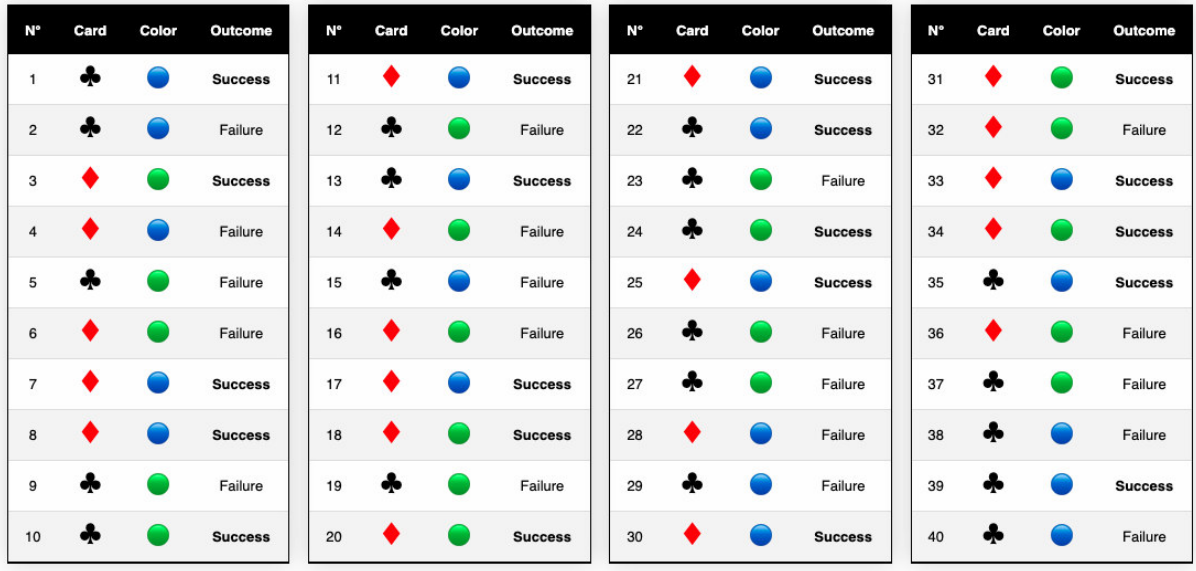
F.1 Datasets for Additional DGPs

Figures F.1 and F.2 show the second-stage datasets for the *SymmCorr* and *SymmUncorr* experiments, respectively. As in *Baseline* (Figure 2), subjects observe 40 past projects with two binary features (Card and Color) and a binary outcome. The experimental interface is identical across all three experiments; only the underlying data-generating process differs.



N°	Card	Color	Outcome	N°	Card	Color	Outcome	N°	Card	Color	Outcome	N°	Card	Color	Outcome
1	♣	green	Failure	11	♣	blue	Failure	21	♣	green	Failure	31	♦	green	Success
2	♦	blue	Success	12	♣	green	Failure	22	♣	green	Failure	32	♦	blue	Success
3	♦	blue	Failure	13	♦	blue	Success	23	♣	blue	Success	33	♦	blue	Success
4	♣	green	Failure	14	♦	green	Failure	24	♣	green	Failure	34	♣	green	Failure
5	♦	blue	Success	15	♦	blue	Success	25	♦	blue	Success	35	♣	green	Failure
6	♣	blue	Failure	16	♦	blue	Success	26	♣	green	Success	36	♦	blue	Success
7	♣	green	Failure	17	♣	green	Failure	27	♦	blue	Failure	37	♣	green	Failure
8	♣	blue	Success	18	♦	blue	Success	28	♦	green	Success	38	♦	blue	Success
9	♣	green	Success	19	♣	green	Success	29	♣	green	Failure	39	♦	green	Failure
10	♦	blue	Success	20	♣	green	Failure	30	♦	blue	Success	40	♦	blue	Failure

Figure F.1 Screenshot of data as observed in the second stage — *SymmCorr*



N°	Card	Color	Outcome	N°	Card	Color	Outcome	N°	Card	Color	Outcome	N°	Card	Color	Outcome
1	♣	blue	Success	11	♦	blue	Success	21	♦	blue	Success	31	♦	green	Success
2	♣	blue	Failure	12	♣	green	Failure	22	♣	blue	Success	32	♦	green	Failure
3	♦	green	Success	13	♣	blue	Success	23	♣	green	Failure	33	♦	blue	Success
4	♦	blue	Failure	14	♦	green	Failure	24	♣	green	Success	34	♦	green	Success
5	♣	green	Failure	15	♣	blue	Failure	25	♦	blue	Success	35	♣	blue	Success
6	♦	green	Failure	16	♦	green	Failure	26	♣	green	Failure	36	♦	green	Failure
7	♦	blue	Success	17	♦	blue	Success	27	♣	green	Failure	37	♣	green	Failure
8	♦	blue	Success	18	♦	green	Success	28	♦	blue	Failure	38	♣	blue	Failure
9	♣	green	Failure	19	♣	green	Failure	29	♣	blue	Failure	39	♣	blue	Success
10	♣	green	Success	20	♦	green	Success	30	♦	blue	Success	40	♣	blue	Failure

Figure F.2 Screenshot of data as observed in the second stage — *SymmUncorr*